A SPATIAL TEMPORAL ANALYSIS OF SURVIVAL AMONG TB AND HIV CO–INFECTED PATIENTS IN ERONGO REGION, NAMIBIA

A THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOSTATISTICS

OF

THE UNIVERSITY OF NAMIBIA

BY

ANDREAS ITASHIPU SHIPANGA

200524526

APRIL 2019

SUPERVISOR: PROF. LAWRENCE KAZEMBE (UNAM – DEPARTMENT OF STATISTICS AND POPULATION STUDIES)

Abstract

Namibia is among the Sub-Sahara African countries with a high burden of TB and HIV coinfection in the world. The interaction between TB and HIV/AIDS makes the diagnosis and management of the co-infection difficult. The purpose of this study is to identify risk factors, assess spatial and space-time clusters mortality among TB and HIV in Erongo Region. A retrospective cohort study from 2003 to 2017 was carried out using data consisting of 3145 subjects from all the 16 health facilities providing ART and TB management in Erongo Region, Namibia. Firstly, the study used the Kaplan Meier to estimate the survival function. Cox PH model was used to model survival and identify risk factors among TB and HIV co-infected patients. Global Moran's I and LISA statistics were used to help analyse the spatial distribution and clusters of the disease across settings. Finally, the study used Bayesian Inference in STAR models to assess the spatial and space-time clusters of TB and HIV Mortality. Of the 3145 patients, 424 (13.5%) defaulted treatment, 42 (1.3%) stopped treatment, 542 (17.2%) were transferred out and 1673 (53.2%) were alive at the end of the study. About 464 (14.8%) patients experienced event of interest (death), during the period. Only marital status and sex were not associated with death in the univariate analysis with a p-value of 0.30 and 0.07 respectively. Mortality was higher in Usakos district as compared to the other 3 health district in the region. Mortality worsened with increasing WHO clinical stage. In multivariate analysis, the covariate Function violated the PH assumptions with a p-value of 0.01. Global Moran's I shows that the EAs tend clusters, 0.17 (p-value<0.001) and the presence of hot spots in Erongo Region. Mortality among TB-HIV co-infected patients accounted for a considerable number of deaths among the cohort. The results have shown the spatial distribution of TB and HIV mortality in the region over the years. The detection of space-time clustering was useful in identifying higher risk areas. From this study it is recommended that there is a need to have targeted intervention among these areas to ensure that Namibia strives to attain its NDP5 which is intended to reduce mortality of TB among HIV patients. Patients aged 50+ and those with bedridden functional status should be strictly followed to reduce mortality. This information can be used by the region for planning purposes, allocation of resources and even dissemination of TB and HIV information.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and/or Acronyms

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
CAR	Conditional Autoregressive
CDC	Centre for Disease and Prevention
CI	Confidence Interval
CIF	Cumulative Incidence Function
DIC	Deviance Information Criterion
DTLC	District Tuberculosis and Leprosy Coordinator
EA	Enumeration Area
ePMS	Electronic Patient Monitoring System
GOF	Goodness of Fit
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
IID	Independently Identically Distributed
INLA	Integrated Nested Laplace Approximation
KG	Kilograms
KM	Kaplan-Meier
LISA	Local Indicators of Spatial Association
MCMC	Markov Chain Monte Carlo
MoHSS	Ministry of Health and Social Services
NDP5	Fifth National Development Plan
NSA	Namibia Statistics Agency
PEPFAR	The U.S. President's Emergency Plan for AIDS Relief
PH	Proportional Hazard
PLHIV	People Living with HIV
SHR	Subdistribution Hazard Ratio
TB	Tuberculosis
TO	Transferred Out
UNAIDS	United Nation AIDS
VB.NET	Visual Basic .NET
WBCG	Walvisbay Corridor Group
WHO	World Health Organization

Acknowledgements

I would like to express my thoughtful appreciation to my supervisor Prof. Lawrence Kazembe for his vital advice and support. He has not only assisted me with his excellent advice, but he also elevated my interest in spatial survival analysis, introduced me to a new statistical software package and improved my comprehension of theoretical Biostatistics. This research would not have been possible without his contribution and encouragement.

I am also grateful all my lecturers and to everyone at the department for adding some pleasure to the work, especially to all my classmates (Alexandrina Petrus, Simon Kashihalwa, Job Shikongo, Jason Nakaluudhe and Adolf Haufiku) for their assistance and the friendship that we have formed during our two years together.

I would also like to thank my parents and my grandparents without whom I would never have been here to complete this work. Thanks to my two uncles Immanuel Kauluma Shipanga and Metumo Shipanga, for their constant support, their understanding during the more difficult spells, and for always keeping me positive and cheerful.

My sincere gratitude to Miss Paulina Ndjambula the Chief Health Programme Officer for Special Program in Erongo Region for the support you have given me during my study and specifically for proving me with information on TB and HIV for Erongo Region.

Declarations

I, **Andreas Itashipu Shipanga**, 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.

…….	……	……..
Name of Student	Signature	Date

Chapter 1

Introduction

1.1 Background of the study

The Human Immunodeficiency Virus (HIV) pandemic presents a significant challenge to global tuberculosis (TB) control. The recent TB fact sheet by the World Health Organization indicates that people who are HIV positive are 20 to 30 times more likely to develop active TB (WHO, 2018). For HIV positive patients, being diagnosed with TB is a sign of progression to AIDS (CDC, 2017). TB is a leading preventable cause of death among people living with HIV, causing more than one third of all AIDS-related deaths in 2015 and almost 60% of the estimated global HIV-related TB cases are not diagnosed and not treated (UNAIDS, 2017). Controlling and preventing HIV/AIDS, tuberculosis and other opportunistic infections are among the key global health priorities, particularly in developing countries such as Namibia.

The World Health Organization (WHO) recommends 12 collaborative HIV/TB activities, including the "Three I's for HIV/TB" (isoniazid preventive treatment [IPT], intensified case finding, and infection control for TB), which should be seen as core prevention, care, and treatment services for HIV infection. Of the 12 activities, there has been progress in implementing testing for HIV infection, providing trimethoprim-sulfamethoxazole preventive therapy, and antiretroviral therapy (ART) (Getahun, Gunneberg, Granich, & Nunn, 2010). Globally, the WHO estimated that, in 2008, only 20% of people living with HIV knew their status (WHO, 2018). In 2017, an estimated 920 000 people living with HIV have fallen ill with TB, while 22% of patients with TB were tested for HIV infection (WHO, 2018).

Namibia is the fourth worst TB affected country in the world, with 9 882 patients diagnosed with the disease in 2014 (MoHSS, 2016a). Namibia also remains one of the country with the

highest HIV prevalence rate in the world, according to the Surveillance Report of the 2016 National HIV Sentinel Survey (MoHSS, 2016c) the prevalence rate was at 17.2%. The country has one of the highest HIV/TB co-infection rates in the world, meaning that on average, about 4 out of every 10 patients with tuberculosis are also HIV-positive (MoHSS, 2016a). Even though the government of Namibia has put up several strategies to reduce TB and HIV/AIDS incidence, the two diseases remain some of the highest causes of death in Namibia (WHO, 2018).

Despite progress made towards reducing the burden of tuberculosis (TB) in Namibia, the disease remains a major public health concern in the country. While there has been a consistent decrease in reported TB cases since 2004, it is noteworthy that the 2016 World Health Organisation (WHO) *Global Tuberculosis Report* included Namibia among the 30 countries with the highest TB burden globally, due to the country's high estimated per capita TB incidence. The country is also included among the countries with very high rates of TB/HIV co-infection, with an HIV prevalence of 38% among TB patients in 2016 (MoHSS, 2017b).

1.2 TB and HIV Epidemiology

Tuberculosis is common in people with impaired immunity such as HIV positive persons, diabetic, malnourished and tobacco smokers and is one of the very infectious bacterial diseases (Narasimhan, Wood, MacIntyre, & Mathai, 2013). The disease is spread through air when an infected person coughs, speaks or sings. It is one of the most common opportunistic infections in persons living with HIV. The recent TB fact sheet by the World Health Organisation indicates that people who are HIV positive are 20 to 30 times more likely to develop active TB (WHO, 2018b). For HIV positive patients, being diagnosed with TB is a sign of progression to AIDS (CDC, 2017a). The World Health Organisation reports that TB is a leading killer of HIV

patients, with one in every three deaths among HIV patients was due to TB (WHO, 2018b). Tuberculosis (TB) is a treatable disease but remains the leading cause of death among persons living with HIV/AIDS (PLHIV). In some settings, up to 50% of patients with both TB disease and HIV infection die during TB treatment with most deaths occurring within 2 months of being diagnosed with TB (Pevzner, Vandebriel, Lowrance, Gasana, & Finlay, 2011).

Coverage of HIV services for TB patients continue to improve, with 95% of TB patients having a known HIV status (MoHSS, 2016a). About 40% of the TB patients were co-infected with HIV in 2015, a marginal decrease from 44% in 2014 (MoHSS, 2016a). The majority (92%) of HIV infected TB patients were on Antiretroviral Therapy (ART). The treatment success rate for TB/HIV patients was however relatively low (81% for the 2015 cohort), primarily due to a relatively high (11%) death rate (MoHSS, 2016a). Other contributing factors include poverty and associated risk factors (overcrowding, poor housing conditions, poor nutrition and delays in seeking health care), smoking, alcohol use and general community transmission due to the high prevalence of the two diseases in the country (MoHSS, 2017b).

Life expectancy at birth in Namibia has in the recent years declined to about 62 years mainly because of the impact of the HIV and AIDs pandemic, the major causes of mortality and morbidity, some of the notable causes of death in Namibia are mostly, Lower Respiratory Infections, followed by Ischemic Heart Disease and TB on its own ranked fourth (CDC, 2017). Tuberculosis and HIV, once thought to be on decline, has of late reportedly a constant trend in the past few years (2010-2016) and this has been exacerbated by HIV and AIDS and other infectious diseases such as TB (MoHSS, 2016c). The global TB treatment success rates range from 30 to 83% but it is much lower in HIV cohorts (Belayneh, Giday, & Lemma, 2015). HIV positive persons receiving ART have reduced risk of developing TB as several studies have

reported the protective effective of ART on opportunistic illness, with a reduction in risk of death ranging from 70 – 90 % (Vijay, Kumar, Chauhan, Rao, & Vaidyanathan, 2011).

In an effort to combat TB occurrence, several campaigns have been initiated all aiming at eliminating the TB disease. These campaigns include adoption of universal ART provision, WHO End TB strategy which targets 35% reduction in TB cases by 2020 and the UN zero campaigns targeting TB, AIDS, Malaria and poverty among others (WHO, 2016a). In Namibia the TB treatment success rate was estimated at 83% in 2015 which is lower than the WHO target of 90%, the main reason for this success rate is due to the reduction in funding the country has received from international donors such as the Global Fund and PEPFAR (Menges, 2018).

1.3 Statement of the problem

HIV co-infection is driving the TB epidemic in many countries and TB in high HIV prevalence areas is a leading cause of morbidity and mortality among the HIV-infected patients (Churchyard & Lawn, 2009). Erongo Region, situated at the coast of Namibia, is considered as one of the hotspot towns of HIV epidemic in Namibia. This is necessitated by the fact that Erongo Region is home to Walvisbay which is a harbour town providing port facility for the import and export of cargo to the rest of Namibia. According to the Surveillance Report of the 2016 National HIV Sentinel Survey (MoHSS, 2016c) the region had a HIV prevalence rate of 15.2% and making it one of the highest affected by the HIV/AIDS epidemic in the country. In 2015, the region reported third highest TB cases in Namibia only behind Khomas and Ohangwena region. This makes it one of the hotspots of the HIV epidemic and other opportunistic diseases in Namibia. Erongo Region started providing TB and Antiretroviral Therapy (ART) treatment to its people in 2003, but there have never been and in-depth analysis of the trends in survival of people on TB and ART management, particularly in relation to

localities within the region. The U.S. President's Emergency Plan for AIDS Relief (PEPFAR) which has supported the region in the provision of general medical support has for the purpose of their project prepared reports that mostly focuses on adherence of patients on treatment. Further studies are therefore required to understand the factors associated with survival and clustering based on individual level locations and investigation of cases and mortality of TB and HIV co-infections.

1.4 Objectives of the study

The main objective of the study is to identify risk factors, assess the spatial distribution, presence of the spatial temporal clustering as well as survival of TB and HIV co-infections in Erongo Region, Namibia.

Specific objectives of the study will be:

a) To model survival of patients on Antiretroviral Therapy (ART) and TB treatment.
b) To investigate spatial pattern of deaths of people infected with both TB and HIV.
c) To examine the temporal distribution of TB and HIV deaths in the region in order to identify possible high-risk areas.
d) To determine factors associated with mortality outcomes adjusting for their multivariate spatial random effects

1.5 Significance of the study

Although several studies reported TB-HIV co-infection in many African countries including Namibia, to our knowledge, there are few reports on TB-HIV co-infection from survival analysis perspective in Namibia. Here, the study will report for the first time, the prevalence of TB-HIV co-infection among pulmonary TB suspects as well as TB patients in the Erongo Region, a predominantly pastoral region in the western of Namibia. It is anticipated that

findings from this study will contribute to the body of knowledge that informs TB/HIV program planners, decision makers, and project implementers by providing predictors of mortality among TB and HIV co-infected patients.

The study will also give the current status of TB and HIV mortality in the region which inline contribute to the overall vision and mission of the National Health Policy Framework 2010-2020, the fifth National Development Plan (NDP5), Vision 2030 and the Harambee Prosperity Plan 2016/17 - 2019/20, which intend to reduce the TB and HIV mortality rate in Namibia. Achievement of these targets will require a paradigm shift from a health sector focused approach to a multi-sectoral approach to HIV and TB care and prevention.

1.6 Organization of the thesis

The report will address an introduction in Chapter 1 as above then data background and description in Chapter 2. The data is analysed in three methods, survival analysis, spatial and space-time modelling, therefore Chapter 3 and Chapter 4 concentrate on data analysis using the three methods respectively, which starts with the introduction of the method then the method modelling, then the application and finally the discussions. Conclusion, recommendations and future research direction comes in Chapter 5. Finally the Appendix includes STATA, R and BayeX codes used for analysis.

Chapter 2

2. Research methods and data overview

2.1 Introduction

Namibia has a three-tier health delivery system consisting of primary (health care), secondary (district level) and tertiary (referral) levels of care. Ownership of the facilities is largely by the public sector (Ministry of Health and Social Services), although reasonable number is also managed by non-governmental organizations, such as NAPPA, WBCG and the private sector. Erongo Region in particular has 24 public health facilities, of which 4 are district hospital, 2 health centres and 18 clinics. There are 16 ART sites and four (4) TB clinics.

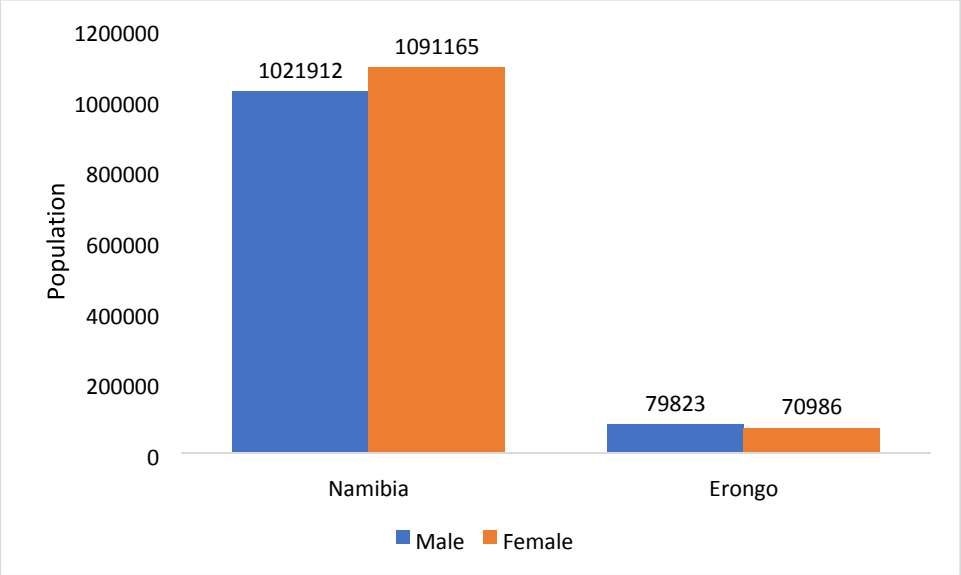
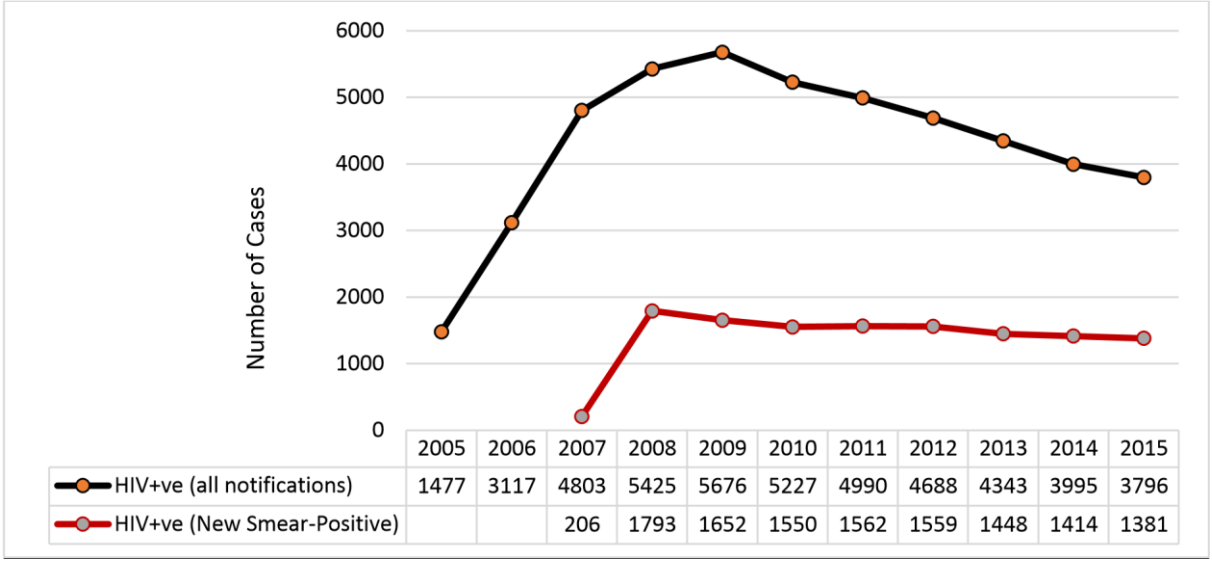


Figure 1: Namibian total population by sex

According to the National Tuberculosis and Leprosy Programme annual report of 2016, 95% of TB patients were tested for HIV, 92% of TB/HIV, co-infected patients were on antiretroviral therapy (ART) and 96% were put on cotrimoxazole preventive therapy (CPT). Figure 1, shows that the number of TB patients who are HIV infected has been declining since 2009. This decline has been however less marked among smear-positive TB patients.



	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
HIV+ve (all notifications)	1477	3117	4803	5425	5676	5227	4990	4688	4343	3995	3796
HIV+ve (New Smear-Positive)			206	1793	1652	1550	1562	1559	1448	1414	1381

Figure 2: Trends in TB/HIV case notifications in Namibia, 2005 – 2015

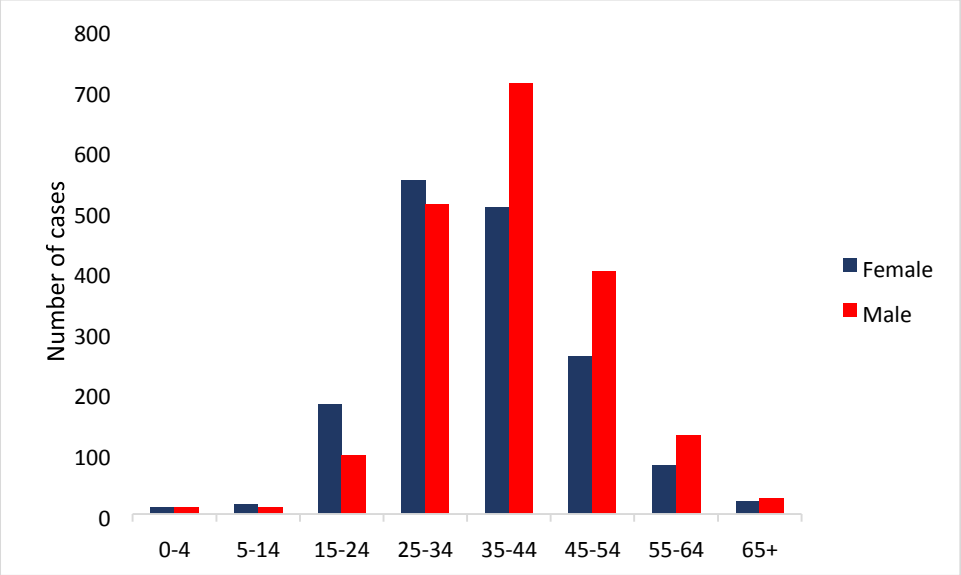


Figure 3: Age-sex distribution among new and relapse TB/HIV cases in Namibia, 2015

The age-sex distribution of TB among PLHIV largely mirrors the overall age-sex distribution of TB in the country. It is however noteworthy that there is a female preponderance in the 5-14, 15-24 and 25-34 age groups as shown in Figure 2. The majority (95%) of patients registered for TB treatment in 2015 had a documented HIV status; 3796 (40%) of these were HIV positive

(MoHSS, 2016a). These has also been significant increase in the number initiated on ART (from 84% in 2014 to 92% in 2015) as indicated by Table 1.

The National Tuberculosis and Leprosy Programme annual report of 2016 showed that the TB treatment success rate among TB/HIV co-infected patients showed a marginal improvement from 79% to 82% for 2013 and 2014 cohorts respectively. The death rate for all forms of TB among TB/HIV patients (11%) is higher than that among all TB patients in Namibia. According to the (NSA, 2016) about 33% of households in Namibia are less than 1 kilometre to the nearest hospital or clinic and 32% are between 2 and 5 kilometre. However, 4.60% have to travel more than 40 km to reach a hospital or clinic. Urban households travel shorter distances: 48.4% within 1 kilometre compared to rural households with 15%. This could be an attributing factor, especially with the high defaulter rate among TB and HIV co-infected patients in Erongo Region.

Table 1: Summary of TB/HIV indicators by region, 2014

Region	**All forms of TB (number)**	**Known HIV Status**		**HIV-Positive TB patients**		**HIV-positive TB patients on CPT**		**HIV-positive TB patients on ART**	
		Number	**Percentage**	**Number**	**Percentage**	**Number**	**Percentage**	**Number**	**Percentage**
Erongo	1074	1014	94	399	39	395	99	348	87
Hardap	677	612	92	216	35	211	98	190	87
//Karas	579	568	98	248	44	248	100	226	91
Kavango	854	780	91	325	42	324	100	325	100
Khomas	1602	1542	96	676	44	663	98	528	78
Kunene	321	315	98	104	33	104	100	101	97
Ohangwena	1254	1246	99	391	31	391	100	387	99
Omaheke	472	446	94	133	30	133	100	129	97
Omusati	717	665	93	252	38	252	100	239	95
Oshana	527	521	99	232	45	232	100	222	96
Oshikoto	747	672	90	311	46	311	100	289	93
Otjozondjupa	601	594	99	208	35	205	99	195	94
Zambezi	528	481	91	301	63	301	100	301	100
Namibia	**9944**	**9456**	**95**	**3796**	**40**	**3770**	**99**	**3480**	**92**

2.2 Research design

A retrospective cohort study was adopted for the study. Moreover, the study followed a cross-sectional research design.

2.3 Population sample

The target population of the study included all the patients on co-therapy TB and ART management who attended the health facilities between, 01st January 2003 and 31st December 2017. However, the study used a sample of all the available TB and HIV co-infected patient records on the Electronic Patient Monitoring System (ePMS) database throughout all the facilities in Erongo Region.

2.4 Procedure

The data were extracted from the ePMS database at MoHSS Erongo Regional Office into Microsoft Excel 2016. TB and HIV data were captured onto the paper based form by the clinician conducting diagnosis during the visit of the patient which results into enrolment into the TB and HIV program. The data forms were then entered into the administrative database in the health management information system section where the database is kept. Considering that subjects who started the treatment outside Erongo Region might have started the treatment under totally different conditions, data relating to patients who were initiated outside Erongo Region (the transfer in) were excluded from the analysis.

2.5 Variable for the study

The outcome variable considered for the study was the survival outcome which was time to death measured from the time of co-infection to death or censored in years. However, time to death were censored for those co-infected patients who were lost to follow up, transferred to

another health facility and did not died at 31st of December 2017 (at end of the study). In general, the independent covariates considered for the separate survival, spatial and space-time modelling are listed in Table 2 below.

Table 2: List of covariates used in the study

Variable name	**Values of the variable**	**Type**
Age	Years (baseline)	Continuous
Sex	Female, male	Categorical
CD4	CD4 counts	Continuous
Still on care	Alive, dead, default, stopped and transferred out	Categorical
Weight	Weight (baseline)	Continuous
Status of patient	Alive, dead	Categorical
Marital status	Single, married, divorced and widowed	Categorical
Function	Ambulatory, working and bedridden	Categorical
WHO clinical stage	Stage I, II, II, IV	Categorical
Facility level	Clinic, health centre and hospital	Categorical

WHO Clinical Stage which is classified into four; I, II, III and IV; where Stage I indicates asymptomatic disease, Stage II indicates mild disease, Stage III indicates advanced disease and Stage IV indicates severe disease. Hence disease severity increases from Stage I to Stage IV. Functional Status of the patients is also categorical covariate with three categories: Working, Ambulatory and Bedridden. Working patients are those patients who are able to work day to day while ambulatory patients are those patients who are able to work some time but bedridden patients cannot able to work due to the infectious disease.

2.6 Research ethics

Ethical approval of this study was obtained from the Research Ethical Committee of the University of Namibia from The Centre of Postgraduate Studies and Ministry of Health and Social Services. Extracted Data from ePMS did not include patient's identity and are kept in locked computer.

Chapter 3

Survival analysis of time-to-death among TB and HIV co-infected patients

3.1 Introduction

TB and HIV co-infection causes a serious bidirectional and synergistic combination of illness in which HIV promotes the progression of latent tuberculosis infection to disease, and TB accelerates the progression of HIV disease to poor prognosis including death (Gesesew, et al., 2016). In many circumstances, HIV has been described as the main reason for failure to meet TB control targets in high HIV settings, and TB is a major cause of death among people living with HIV in similar settings (WHO, 2018). There is a need to model for mortality using Time-to-event data as it shows how the two diseases are associated with each other and how they affect outcome of patients on co-therapy of TB and HIV. Time-to-event data are unique because the outcome of interest is not only whether or not and event occur, but also when the event occurred (Gesesew, et al., 2016).

Some studies have concluded that early initiation of ART during treatment for TB improves survival among patients, (Karim, et al., 2011) (Tabarsi, et al., 2009) (Odone, et al., 2014) whereas other studies have not found this same benefit (Mfinanga, et al., 2014) (Rosen, et al., 2016). Based on the best available evidence and expert opinion, the World Health Organization (WHO) issued guidelines in 2017 recommending that ART should be initiated for all HIV-infected patients with active TB disease within the first 8 weeks of starting TB treatment regardless of the CD4+ T cell count (WHO, 2009). Apart from the effects on mortality early initiation of ART in TB-HIV co-infected patients reduces the incidence of tuberculosis across all CD + T cell count levels (Suthar, et al., 2012).

A wide range of factors that influence TB/HIV coinfection mortality have been reported. These include: age, gender, marital status, level of education, religion, occupation, residence, weight, AIDS staging, TB clinical presentation and calendar year (Catala, et al., 2011). Previous studies have focused mainly on the survival rates of general HIV patients, giving less attention to TB/HIV co-infected patients (Tadesse, Haile, & Hiruy, 2014). Some studies (O'Donnell, et al., 2013) (Manda, Masenyetse, Lancaster, & van der, 2013) took into account drug susceptibility patterns of Mycobacterium tuberculosis outside Namibia. However, these studies assessed the mortality rates of HIV positive and negative patients on anti-TB treatment. To date, there have been limited clinical data regarding mortality rates among TB/HIV co-infected patients and the impact of antiretroviral therapy (ART) on clinical outcomes in Namibia.

Competing risks are frequently encountered in survival analysis, it arises when a failure can result from one of several causes and one cause precludes the others. It alters the probability of occurrence of other event. For example, a HIV/AIDS patient who is at risk of dying from tuberculosis, he/she is also at risk of dying due to causes like diarrhoea or other infection (Grover, Swain, & Ravi, 2014). Several authors have attempted to study competing risks theory and the estimation of cumulative incidence function of an event of interest. When estimating the crude incidence of outcomes, analysts should use the cumulative incidence function, rather than the complement of the Kaplan-Meier survival function. The use of the Kaplan-Meier survival function results in estimates of incidence that are biased upward, regardless of whether the competing events are independent of one another (Bakoyannis & Touloumi, 2012).

When fitting regression models in the presence of competing risks, researchers can choose from two different families of models: modelling the effect of covariates on the cause-specific hazard of the outcome or modelling the effect of covariates on the cumulative incidence function

(Austin, Lee, & Fine, 2016). The former allows one to estimate the effect of the covariates on the rate of occurrence of the outcome in those subjects who are currently event free. The latter allows one to estimate the effect of covariates on the absolute risk of the outcome over time. The former family of models may be better suited for addressing etiologic questions, whereas the latter model may be better suited for estimating a patient's clinical prognosis. Competing risks regression modelling requires that one consider the specific question of interest and subsequent choice of the best model to address it (Dignam, Zhang, & Kocherginsky, 2012).

In competing risks survival model, the main interest is in the time until a first event and also the type of the event that occurs, subjects are at risk of several different types of events during the follow-up period. Unlike censoring, which merely obstructs one from viewing the event, a competing event prevents the event of interest from occurring altogether, and analysis is adjusted accordingly (Dignam et al., 2012). In Cox regression, you focus on the survivor function, which indicates the probability of surviving beyond a given time. In competing-risks regression, you instead focus on the cumulative incidence function, which indicates the probability of the event of interest happening before a given time, the hazard function is specific to a particular event in competing risks (Kleinbaum & Klein, 2010).

Grover et al., (2014) Estimated the probability of death of AIDS patients in the presence of competing risks for complete data and in the absence of covariates. This work is a further extension of their work by considering censoring and covariates. Cumulative Incidence Function (CIF) has been used to estimate probability of death and cause specific hazard, subdistribution hazard and flexible parametric model have been applied to assess the covariate effects on CIF (Twabi & Mukaka, 2018). Austin et al., (2016) discussed likelihood inference to examine the effect of prognostic factors on the event of interest in the presence of competing

risk events. Dignam et al., (2012) described various probability models for summarizing competing risk data. Bakoyannis & Touloumi (2012) reviewed some basic statistical methods for analysing the competing risk data.

This chapter present survival models of patients on ART and TB treatment who developed TB after commencing ART in order to gain a better insight of the associated factors. The chapter further present Kaplan-Maier analysis for univariate analysis, Cox Proportional Hazard Model was used to model for risk factors associated with mortality of TB and HIV and lastly, the chapter gave insight Competing risk model as outcomes for patients on TB and ART are the possible competing risks events that can occurs.

3.2 Methods

3.2.1 Kaplan Meier curves

The interest in this study is time to event and in this case event is death. Data is censored in three categories of Alive if the subject managed to reach a cut-off point of 31st December 2017, Lost to follow or default up if the patient could not be traced. The type of censoring in this study is right censoring which means subjects ceased to be observed before their death occurs but each of them is at least observed for some time, thus some data is collected about each subject's survival. The Kaplan-Meier (K-M) method (Kaplan & Meier, 1958), is the most widely used non-parametric method of estimating the survival function *S(t)*, which is the probability that a subject survives longer than time *t*. This method utilizes information from both subjects who have experienced an event as well as right censored subjects. The Kaplan-Meier estimates at time *t* is given by:

$$S(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \tag{3.1}$$

where t_j is the time point where the survival is calculated and is within the period a subject is observed, n_j is the number of subjects at risk of death at time t_j and d_j is the number of deaths at time t_j .

The Kaplan-Meier curve is based on several assumptions namely:

- Sample is chosen randomly and independently from a larger population.
- Survival probabilities are the same for subjects entered early and late in the study.
- Time to censoring and survival times are independent.
- Deaths occurred at the times specified.
- Censored subjects have the same survival prospects as uncensored subjects.

The survivor function is normally presented as a Kaplan-Meier curve which is a plot of probability of survival function *S(t)* on the vertical axis against survival time *t* on the horizontal axis. Vertical drops indicate times at which an event (in this case, death) was observed. The survival probability at a particular time, median survival time, mean survival time and other quartiles are summary statistics that can be extracted from this survival curve (Kleinbaum & Klein, 2015). Sometimes we might be interested to ascertain whether the survival of subjects from one area (*e.g.* from Walvisbay District) is different from another for instance Swakopmund District. This type of comparison can be achieved visually by comparing the survival curves or through statistical tests. The log-rank test is the method used in this study in statistically comparing the overall difference between the survival curves (Kleinbaum & Klein, 2012).

The survivor function *S(t)* gives the probability that a person survives longer than some specified time *t*: that is, *S(t)* gives the probability that the random variable *T* exceeds the

specified time *t*. The hazard function *h(t)* gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time *t*. In contrast to the survivor function, which focuses on not failing, the hazard function focuses on failing, that is, on the event occurring. Thus, in some sense, the hazard function can be considered as giving the opposite side of the information given by the survivor function.

3.2.2 The Log-rank test

The log-rank test is a large-sample chi-square test that uses, as its test criterion a statistic that provides an overall comparison of the KM curves being compared. Kleinbaum & Klein (2015, p15) suggested that this (log-rank) statistic, like many other statistics used in other kinds of chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log-rank statistic are defined by each of the ordered failure times for the entire set of data being analysed. The log-rank statistic tests the null hypothesis that at all time points, the survival functions for all groups are equal, against the alternative hypothesis that at least one survival function is different from the others for some time periods (Goel, Khanna, & Kishore, 2010).

In other words, for *g* groups, the log-rank statistic tests:

H_0: There is no difference between survival curves

H_1: At least one survival curve is different from the others

The log-rank statistic is given by:

$$X^2 = \sum_{i}^{n} \frac{(O_i - E_i)}{E_i} \tag{3.2}$$

Where n is the number of groups, O_i is the observed number of deaths in each group and E_i is the expected number of deaths in each group i, assuming a null hypothesis of no difference in survival between the groups. O_i and E_i are calculated for each time when an event occurs. The log rank test is based on the same assumptions as the K-M given above, and under H_0, the log-rank statistic has n - 1 degrees of freedom, where n is the number of groups being compared. The decision whether to reject the null hypothesis is made using the chi-square tables with the appropriate degrees of freedom.

3.2.3 Multivariable survival model

The K-M curves and the log-rank test described above provide univariate analyses useful in assessing whether a covariate affects survival and are most suitable for descriptive purposes. They are particularly handy when the predictor variables are categorical and do not work easily with continuous predictors. However, they do not allow us to see how survival of a group is affected with the influence of other covariates included in the model. The study used Cox Proportional Hazard model to estimate effects of covariates used on survival which are were Health District, Facility Name, Residential address, age, sex, still on care, weight (kg), marital status, CD4 cells, function, WHO clinical stage, all measured at the start of the treatment. The Cox Proportional Hazard model (Cox, 1972) is commonly employed in analysing survival data in a multivariate way, allowing the effects of a set of covariates on survival time to be assessed. The Cox Proportional Hazard model also handles censored data, categorical and continuous variables as well as variables that change over time, all of which may influence survival (Hazra & Gogtay, 2017). The Cox Proportional Hazard model also allows for frailty to be included at various levels. The study also fitted the competing risk model, as there were outcome events that were competing with the event of interest.

3.3 Model Description

3.3.1 Cox Proportional Hazard model

The most common approach to model covariate effects on survival among other methods is the Cox regression model, which takes into account the effect of censored observations. The Cox model was introduced by (Cox, 1972). Let $x_1, x_2, ..., x_p$ be the values of p covariates X_1, X_2, ..., X_p. According to the Cox regression model, the hazard function is given as follows:

$$h(t, X) = h_0(\mathrm{t})\exp(\sum_{i=1}^{p} \beta_i x_i) \tag{3.3}$$

Where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is a *1xp* vector of regression coefficients and $h_0(t)$ is the baseline hazard function at that time. The model will be used to model mortality among TB and HIV co-infected patients in the study. Since the Cox regression model relies on the hazards being proportional, i.e. on the effect of a given covariate not changing over time, it is very important to verify that the covariates satisfy the assumption of proportionality (Cox, 2018).

The measure of the effect of the given covariates on survival time is given by the hazard ratio denoted as HR. Consider a categorical variable with two levels say $X = 1$ and $X = 0$, then the hazard ratio for the two groups is defined as:

$$HR = \frac{h(t|X = 1)}{h(t|X = 0)} = \exp(\beta) \tag{3.4}$$

When HR = 1, it implies that the individuals in the two categories are at the same risk of getting the event, when HR > 1, it implies that the individuals in the first category ($X = 1$) are at a high risk of getting the event and if HR < 1, the individuals in the second category ($X = 0$) are at a high risk of getting the event. The study will therefore utilize the global test to test if the hazard is proportional for all the covariates. If this assumption is violated, the Cox regression model

is invalid and more complicated analyses such as the Stratified Cox Regression model or the Extended Cox regression model will be required.

3.3.2 The Stratified Cox regression model

The stratified Cox regression model is a modification of the Cox regression model by stratification of covariate that does not satisfy the proportional hazards assumption (Kleinbaum & Klein, 2015). Covariates that are assumed to satisfy the proportional hazards assumption are included in the model, whereas the covariate stratified is not included (Therneau & Grambasch, 2014).

Let k covariates fail to satisfy the proportional hazards assumption, and p covariates satisfy proportional hazards assumption. The covariates not satisfying the proportional hazards assumption are denoted by $Z_1, Z_2, \ldots, Z_k$ and the covariates satisfying the proportional hazards assumption are denoted by $X_1, X_2, \ldots, X_p$. To form the stratified Cox regression model, a new variable is defined from z variables and denoted by z^*. The stratification variable z^* has k^* categories, where k^* is the total number of combinations (strata) formed after categorizing each of z's. There are interaction and no-interaction models defined in the concept of the stratified Cox regression model (Qin, Knol, Corpeleijn, & Stolk, 2010).

The models are defined by:

$$h_t(t,x) = h_{0g}(t)exp\left[\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p\right], g = 1,2,\ldots,k^* \; strata \qquad (3.5)$$

The stratified partial likelihood function is:

$$L = L_1 + L_2 + \ldots + L_{k^*} \qquad (3.6)$$

The above form means that you compute likelihood for each strata and multiply the likelihoods. There are two types of stratified cox models that can be built. One where there is no interaction,

i.e. β's in the model are the same for each subscript g (Kleinbaum & Klein, 2015). The no-interaction assumption means that the variables being stratified are assumed not to interact with *X's* in the model. In a model that allows interaction, you fit separate models for each strata.

For a non-interaction model, where the subscript g represents the strata and,

$$h_t(t,x) = h_{0g}(t)exp\left[\beta_{1g}x_1 + \beta_{2g}x_2 + \ldots + \beta_{pg}x_p\right], g = 1,2,\ldots,k \quad (3.7)$$

For an interaction model. The stratified Cox regression model contains regression coefficients that do not vary over the strata (Therneau & Grambasch, 2014). This property of the model is known as the no-interaction assumption. If interaction is allowed for, different coefficients for each of the strata are obtained. The test that is used to examine the no-interaction assumption is the likelihood ratio test statistic (Kleinbaum & Klein, 2015). For this test statistic, log likelihood functions of the interaction and no-interaction models are used. The interaction model differs from the no-interaction model by containing product terms. Thus, the null hypothesis is that the coefficients of the product terms are equal to zero. The likelihood ratio test statistic shows a Chi-square distribution with (k*−1) degrees of freedom under the null hypothesis (McHugh, 2013).

3.3.3 Competing risk survival models

Competing Risk models are applicable to situations when there are more than a single event types under consideration and each individual under study can undergo exactly one type of event. For example, there could be four types of events, each having a specific intensity rate $\alpha_i, i \in \{1,2,3,4,\}$. Estimating these intensities is not a problem but converting these intensities into probabilities is not a straightforward exercise. The estimation of cumulative hazard function of a particular event type i can be done easily by assuming the other events as censored

observations, the real problem is estimating individual survival curves based on event (Kleinbaum & Klein, 2015).

The outcomes for patients on TB and ART management are the possible competing risks events that occurs. The outcomes were modelled from the date they initiate ART. The competing risks for patients on TB and ART are the following:

1. Interrupted-Defaulted
2. Transfer-Out
3. Dead
4. Stopped

This study focuses on default and transfer out as the competing events. Similar to the Cox model, date of entry into follow-up in the competing risks model was date of ART registration. This study was censored on 31 December 2017. Subjects were also censored if they had died or stopped. Transfer-outs were censored in modelling the sub-hazard of default while default was censored in modelling the sub-hazards of default.

Suppose we have *n* observations consisting of four pieces of information each:

$$(t_i, d_i, j_i, x_i)$$

Where t_i is the observation time, d_i is a death indicator (1 if dead, 0 if censored), j_i is a cause of death index (takes a value between *1* and *m* for deaths and is undefined for censored cases), and x_i is a vector of covariates.

Suppose the j-th hazard function follows a proportional hazards model with Weibull baseline, say

$$\lambda_j(t, x) = \lambda_{j0}(t) e^{x'\beta}, \tag{3.8}$$

where the baseline hazard is

$$\lambda_{j0}(t) = \lambda_{jp_j}(\lambda_j t)^{p_j - 1}. \quad (3.9)$$

In view of the above results, we can estimate the parameters $(p_j,\ \lambda_j, \beta_j)$ using the techniques discussed before, simply by treating failures for causes other than j as censored cases.

Note that we have allowed all parameters to depend on the cause of death. We could, if we wanted, use different x's for each type of failure. If we wanted to restrict all the Weibulls to have the same index, for example, so $p_j = p\ \forall j$ then the overall likelihood function would not factor out and we would not be able to use this simplification. The same would be true if we wanted to force some β's to be equal across causes. In either case one would have to maximize the full likelihood.

3.3.4 Acceleration failure time model

Although parametric PH models are very applicable to analyse survival data, there are relatively few probability distributions for the survival time that can be used with these models (Qi, 2009). In these situations, the accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models we measure the direct effect of the explanatory variables on the survival time instead of hazard, as done in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. Currently, the AFT model is not commonly used for the analysis of clinical trial data, although it is fairly common in the field of manufacturing. Similar to the PH model, the AFT model describes the relationship between survival probabilities and a set of covariates (James, 2014).

Under an accelerated failure time model, the covariate effects are assumed to be constant and multiplicative on the time scale, that is, the covariate impacts on survival by a constant factor

(acceleration factor). According to the relationship of survival function and hazard function, the hazard function for an individual with covariate $X_1, X_2, \ldots, X_p$ is given by:

$$h(t|\mathrm{x}) = \left[\frac{1}{\eta(\mathrm{x})}\right] h_0 \left[\frac{n}{\eta(\mathrm{x})}\right] \tag{3.10}$$

The corresponding log-linear form of the AFT model with respect to time is given by

$$\log T_i = \mu + a_1 X_{1i} + a_2 X_{2i} + \ldots + a_p X_{pi} + \sigma\varepsilon_i \tag{3.11}$$

where μ is intercept, $\sigma\varepsilon_i$ is scale parameter and ε_i is a random variable, assumed to have a particular distribution. This form of the model is adopted by most software package for the AFT model. Furthermore, the widely used distributions of Weibull and Gompertz have hazards functions that are either monotonically increasing or decreasing (Qi, 2009).

Log-logistic

The uni-modal form of the hazard function is defined by a log-logistic. The log-logistic survival and hazard function are given by:

$$h(t) = \frac{e^{\theta} k t^{k-1}}{1 + e^{\theta} t^{k}} \tag{3.12}$$

where $0 \leq t < \infty$, $k > 0$

The hazard function above decreases monotonically if $k \leq 1$ but when $k > 1$ has a single mode (Qi, 2009). The corresponding survivorship function is defined by:

$$(t) = [1 + e^{\theta} t^{k}]^{-1} \tag{3.13}$$

Log-normal

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by:

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right) \tag{3.14}$$

where μ and σ are unknown parameters and $\Phi(.)$ is a standard normal distribution function, defined by:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \frac{-u^2}{2} du \tag{3.15}$$

3.3.5 Frailty survival model

Survival analysis models presented this far require that the survival times be independent. There are however situations when survival times are dependent. Dependant survival time data arise when the individuals have some feature in common. For instance, there is some characteristic that is shared by more than one observed event time. Account can be taken of effects of this characteristic by introducing a corresponding term into the model (Qi, 2009). The hazard function of a shared frailty survival model is defined by:

$$h_{ij}(t) = h_0(t) e^{(\beta x_{ij} + \zeta_i)} \tag{3.16}$$

where $h_0(t)$ is a baseline hazard; ζ_i is the frailty term and it is unobservable. $^{(\beta x_{ij} + \zeta_i)}$ is the covariate term. $i = 1, ..., n$ are group with $j = 1, ..., n_i$ observation in group i. A frailty survivorship function is defined by:

$$S_{ij}(t_{ij}) = e^{[(\beta x_{ij} + \zeta_i) H_0(t_{ij})]} \tag{3.17}$$

where $H_0(t_{ij})$ is the cumulative hazard function evaluated.

3.4 Measurement error

When covariates are measured with error, characterizing the association of mismeasured covariates and the underlying covariates is necessary for valid inference; different measurement error mechanisms require different adjustments of naive inference procedures to account for error effects. In this section, we introduce three widely used measurement error

models, distinguish two measurement error. For a comprehensive overview, we refer to (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006).

3.4.1 Measurement error models

In the covariate vector $Z_i(t) = (X_i^T, V_i^T(t))^T$ we let X_i represent time-dependent but error-prone covariates, suppose (t) be covariates that are precisely measured and possible time dependent. Suppose X_i is not observed, but its surrogated measurement W_i is collected.

Classic Additive Models

The classical additive measurement error model assumes that

$$W_i = X_i + \epsilon_i \tag{3.18}$$

where the ϵ_i are independent and identical distributed (i.i.d.) with mean 0 and a positive definite variance matrix Σ_0, and are independent of X_i. Often, a multivariate normal distribution is assumed for ϵ_i.

Berkson Model

The Berkson measurement error model assumes that

$$W_i = X_i + \epsilon_i \tag{3.19}$$

where the ϵ_i are independent and identical distributed (i.i.d.) with mean 0 and a positive definite variance matrix, and are independent of W_i. Often, a multivariate normal distribution is assumed for ϵ_i.

Multiplicative Model

The multiplicative measurement error model is given by

$$W_i = X_i \epsilon_i \tag{3.20}$$

where the ϵ_i are independent and identical distributed (i.i.d.) with mean 0 and a positive definite variance matrix Σ_0, and are independent of X_i.

The classical additive model (3.18) is perhaps the most popular error model, especially in modelling covariate measurement error in survival data. However, it is important to note that the choice of an error model is determined by the data at hand. See Carroll et al. (2006) for details. When discrete variables are subject to error, it is usually called a misclassification problem, and error modelling strategies are different from the three models introduced. Buonaccorsi (2010) provided a detailed treatment for misclassification problems.

Error mechanism

Two error mechanisms: nondifferential error mechanism and differential error mechanism are often distinguished in survival analysis with covariate measurement error. Nondifferential error mechanism occurs when W_i is independent of the underlying failure time T_i and censoring time C_i given the true covariates $(X_i^T, V_i^T(\cdot))^T$. Equivalently, the nondifferential error mechanism means that the distribution of T_i and C_i given X_i, W_i and $V_i(\cdot)$, does not depend on W_i:

$$f(t_i, c_i | x_i, w_i, v_i(\cdot)) = f(t_i, c_i | x_i, v_i(\cdot)) \tag{3.21}$$

where t_i, c_i, w_i and $v_i(\cdot)$ are realized values of T_i and C_i, X_i, W_i and $V_i(\cdot)$, respectively. Thus the surrogate W_i is noninformative in that it does not contribute information about T_i and C_i, X_i were known. Measurement error if differential if (1.5) is not true. In this thesis, the study assumes the nondifferential error mechanism, as consistent with the treatment done by most authors.

3.5 Data analysis

There are three stages of analysis considered in this study. The first involves preliminary descriptive analysis relating to variables to be considered using Kaplan Meier test. The second stage involves investigation of spatial correlation structure of mortality. The last stage of analysis involves fitting models to determine how particular independent variable contributes to explaining the dependent variable, as well as accounting for frailty effects and mapping the risk of mortality. Multivariate analysis is employed by fitting Cox Proportion Hazard Model to estimate survival trends of subjects as well as to estimate the effects of covariates on survival and a Bayesian model is also fitted to show spatial and temporal in mortality hazard by EAs in the region. Analysis were done using STATA v14.2, R v3.5.0 and BayesX v3.0.2.

3.6 Results

3.6.1 Descriptive statistics of selected variables

Table 3 present descriptive statistics disaggregated by patient's survival status of patients, the corresponding p-values for log-rank test and the Cox univariate analysis for variable age, CD4 and weight. The extracted and final dataset utilized in this study comprised of information of 3145 subjects of whom 464 (14.8%) had experienced an outcome event (died) of interest during the study period. The median age of the subjects considered in this study was similar in most of the category with Default and Transferred out all had 41 years while Alive, Dead and Stopped had 40, 44 and 38 years respectively. The overall median age for all the patient considered in the study was 41 years. The overall median weight for the study subjects was 59.0kg, with those that were alive at the end of the study period had 60.0kg, those ones that experienced an event (dead) had 58.8kg, the one that defaulted treatment had 57.3kg, those one that got stopped medication due to other medical condition had 61.5kg and those that got transferred to other facilities had 58.8kg.

Table 3: Descriptive summary of explanatory numeric variables by outcome status considered in this study and log rank test

Outcomes Categories	Variables		
	Age	**CD4**	**Weight**
Alive (1673)	40 years	179.0 cells/μL	61.8 kg
Dead (464)	44 years	143.5 cells/μL	58.8kg
Defaulters (424)	41 years	153.5 cells/μL	57.3 kg
Stopped (42)	38 years	187.0 cells/μL	61.5 kg
Transferred (542)	41 years	155.0 cells/μL	58.8kg
Total (3145)	**41 years**	**167 cells/μL**	**60.0 kg**
P-Value	**p<0.001**	**p<0.001**	**p<0.001**

The dataset is skewed towards males (53.2%), who make up more than half of the subjects in the study and female only making up 46.8% Table 4 illustrates this. A high percentage of the patient were enrolled on TB/HIV treatment in Erongo Region came from the Walvisbay district with 41.1%, followed by Swakopmund district with 34.8%, followed by Usakos district with 12.4% and the last being Omaruru district with 11.8%. The analysis further shows that most of the patients were enrolled on treatment through district hospital with 74.1% of them being enrolled through hospitals, 18.0% of the patients were enrolled at health centres with only 7.9% who were enrolled at clinics in the region.

The analysis in Table 4 further shows that most of the patient (87.1%) are single, 7.4% and 3.6% were married and divorced respectively, while only about 2.0% were widowed at the time of enrolment in the TB and ART programme in the Erongo Region. Another variable of interest which is captured during admission of subjects into the TB/HIV program is the patients CD4 Category at initiation, of those enrolled and had their CD4 cell measured, 73.3% had CD4 cell count of less than 250 cells/μL while the other 26.7% had CD4 count more than 250 cells/μL. The data had numbers of missing values of CD4 cell count of 262 (8.3%) subjects.

Table 4: Descriptive summary of explanatory categorical variables by outcome status considered in this study and log rank test

Variables	**Categories**	**Outcome Categories**					Total (3145)	**P-Value**
		Alive (1673)	Dead (464)	Defaulters (424)	Stopped (42)	Transferred Out (542)		
Sex	Female	783(46.8)	191(41.2)	197(46.5)	17(40.5)	245(45.2)	1433(45.6)	0.09
	Male	890(53.2)	273(58.8)	227(53.5)	25(59.3)	297(54.8)	1712(54.4)	
Marital Status	Divorced	60(3.60)	6(1.30)	8(1.90)	2(4.80)	9(1.70)	85(2.70)	0.30
	Married	123(7.40)	27(5.80)	18(4.30)	1(2.40)	44(8.10)	231(6.80)	
	Single	1457(87.1)	426(91.8)	39(92.9)	39(92.9)	478(88.2)	2793(88.8)	
	Widowed	33(2.0)	5(1.10)	5(1.20)	0(0.0)	11(2.0)	54(1.70)	
Facility Level	Clinic	162 (9.70)	13(2.80)	34(8.0)	4(9.50)	34(6.30)	247(7.90)	p<0.001
	Health Centre	361(21.6)	35(7.50)	65(15.3)	4(9.50)	102(18.8)	567(18.0)	
	Hospital	1150(68.7)	416(89.7)	325(76.7)	34(81.0)	406(74.9)	2331(74.1)	
District Name	Omaruru	98(5.90)	89(19.2)	107(25.2)	3(7.10)	73(13.5)	370(11.8)	p<0.001
	Swakopmund	672(40.2)	142(30.6)	95(22.4)	27(64.3)	158(29.2)	370(11.8)	
	Usakos	143(8.60)	100(21.6)	43(10.1)	3(7.10)	100(18.5)	389(12.4)	
	Walvisbay	760(45.4)	133(28.7)	179(42.2)	9(21.4)	211(38.9)	1292(41.1)	
WHO Stage	T1	856(51.2)	154(33.2)	195(46.0)	23(54.8)	239(44.1)	1467(46.7)	p<0.001
	T2	190(11.4)	67(14.4)	42(9.90)	4(9.50)	55(10.2)	358(11.4)	
	T3	494(29.5)	187(40.3)	124(29.3)	11(26.2)	197(36.4)	1013(32.2)	
	T4	133(8.0)	56(12.1)	63(14.9)	4(9.50)	51(9.40)	307(9.80)	
Function	Ambulatory	18(1.10)	90(19.6)	48(11.4)	3(7.10)	58(10.8)	217(6.90)	p<0.001
	Bedridden	3(0.20)	41(8.90)	10(2.40)	0(0.0)	2(0.40)	56(1.80)	
	Working	1647(98.7)	328(71.5)	364(86.3)	39(92.9)	478(88.9)	2856(91.30)	
Age Group	0-14 Years	33(2.0)	6(1.30)	9(2.10)	0(0.0)	12(2.20)	60(1.90)	p<0.001
	15-49 Years	1388(82.9)	321(69.2)	325(76.7)	38(90.5)	416(76.8)	2488(79.2)	
	50+ Years	252(15.1)	137(29.5)	90(21.2)	4(9.50)	114(21.0)	597(19.0)	
CD4 Category	< 250 cells/μL	1067(69.0)	356(82.0)	292(76.8)	24(75.0)	116(23.6)	2114(73.3)	0.001
	> 250 cells/μL	479(31.0)	78(23.2)	88(23.2)	8(25.0)	116(23.6)	769(26.7)	

The majority of the patients seemed to have come to treatment when they were not showing any symptom condition i.e. 51.2% of the subjects came while their medical condition had not progressed to any WHO clinical stage, hence their condition was classified under stage 1, with the subject only show asymptotic symptoms. 11.4% and 32.2% came while in stage 2 and stage 3 respectively while only about 9.8% came for medical attention while the have already progressed to stage 4 which is classified as severe medical condition. The median CD4 cell count for subjects who were alive was 179.0 cells/μL, median CD4 cell count for those who died was 143.5 cells/μL, median CD4 cell count for those who defaulted was 153.5 cells/μL and median CD4 cell count for those who were stopped and transferred out was 187.0 and 155.5 cells/μL respectively. The overall median CD4 cell count was 167.0 cells/μL.

It is not feasible to calculate a Kaplan-Meier curve for the continuous predictors since there would be a curve for each level of the predictor and a continuous predictor simply has too many different levels (Allison, 2014). Instead, the study considered the Cox proportional hazard model with a single continuous predictor. Unfortunately, it is not possibly to produce a plot when using the **stcox** command in STATA (StataCorp, 2013). Instead the study considered the Chi-squared test for age, CD4 count and weight, thus weight a potential candidate for the final model since the p-value < 0.001 is less than our cut-off value of 0.20 – 0.25.

3.6.2 Risk factors for death and survival among patients

A measure of whether the survival patterns are significantly different within a variable level is shown in terms of p-value generated from a log-rank test in Table 3 and 4. The table shows that there are significant differences in the survival times of subjects for in most of the covariates such as facility level, district name, function, WHO stage, CD4 category and age group while covariates such as sex and marital status, were not significant at 5% significant

level. In summary, Figure 4 is a Kaplan-Meier curve illustrating the survival trend for the entire region. The curve shows that mortality hazard is high in the first 8 years, such that by 8 years, the proportion surviving had reduced to 0.76 which had only decreased to 0.73 by 10 years and finally decreased to 0.70 by the end of the study period.

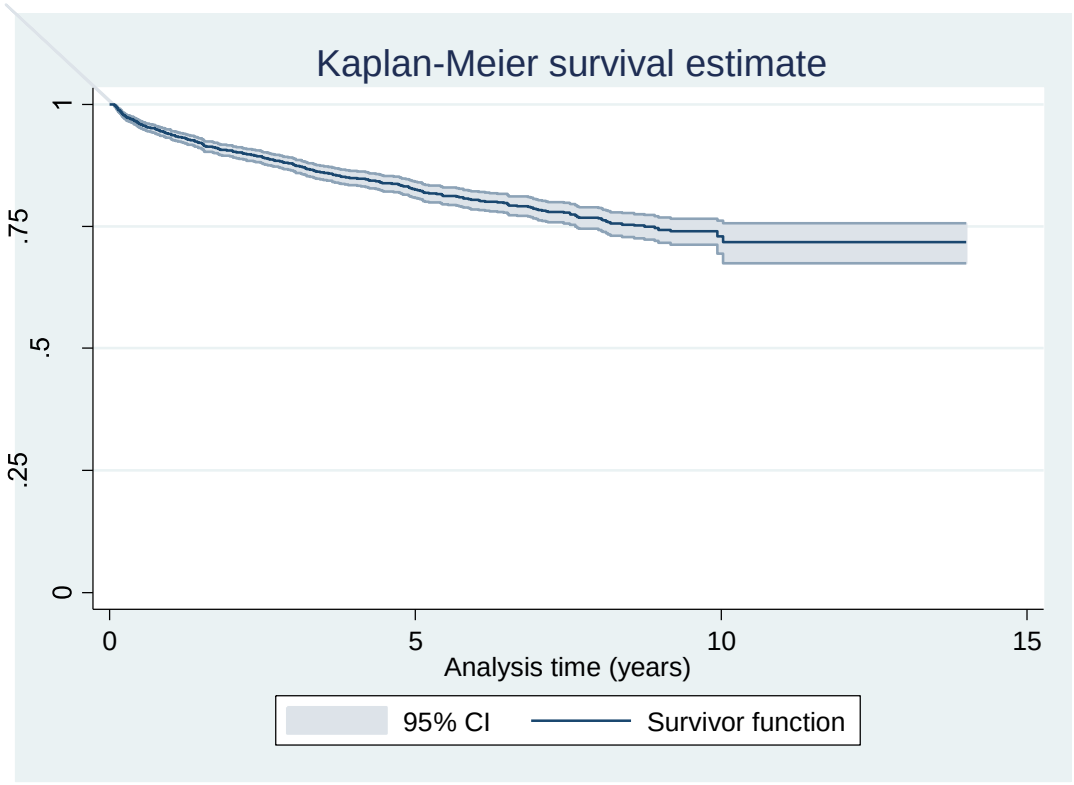


Figure 4: Kaplan-Meier survival curve for Erongo Region

Figure 5, is a Kaplan Meier curve describing survival time distribution by sex. K-M graph on sex shows that females and male survivor probability were approximately proportional. Even though female patients seem to have a slightly better survival experience than their male counterpart for most of the study period. Overall, the survival probability appears very similar across the two categories of sex. This is consistent with the log-rank test results that shows that there is no difference in the two survival the categories (χ^2 = 3.34, df = 1, p-value = 0.07). The probability of patient surviving beyond 10 years was the same (0.70) across the two category.

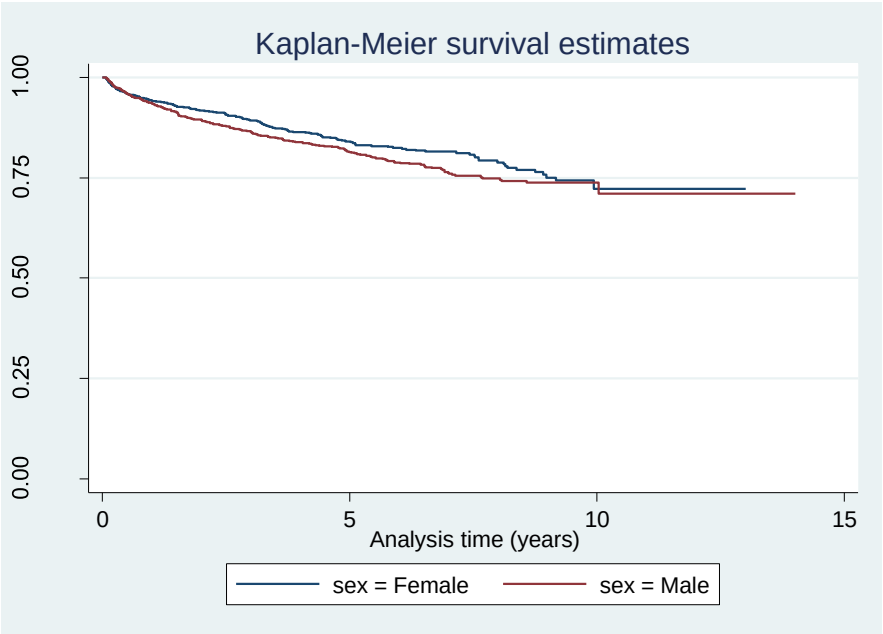


Figure 5: Kaplan-Meier survival curve for sex

Age group was significant in the log-rank test of equality across category. Therefore, there was enough evidence to support the statement that, at least one survival curve is different from the others. The analysis shows that there were significant differences among patients with respect to their age group category. There were differences in survival among younger patient relative to older patient. Age group category with (X^2 = 28.7, df = 2, p-value < 0.001) was therefore included in the model. Figure 6, is a Kaplan Meier curve describing survival time distribution by age group. Across the levels of age group, the survivor experiences of patient were proportional, the survival experience was similar at the beginning of the study across the age group category of 0-14 years and 50+ years. There were also few deaths recorded among the age group of 15-49 Years during the first 3 years. The survival probability of 15-49 years was higher after 3 years onwards comparing to the 0-14 years old and 50+ years had the lowest probability of survival.

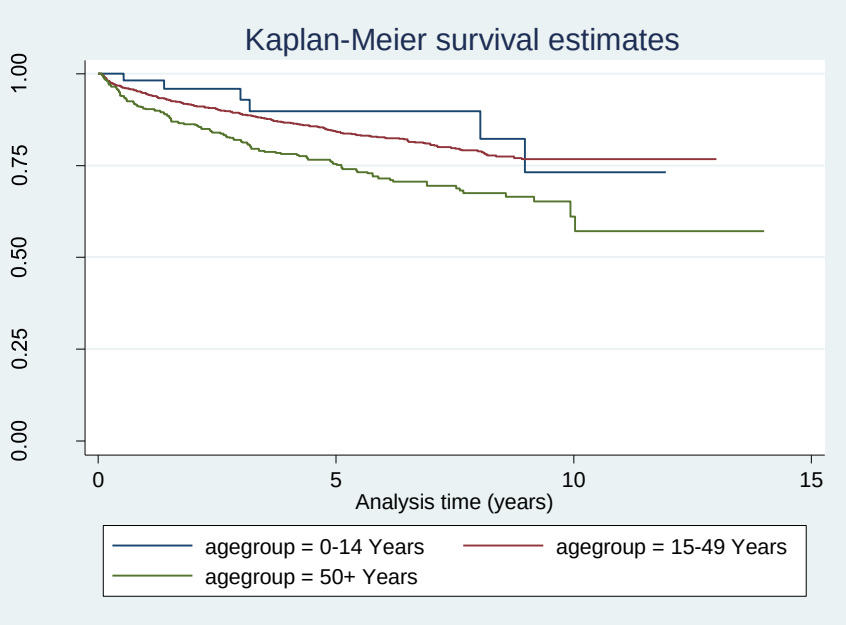


Figure 6: Kaplan-Meier survival curve for age group

The results of the log-rank test of equality across strata shows that there were no differences in survival across the categories of marital status, ($\chi^2 = 3.64$, df = 3, p-value = 0.30). Marital status was not significant with a p-value of 0.30, *refer back to Table 4*. The survival chances among patient who are single, married, divorced and widowed was similar. Marital was therefore excluded in the fitted models. Figure7, is a Kaplan Meier curve describing survival time distribution by marital status. The survival curve clearly supports the log-rank test results that the curves for the four categories are identical statistically.

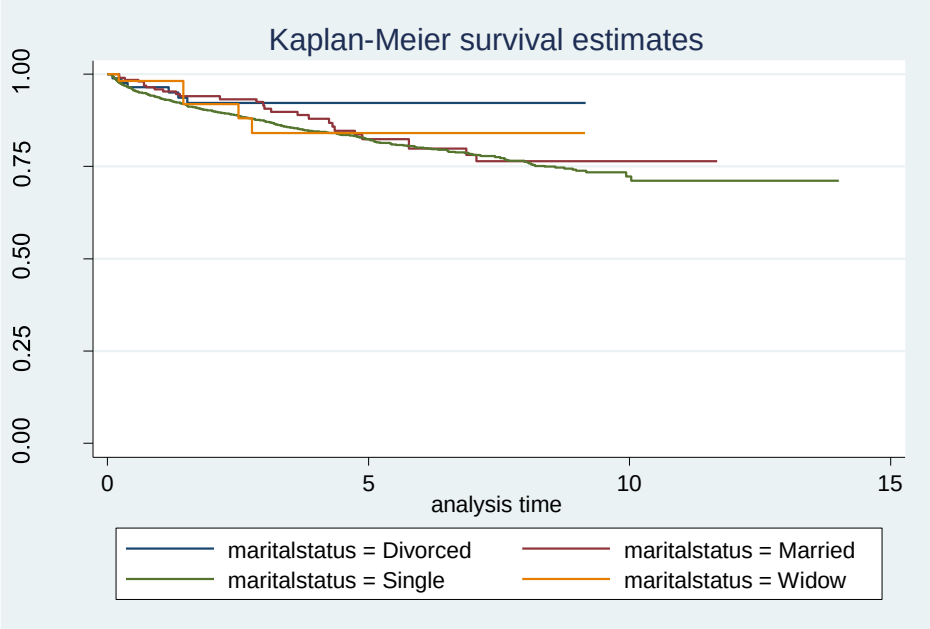


Figure 7: Kaplan-Meier survival curve for marital status

Facility level of service provision was significant in the log-rank test of equality across strata ($\chi^2 =$ 20.78, df = 2, p-value < 0.001). There were therefore significant differences among survival of TB and HIV co-infected patient in different categories of health service provision. Patients accessing treatment through clinics, health centres and hospitals had different chances rates of surviving. Facility level of accessing treatment with a p-value of less than 0.001 was therefore included in the model. Figure 8, is a Kaplan Meier curve describing survival time distribution by level of service provision. Across the level of service provision, the survivor chance of TB and HIV patients was proportional. Similar to most of the predictors, the survival probability was similar in the beginning after initiation in the health centres and clinics. Patients accessing treatment through health centres had a higher survival probability comparing to those that are accessing services at hospitals and health centres.

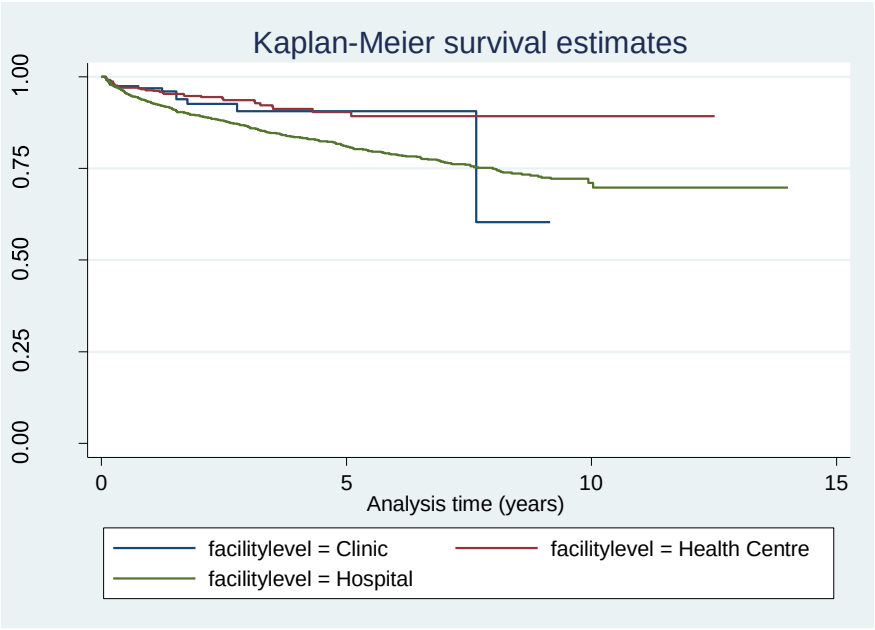


Figure 8: Kaplan-Meier survival curve for facility level

World Health Organization's (WHO) staging which is based on subjects medical condition also shows that there is a significantly difference in survival pattern for subjects in each of the four stages (χ^2 = 26.35, df = 3, p-value < 0.001). The probability of survival for the different WHO stages worsen with increased stage. Stage T1 is asymptomatic, T2 is mild, T3 is advanced and T4 severe medical condition. TB and HIV patients on WHO clinical stage 4 seems to have the lowest survival probability. The Kaplan-Meier curve Figure 9 shows this pattern. The figure further indicates that stage T1 as expected had the highest survival probability comparing to the other three WHO stages who have the same survival pattern.

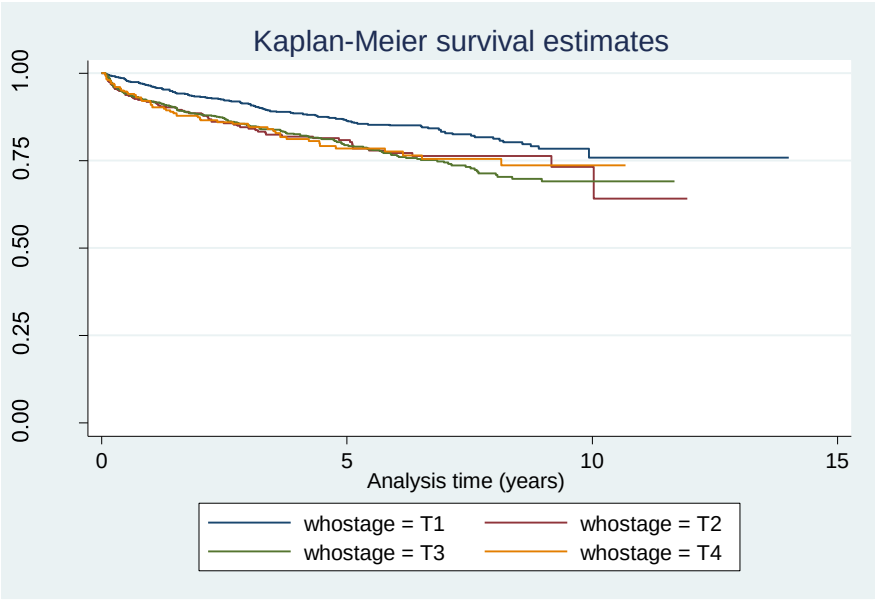


Figure 9: Kaplan-Meier survival curve for WHO clinical staging

The analysis from Table 4, indicated that at least one of the three survival curve of the functional status is different from the other ones ($\chi^2 = 599.57$, df = 2, p-value < 0.001). Figure 10, illustrates the survival patterns of the three category that are, working, ambulatory and bedridden TB and HIV co-infected patients. Patients that are categorised under working category had a high probability of survival comparing to the one that are ambulatory and bedridden. As expected patients that are classified as bedridden had the lowest survival probability. The results from the log-rank test are literally supporting the survival curve in Figure 10.

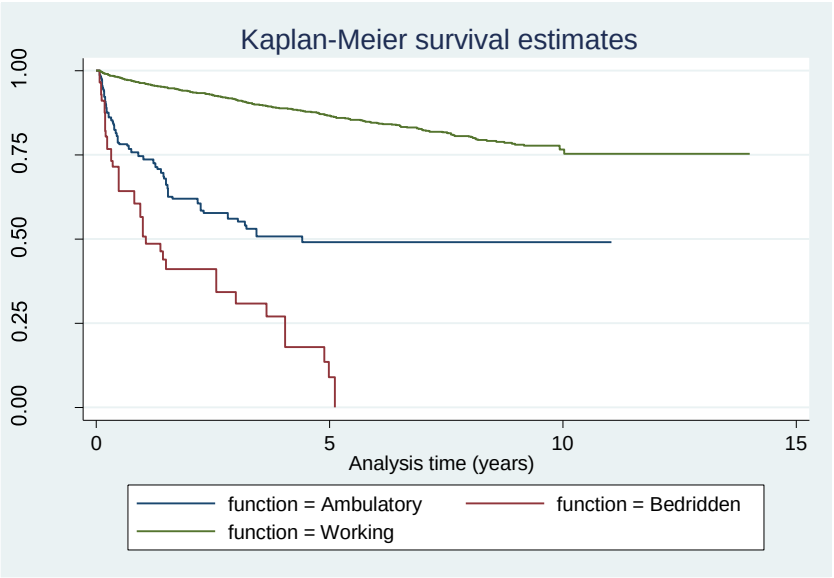


Figure 10: Kaplan-Meier survival curve for functional status of patient

Survival pattern for CD4 cell count category shows that the considered categories i.e. CD4 cell count of less than 200c/μL and cd4 cell count of at least 200c/μL highly significantly (χ^2 = 12.1, df = 1, p-value < 0.001), the variable CD4 category was included the multivariate model since its significant. The Kaplan-Meier curve Figure 11 for CD4 category shows that subjects with CD4 cell count of less than 200c/μL have different survival probability with those with at least 200c/μL. Those with CD4 cell count more than 250c/μL have higher probability of surviving comparing to those with CD4 cell less that 250c/μL.

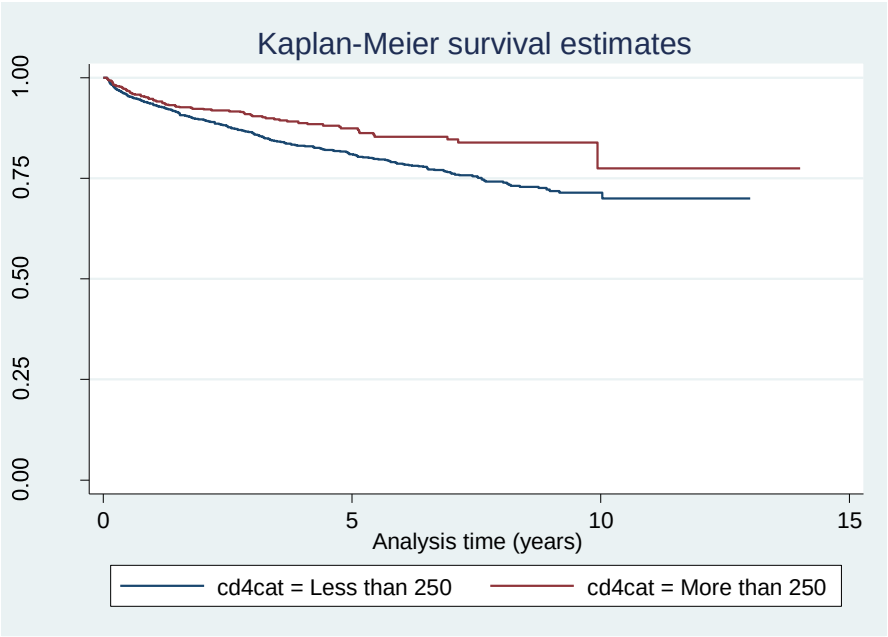


Figure 11: Kaplan-Meier survival curve for CD4 category

The health district of service provision was significant in the log-rank test of equality across the four health district. There were therefore significant differences among TB and HIV Co-infected patient different district of health service provision ($\chi^2 = 134.42$, df = 3, p-value < 0.001). Health district level of accessing treatment with a p-value<0.001 was therefore included in the model. Figure 12, is a Kaplan Meier curve describing survival time distribution by level of service provision. The K-M graph on district of service provision treatment was accessed. Across the level of service provision, the survivor probability of was proportional. Even though Swakopmund and Walvisbay district recorded the highest event during the duration of the study, the two district had the highest survival probability compared to Usakos and Omaruru district who have the lowest chances of surviving.

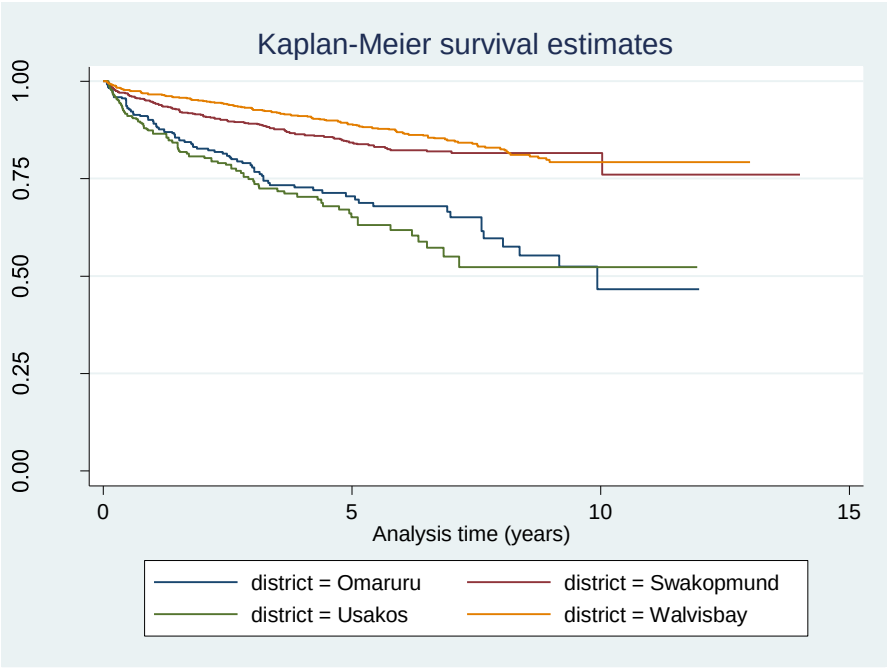


Figure 12: Kaplan-Meier survival curve for health district name

3.6.3 Cox Proportional Hazards model

The study fitted different Cox Proportional Hazard Models to determine independent variables that are assisting in explaining the outcome event of the study. Models were compared using Likelihood Ratio test and the variables added turned out to be improving the model fit. The following was the fitted model in multivariable form:

$$Y_{i1}(status) = \alpha + \beta_{i2}(weight) + \beta_{i3}(sex) + \beta_{i4}(agegroup) + \beta_{i5}(facilitylevel) + \beta_{i6}(district) + \beta_{i7}(whostage) + \beta_{i8}(function) + \beta_{i9}(cd4cat)$$

Where, $i1 = 1, \ldots, 5$ $i2 = 1$ $i3 = 1, 2$ $i4 = 1, 2, 3$, $i5 = 1, 2, 3$ $i6 = 1, 2, 3, 4$ $i6 = 1, 2, 3, 4$ $i7 = 1, 2, 3$ $i8 = 1, 2$

The model above was adopted for analysis in this study with LR chi-square (15) = 353.01, prob > χ^2 < 0.001

3.6.3.1 Checking the Proportional Hazard Assumption

The study has used Schoenfeld and scaled Schoenfeld residuals in checking the adherence to proportional hazard assumption. Individual variables are subjected to the test and finally, a global test examine the whole model. The proportional hazard assumption was tested at 5% level of significance. Thus any result with p-value of less than 0.05 means the proportionality assumption is violated. Table 5 below summarizes the results of the test of proportional hazard assumption, all independent variables except Function did not violate the proportional hazard assumption (p-value= 0.01).

Table 5: Test of Proportional Hazard assessment of assumption

	Rho	χ^2	df	Prob>χ^2
Weight	-0.07	2.17	1	0.14
Sex	-0.02	0.24	1	0.63
Age Group	0.05	0.97	1	0.33
Facility Level	0.04	0.54	1	0.46
District	0.03	0.34	1	0.56
WHO Stage	-0.07	2.05	1	0.15
Function	0.12	6.23	1	0.01
CD4 Category	-0.05	1.10	1	0.30
Global Test		13.34	8	0.10

The covariate function needed a closer examination and graphical assessment was used to closely look at the variable. Function illustrates a violation of proportional hazard assumption since it is not parallel to the zero line at inception, refer to Figure 13 below.

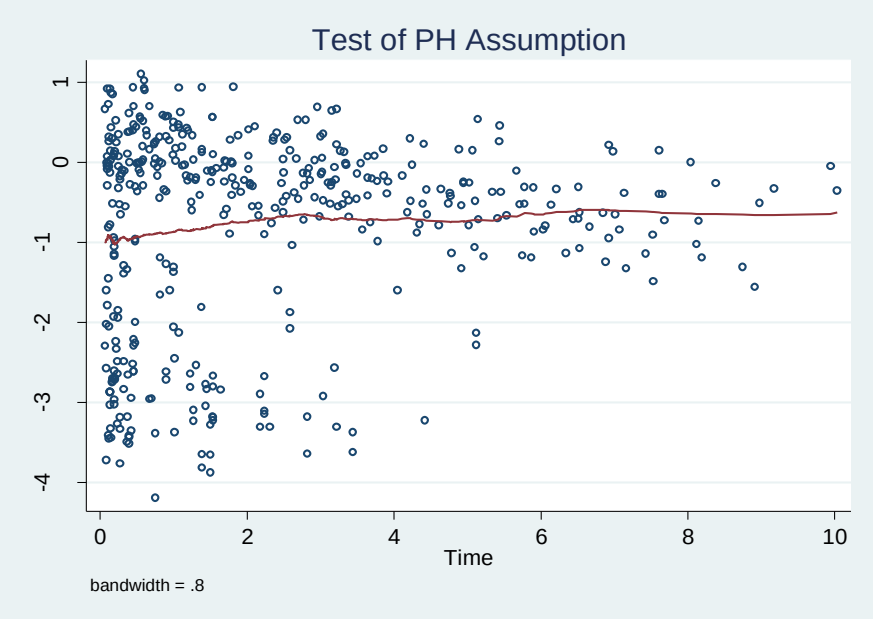


Figure 13: Test of Proportional Hazard assumptions

3.6.3.2 Goodness of the model fit

A graph of Nelson-Aalen cumulative hazard function and the Cox Snell variable the hazard function is plotted to check model fit. Refer to Figure 14 below. It is noted from the graph above that the hazard function follows the 45° line very closely except for very large values of time. The deviations at the end of the time are due to a large volume of censored data and this is causing some wiggling at large values of time. If the hazard function follows the 45° line then it's know that it approximately has an exponential distribution with a hazard rate of one and that the model fits the data well (Therneau & Grambsch, 2013).

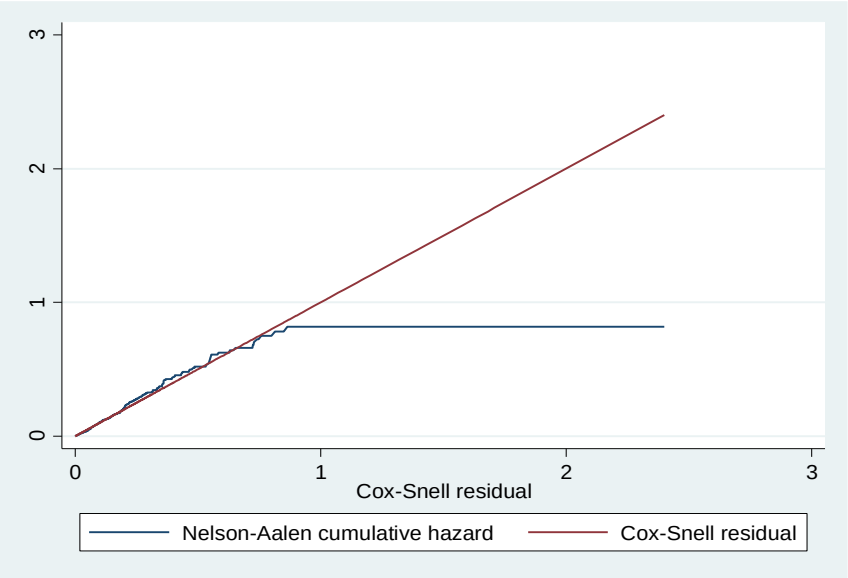


Figure 14: Cox-Snell Residual for Cox PH model goodness of fit

3.6.4 Stratified Cox Proportional Hazard model

The violation of proportional Hazard by Function required fitting a Cox regression model stratified by function. Table 6, below is an output of estimates of Stratified Cox regression model in multivariate level. During follow-up, mortality was significantly higher in children (0-14 years) by a factor of 1.10 (p-value = 0.83, 95% CI = 0.47, 2.58) as compared to the young adults (15-49 years), this means that children (0-14 years) category where 9.6% more likely to experience death compared to young adults (15-49) years. For subjects who are Adults (50+ years) there is no evidence that their mortality hazard is different from that of young adults (15-49 years), (p-value = 0.96, 95% CI= 0.47, 2.58). The HR is slight lower for male patients compare to the female counterparts, HR = 0.98 (p-value=0.85, 95% CI = 0.80, 1.20).

The hazard ratios of mortality of subjects living in Usakos district are significantly higher by 46.8% as compared to that are living in Omaruru district (p-value=0.02, 95 % CI=1.08, 2.01). Those once that are living in Swakopmund and Walvisbay district mortality is lower by 41.5% and 26.8% respectively hazard ration of 0.59 for Swakopmund (p-value<0.001, 95% CI=0.44,

0.78) and 0.73 for Walvisbay (p-value=0.04, 95% CI=0.54, 0.99). The hazard ratio of patients in health centre relative to patients in clinics was not significantly different, p-value of 0.27. Patients in health centre were 32.4% less likely to die comparing to patients in clinics. Patients in health centre had a hazard ratio of HR = 0.68 (95% CI = 0.34, 1.36). TB and HIV patients in hospitals was also not significantly different with relative to those from clinics, p-value = 0.40. Patients in hospital were 23.2% less likely to experience an event (dead) compare to patients from clinic, patients from hospitals had a HR = 0.77, 95 % CI = (0.42, 1.42).

The hazard ratios increased progressively with increasing World Health Organization's immunological staging (WHOa, 2016). In this study, the results stated otherwise, the HR decreases with each increasing WHO immunological staging. The hazard ratios for subjects who started the treatment while in stage 2 is 1.07 (p-value=0.66, 95% CI=0.79, 1.46), decreases even more for subjects who started treatment while in stage 3 by a factor of 1.06 (p-value=0.65, 95% CI=1.93, 3.29) and further decreasing for subjects who had already progressed to stage 4 by a factor of 0.87 (p-value = 0.41, 95% CI = 0.62, 1.22) as compared to subjects who started the treatment while in stage 1, refer to Table 6 below. The hazard is also slightly higher for subjects who started the treatment with CD4 cell count of at least 250 cells/μL, by a factor of 6.19% (p-value= 0.65, 95% CI=0.82, 1.40) as compared to those who started the treatment with CD4 cell count of less than 250cells/μL.

Table 6: Estimates of Stratified Cox Regression

Variables				
	Haz. Ratio	Std. Error	p-value	95% Conf. Interval
Weight	0.99	0.04	0.98	0.99, 1.01
Sex				
Female	Ref			
Male	0.98	0.10	0.85	0.80, 1.20
Age group				
0-14 years	Ref			
15-49 years	1.10	0.48	0.83	0.47, 2.58
50+ years	0.98	0.43	0.96	0.41, 2.33
Facility level				
Clinic	Ref			
Health centre	0.68	0.24	0.27	0.34, 1.36
Hospital	0.77	0.24	0.40	0.42, 1.42
District				
Omaruru	Ref			
Swakopmund	0.59	0.09	p<0.001	0.44, 0.78
Usakos	1.47	0.23	0.02	1.08, 2.01
Walvis bay	0.73	0.11	0.04	0.54, 0.99
WHO stage				
Stage 1	Ref			
Stage 2	1.07	0.17	0.66	0.79, 1.46
Stage 3	1.06	0.12	0.65	0.84, 1.33
Stage 4	0.87	0.15	0.41	0.62, 1.22
Function				
Ambulatory	Ref			
Bedridden	1	omitted		
Working	1	omitted		
CD4 category				
Less than 250 cells/μL	Ref			
More than 250 cells/μL	1.06	0.14	0.65	0.82, 1.38

Note: Ref is the reference category

3.6.5 Competing risks model

Cumulative incidence are used to describe the survival experience when there are competing events, Kaplan-Meier estimate ignores events of all types other than the one of interest and the probability is calculated whenever the event of interest occurs (StataCorp, 2013). The Cumulative Incidence Function (CIF) estimator for a particular event depends not only on the number of individuals who have experienced the event of interest, but also on the number of individuals who have not experienced any other type of event. The CIF represents the probability that an individual will experience an event of the particular type given that the individual has not experienced any event.

In the fitted competing risks model, the main events of interest were default from treatment and transfer out. Figure 15, presents a plot of cumulative incidence of defaulting and transferring out.

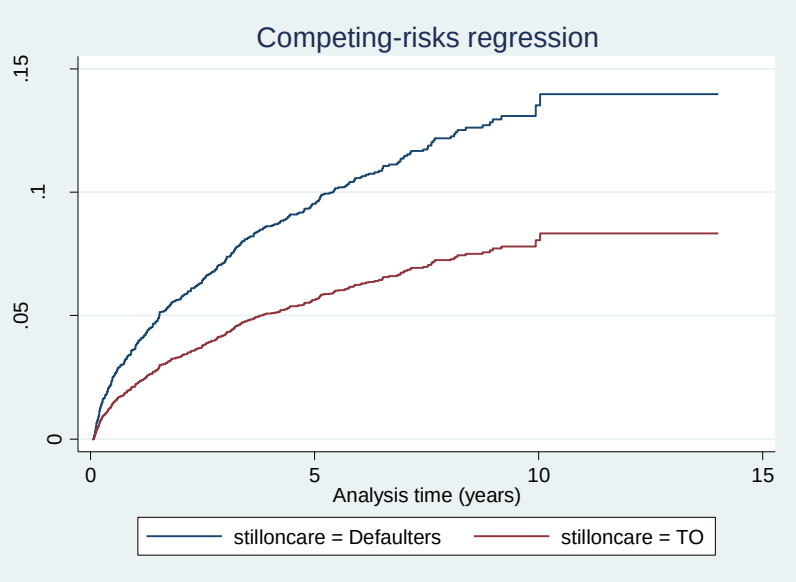


Figure 15: Cumulative Incidence of TO and default

The estimated probability that an individual will experience the event of default is higher over the whole of those who transferred out. There incidences of both events started on the rise at

the beginning of the study. The chance of a TB and HIV patient experiencing either event was therefore very small. The probability of defaulting increases steeply from 5 years then the increased until from around 10 years then it remained constant. The probability of transfer outs, however, remains relatively low over the entire follow-up period comparing to the defaulters. A competing risks model was then fitted. This was to further study the distribution of hazards among TB and HIV patients in the context of competing events. Table 6, presents the fitted model of which the results are discussed.

3.6.5.1 Defaulters and transferred out

In the fitted competing risks model of defaulters, the estimated sub-hazard ratios take into account the possibility that transfer out can also occur and vice versa. The sub-hazard ratios of categories of Swakopmund, Usakos and Walvisbay district, WHO stage 3, bedridden, working and CD4 category of more than 250 cells/μL were significantly associated with the risk of patient defaulting from treatment and being transferred out to other facility. Only, Usakos district is a risk factor a risk factor in defaulting and not a risk factor with respect to transfer out. Male patients were 9% more likely to default treatment compare to female patients (SHR = 1.09, p-value = 0.38, 95 % CI = 0.89, 1.34). The sub-hazards of default were similar to the cause specific hazards. Similarity is observed in most of the categories, as mentioned above Usakos district is a risk factor in defaulter and not a risk factor in transferred out with a p-value of 0.02 and 0.56 respectively. The risk is 50% more with respect to Omaruru in defaulters (SHR = 1.50, 95 % CI = 1.08, 2.07) and merely 11% more between the two categories in transferred out (SHR = 1.11, 95% CI = 0.80, 1.54). However, male, all the age group categories, health centre, hospital, WHO stage 2 and stage 4 were not found to be a risk factor in defaulting and being transferred out.

The tests of statistical significance are largely in agreement in the cause-specific and sub-hazards of default and transferred out in most of the categories. This similarity is likely due to low incidence of transfer-out. Figure 12 above show that the incidence of transfer-out remained relatively low over the entire follow-up period. The estimated sub-hazard ratios of default were affected but low incidence. This considering that the estimates take into account the possibility that transfer-out can also occur.

Table 7: Competing risks model

	Defaulters			Transferred out		
Predictors	SRH	p-value	95% CI	SRH	p-value	95% CI
Weight	0.99	0.02	0.983, 0.998	0.99	0.02	0.98, 1.00
Sex						
Female	Ref					
Male	1.10	0.38	0.89, 1.34	1.06	0.56	0.87, 1.30
Age group						
0-14 years	Ref					
15-49 years	1.60	0.29	0.67, 3.82	1.56	0.31	0.67, 3.66
50+ years	2.30	0.06	0.95, 5.57	2.233	0.07	0.94, 5.31
Facility level						
Clinic	Ref					
Health centre	0.95	0.88	0.46, 1.94	0.86	0.70	0.42, 1.80
Hospital	1.68	0.10	0.90, 3.13	1.64	0.13	0.87, 3.09
District						
Omaruru	Ref					
Swakopmund	0.65	0.01	0.47, 0.88	0.55	p<0.001	0.41, 0.75
Usakos	1.50	0.02	1.08, 2.07	1.11	0.55	0.80, 1.54
Walvis bay	0.65	0.01	0.47, 0.89	0.56	p<0.001	0.41, 0.76
WHO stage						
T1	Ref					
T2	1.32	0.08	0.97, 1.81	1.30	0.11	0.94, 1.78
T3	1.45	0.02	1.15, 1.82	1.41	0.03	1.12, 1.77
T4	1.26	0.22	0.87, 1.81	1.35	0.11	0.94, 1.93
Function						
Ambulatory	Ref					
Bedridden	2.42	p<0.001	1.51, 3.87	3.32	p<0.001	2.07, 5.32
Working	0.28	p<0.001	0.21, 0.37	0.31	p<0.001	0.23, 0.41
CD4 category						
Less than 250 cells/μL	Ref					
More than 250 cells/μL	0.64	p<0.001	0.50, 0.83	0.65	p<0.001	0.51, 0.85

Note: Ref is the reference category

3.6.5.2 Competing Risk Model Test of PH Assumptions

Cox proportional hazard assumptions were then assessed in the competing risks model. Similar to Cox proportion hazards model, proportionality was checked by including time-varying covariates in the model. Time dependent covariates are interactions of the predictors and time, if a time-dependent covariate is significant, it is an indication a violation of the proportionality assumption for that specific predictor (StataCorp, 2013). Table 8 presents the competing risks model of defaulter and transferred out with time-dependent covariates.

Table 8: Testing of assumption of competing risks model

Predictors	Defaulter			Transferred out		
	Coef.	p-value	95 % CI	Coef.	p-value	95 % CI
Weight	-0.02	0.19	0.99, 1.01	-0.02	0.16	0.99, 1.01
Sex	-0.03	0.52	0.89, 1.06	-0.04	0.42	0.88, 1.05
Age group	0.03	0.54	0.98, 1.13	0.04	0.40	0.95, 1.15
Facility level	0.12	0.20	0.94, 1.37	0.14	0.17	0.94, 1.40
District	0.02	0.40	0.98, 1.06	0.02	0.49	0.97, 1.06
WHO stage	-0.03	0.12	0.94, 1.01	-0.03	0.14	0.94, 1.01
Function	0.23	p<0.001	1.15, 1.37	0.24	p<0.001	1.17, 1.38
CD4 category	-0.07	0.30	0.82, 1.06	-0.06	0.34	0.83, 1.07

In the competing risk model of defaulter, the log-time interactions with weight, sex, age group facility level, district WHO stage and CD4 category were all non-significant. Similarly, in the test of proportional hazard assessment of assumption in Table 5, the interaction of log-time function was significant. Same could be seen in the competing risk model of transferred out. This indicates that function violates the Cox model assumption of proportional hazards. The Cox proportional hazards assumptions, therefore, are valid only in the 7 of the 8 covariates included in the model.

3.6.6 Stratified Competing risks model

Stratified competing risks model was fitted to in-order to adjust for violation Cox model assumptions. Function violated the Cox model assumptions for both defaulter and transferred out. A Competing risks model with the sub-hazard of defaulters and transferred out was fitted. The model was stratified on function. Table 9 presents the stratified Competing risks model.

Most of the sub-hazards for covariates included in the stratified competing risks for defaulter were significant, refer to Table 9. The sub-hazard default from treatment was the same for men and women who were enrolled on TB and HIV management in MoHSS health facilities (SHR = 0.99, 95 % CI = 0.82, 1.22). Patients aged 15-49 years were 2.14 times more likely to default compared to 0-14 years old patient (SHR =2.14, 95 % CI = 0.98, 4.68). Similarly, adults aged 50+ years were also 2 times more likely to default treatment in comparison to the 0-14 years old counterpart (SHR = 2.17, CI = 0.97, 4.88). The sub hazard ratio for patients from health centres was 45.6% less likely to default treatment relative to the clinic patients (SHR = 0.54, 95 % CI = 0.58, 1.35), while patients from hospital were 58% more likely to default treatment in contrast to the one from clinics (SHR = 0.42, 0.29).

In the stratified competing risks model for transferred out, men in were 6% more likely to be transferred out to other facilities relative to female patients (SHR = 1.06, 95% CI = 0.89, 1.27). The sub-hazard of patients aged 15-49 years old was 66% higher relative to 0-14 years old patients (HR = 1.66, 95% CI = 0.88, 3.13), while the sub-hazard for adults aged 50+ years was also 83% higher relative to children aged 0-14 years (SHR = 1.83, 95% CI = 0.95, 3.54). Patient who were receiving their treatment at health centre were at 11% lower risk of transferred out to other facilities comparing to patient at clinics (HR = 0.89, 95% CI = 0.58; 1.35), on the other

hand, patients who are receiving treatment at hospitals in the region were 45% less likely to be transferred out relative to patients from clinics (HR = 0.55, 95% CI = 0.38, 0.80).

Table 9: Stratified competing risk model

	Defaulters			**Transferred out**		
Predictors	SHR	p-value	95% CI	SRH	p-value	95% CI
Weight	0.99	0.03	0.98, 0.99	0.99	p<0.001	0.98, 0.99
Sex						
Female	Ref					
Male	0.99	0.97	0.82, 1.22	1.06	0.49	0.89, 1.27
Age group						
0-14 years	Ref					
15-49 years	2.14	0.06	0.98, 4.68	1.66	0.11	0.88, 3.13
50+ years	2.17	0.06	0.97, 4.88	1.83	0.07	0.95, 3.54
Facility level						
Clinic	Ref					
Health centre	0.54	0.01	0.35, 0.85	0.89	0.57	0.58, 1.35
Hospital	0.42	p<0.001	0.29, 0.61	0.55	0.02	0.38, 0.80
District						
Omaruru	Ref					
Swakopmund	0.27	p<0.001	0.20, 0.36	0.74	0.05	0.55, 0.99
Usakos	0.425	p<0.001	0.29, 0.62	1.49	0.02	1.07, 2.07
Walvis bay	0.392	p<0.001	0.30, 0.52	0.70	0.02	0.52, 0.94
WHO stage						
T1	Ref					
T2	0.85	0.37	0.61, 1.20	0.78	0.10	0.58, 1.05
T3	0.94	0.59	0.74, 1.18	1.13	0.23	0.93, 1.37
T4	1.63	0.02	1.20, 2.18	0.92	0.62	0.68, 1.26
Function						
Ambulatory	Ref					
Bedridden	1	omitted				
Working	1	omitted				
CD4 category						
Less than 250 cells/μL	Ref					
More than 250 cells/μL	0.62	0.02	0.43, 0.91	0.72	0.06	0.51, 1.01

Note: Ref is the reference category

Furthermore, in the stratified competing risks model for transferred out, all the sub-hazards for covariates for health district were significant. In relative to Omaruru district, patients from Swakopmund district were 26% less likely to transferred out (SHR = 0.74, 95% CI = 0.55, 0.99), Usakos district were 49% more likely to be transferred out (SHR = 1.49, 95% CI = 1.07, 2.07), while Walvisbay district were 30% less likely to be transferred out to other facilities (SHR = 0.70, 95% CI = 0.52, 0.94). In relation to WHO clinical staging, patients who were categorized under stage 2 were 22% less likely to seek for a transfer in relative to stage 1 patient (SHR = 0.78, 95% CI = 0.55, 1.05), on the other hand patients WHO are classified under stage 3 were 13% more likely to be transferred out, lastly patient who are categorized under stage 5 were 8% less likely to be transferred out (SHR = 0.92, 95% CI = 0.68, 1.26) to other health facilities. The last covariate under consideration was CD4 category, here patients who had more than 250 cells/μL were 30% less likely to be transferred out to other health facilities (SHR = 0.72, 95% CI = 0.51, 1.01).

Models fitted above, shows that the sub-hazard ratios for sex was non-significant at defaulter level, while sex and WHO stage was non-significant at transferred out level. These results suggest that weight, age group, facility level, health district, WHO stage and CD4 category are important factors in predicting default rate in health facilities in Erongo Region. On the other hand, weight, age group, facility level, health district, CD4 category are important factors in predicting transferred out rate in health facilities in Erongo Region. However, the estimates of defaulter and transferred out had almost the same confidence intervals. 424 (13.5%) in defaulter and 542 (17.2%) presented in Table 4 is however a small sample to provide estimates with reasonable confidence intervals.

3.7 Discussion

This 14-year retrospective study of patients on both TB and ART management gives an insight into survival and its determinants in Erongo Region for the period January 2003 to December 2017. Data for this study started to be captured by Heath personnel from the first day of enrolling patients into the ART program, patients are then followed up. The study considered the following risk factors; age, sex, CD4, weigh, marital status, facility level, health district name, WHO clinical stage, functional status of patient, age category as well as CD4 category of patients. Most of the risk factors considered for the study were significantly explaining the outcome in both univariate as well as multivariate analysis, with only marital status variable which was not associated with death in the univariate analysis with a p-value of 0.3032. This study revealed the overwhelming problem of the high mortality of TB-HIV co-infected patients during TB treatment. The study found that mortality rate was high (15%) among TB and HIV co-infected patients during TB treatment. This finding is similar to a study conducted in Northwest of Ethiopia by Sileshi, Deyessa, Girma, Melese, & Suarez (2013), where death occurred in 22% of patients exposed to ART during TB treatment.

In a multivariate cox proportional hazard model, the variable function violated the proportional hazard assumption, hence required the study to fit a stratified cox regression model. Mortality hazard for male patient of 0.98 (p-value=0.85, C.I. = 95% 0.80, 1.20) was slightly lower comparing to female patient only by 1.9%. In addition to sex, the study found other immunological factors associated with mortality. For example, mortality rates increased in TB-HIV co-infected patients with lower CD4 counts, patients with CD4 count more than 250 cells/μL are less likely to die comparing to patients with CD4 count less than 250 cells/μL. The results are similar to that of (Albuquerque, et al., 2014), oppositely, a study conducted in

southern India showed that a CD4 count below 200/mm^3 was not associated with a higher rate of mortality (Vijay, Kumar, Chauhan, Narayan , & Vaidyanathan, 2011). The difference between our results and these others may be that we categorized CD4 counts different intervals, which better enabled us to see the effect of CD4 counts on mortality in the Namibian context.

Like documented in the literature (Balcha, Skogmar, Sturegard, Bjorkman, & Winqvist, 2015), finding of this study also showed that those co-infected patients who were on stage 4 WHO clinical staging are significantly associated with mortality outcome compared to those who were classified as stage 1 condition. It is also consistent with a study conducted in Bair Dar, Ethiopia where more TB/HIV co-infected patients who are classified under stage 4 are more likely to experience mortality compare to those classified under stage 1, 2 and 3.

The risk of death during TB treatment was lower in patients who were initiated at a hospital compared to those who were initiated on treatment at clinic and a health centre. The reason could be that only one hospital in the region is an ART and TB site at the same time out of 4 hospitals in the region, a higher number of patients are receiving their treatment at ART and TB site which are based at clinics in the region. As a result, the severely ill clinics and health centre patients appeared to have a greater incidence of mortality, as compared to the less ill hospital patients. Mortality was higher among TB and HIV patients receiving treatment in Usakos District with hazard ration of 1.47 (p-value = 0.02, 95% CI = 1.08, 2.01) comparing to patients receiving treatment in Omaruru District. Swakopmund District patients are less likely to die comparing to the reference district with a hazard ratio of 0.59 (p-value < 0.001, 95% CI = 0.44, 0.78), patients are 41% less likely to die comparing to patients who are receiving treatment in Omaruru District.

Various models were used in the reviewed studies. Boyles, et al., (2011) used multiple Cox proportional hazard regression analysis, adjusted for competing risks. Tenthani, et al., (2014) used multiple logistic regressions with a random effect for the cohorts and competing risk regression to examine variables associated with attrition. Similarly, this study used competing risks survival model.

Across the studies, modelling sub-hazards was probably an appreciation that the other possible events are competing with the event of interest during the follow up time in an ART and TB cohort. Du Toit, et al., (2014) used logistic regression to explore factors associated with retention in HIV care in the pre-ART group. In logistic regression, odds ratio of an outcome given the predictors is modelled. The study considered default and transfer out as competing events, the overall conclusion was slightly the same. Again after fitting the competing risk model, function was found to have violated the proportionality assumption, hence the study went further to fit a stratified competing risk model by variable function. Considering transfer out as a competing, male patients were less slightly likely to default with hazard ratio 0.99 (p-value = 0.99, 95% CI = 0.82, 1.22), while with transferred out, male patients were more likely to be transferred out to other health facility.

Older patients are at higher risk of defaulting treatment comparing to young patients and they are also a risk of being transferred out to other facilities. With WHO clinical staging, all the four stage the risk increases as the stage increase for all the two competing event with the event of interest. Alarming, patients who are in stage 4 are at higher risk to default treatment with hazard ratio of 1.62 (p-value = 0.02, 95% CI = 1.20, 2.18), while on the other hand patient in stage 3 more likely to be transferred out with hazard ratio 1.13 (95% CI = 0.93, 1.37). Patient

with CD4 counts less than 250 cells/μL were more likely to be transferred out and also more likely to default treatment, this is in support with (MoHSS, 2017).

Cause-specific proportional hazards and proportional sub-hazards modelling approaches produced similar results in modelling default rate from HIV and TB treatment. However, proportional sub-hazards modelling accounts for the possibility that other events can occur. Accounting for competing events presents a more realistic reflection of reality. Therefore, a thorough understanding of the characteristics affecting outcomes can be gained through fitting sub-hazards with transfer out as competing events. Boyles and colleagues considered transfer outs as still in follow-up on ART. Transfer outs were therefore neither censored nor competing events (Boyles, Wilkinson, Leisegang, & Maartens, 2011).

Chapter 4

Modelling spatial and spatial-temporal patterns of TB and HIV mortality

4.1 Introduction

Disease mapping goes under a variety of names, some of which are: spatial epidemiology, environmental epidemiology, disease mapping, small area health studies. However, at the centre of these different names are two characteristics. First a spatial or geographical distribution is the focus and so the relative location of events is important. This brings the world of geographical information systems into play, while also including spatial statistics as a key component. The second ingredient is disease and the spatial distribution of disease is the focus. Hence the fundamental issue is how to analyse disease incidence or prevalence when we have geographical information. Sometimes this is called geo-referenced disease data, specifying the labelling of outcomes with spatial tags (Lawson, 2013).

It is apparent that none of the names listed above include the term `statistics'. This is unfortunate as it is often the case that statistical methodology (especially methodology from spatial statistics) is involved in the analysis of maps of disease. A more appropriate description of the area of focus of this work is spatial biostatistics as this emphasizes the broad nature of the focus (Lawson, 2018).

The use of survival models involving a random effect or “frailty” term is becoming more common. Usually the random effects are assumed to represent different clusters, and clusters are assumed to be independent. We consider random effects corresponding to clusters that are spatially arranged, such as clinical sites or geographical regions (Banerjee, Carlin, & Gelfand, 2015). That is, we might suspect that random effects corresponding to strata in closer proximity to each other might also be similar in magnitude. Very often, time-to-event data will be grouped

into strata (or clusters), such as clinical sites, geographic regions, and so on. In this setting, a hierarchical modelling approach using stratum-specific parameters called frailties is often appropriate, this is a mixed model with random effects (the frailties) that correspond to a stratum's overall health status (Banerjee, Bradley, & Gelfand, 2014).

Uthman, Yahaya, Ashfaq, & Uthman (2009) shows that there is no decline in growth in the number of deaths due to TB among HIV positive in most Africa countries. There is presence of 'hot-spots' and very large differences persist between sub-regions. Only by tackling TB and HIV together will progress be made in reversing the burden of both diseases. There is a great need for scale-up of preventive interventions such as the World Health Organization '3I's strategy' (intensified case finding, isoniazid preventive therapy and infection control). The results of the study states that Eastern, Southern, Western, and Middle Africa experienced an upward trend in the number of reported TB-HIV deaths. The spatial distribution of TB cases was non-random and clustered, with a Moran's $I = 0.45$ ($p < 0.001$). Spatial clustering suggested that 13 countries were at increased risk of TB-HIV deaths, and six countries could be grouped as "hot spots" (Uthman et al., 2009).

A study conducted in a rural region of north–eastern South Africa shows that the value of conducting high resolution spatial analyses in order to understand how local micro–epidemics contribute to changes seen over a wider area (Mee, et al., 2014). Comparing the two periods, there was a 30% decrease in age and gender standardised adult HIV–related and TB (HIV/TB) mortality with no change in mortality due to other causes. There was considerable spatial heterogeneity in the mortality patterns. Areas separated by 2 to 4 km with very different epidemic trajectories were identified. There was evidence that the impact of ART in reducing

HIV/TB mortality was greatest in communities with higher mortality rates in the earlier period (Mee, et al., 2014).

Spatio-temporal models are used when data is collected over space and time (Gomez-Barroso, Rodriguez-Valin, Ramis, & Cano, 2013). These approaches do not detect the cause of disease but rather provides an insight of the areas with high incidences which can be investigated further for causal relationship (Lawson, 2013). In addition, the method can identify the source of disease occurrence with respect to demographic, time and space. Spatial temporal methods are useful in monitoring of disease status in a community as they may provide information on changing disease pattern over time.

Zulu, Kalipeni, & Johannes (2014) demonstrated that spatial analysis enhanced understanding of local spatiotemporal variation in HIV prevalence, possible underlying factors, and potential for differentiated spatial targeting of interventions. Findings suggest that intervention strategies should also emphasize improved access to health/HIV services, basic education, and syphilis management, particularly in rural hotspot districts, as further research is done on drivers at finer scale. Analysis revealed wide spatial variation in HIV prevalence at regional, urban/rural, district and sub-district levels. However, prevalence was spatially levelling out within and across 'sub-epidemics' while declining significantly after 1999. Prevalence exhibited statistically significant spatial dependence nationally following initial (1995-1999) localized, patchy low/high patterns as the epidemic spread rapidly. The epidemiologic and spatial temporal analysis of TB epidemic and TB/HIV co-infection epidemic demonstrates a potential connection between TB and HIV in Urumqi. Demographic, temporal, geographic factors are the reasons of causing TB and TB/HIV co-infection epidemic (Albuquerque, et al., 2014).

The aim of this chapter is therefore, to assess the spatial distribution, presence of the spatial temporal clustering as well as survival of TB and HIV co-infections in order to identify possible high-risk areas in the region. The study used LISA and Moran's I to assess the spatial distribution in the region. This section also presented the spatial survival model to investigate spatial pattern of deaths of people infected with both TB and HIV. Lastly, the chapter discussed the space-time model to examine the temporal distribution of TB and HIV deaths.

4.2 Methods

Spatial data are any form of data attached to geographical location (Lawson, 2013). In statistical literature, there are three main forms of spatial data, namely, areal data, point referenced data, and point pattern data. Modelling such data can be considered from a hierarchical perspective, where the random effects are introduced to account for spatial dependence unexplained by the observed data. Modelling in the Bayesian framework is commonly preferred over the non-Bayesian approach that assumes parameters to be fixed but unknown, whereas the former considers all parameters to be stochastic (Ntirampeba, Neema, & Kazembe, 2017). Hence, each parameter is assigned with a probability function known as a prior function. Likelihood models and prior functions are the most important parts of Bayesian inference. The likelihood function describes the dependence of a set of parameters on sample values and it is believed to portray in its totality the information contained in a data set, while the latter provides extra information about parameters through beliefs or assumptions as they are assigned before seeing data (Lawson, 2013). The conventional likelihood model formulation assumes that data are conditionally independent. This assumption allows formulating the likelihood function as a product of individual contributions of each observation y_i as follows. Let y_i, for $i = 1, \ldots, n$, be a sample of observed values. Then the likelihood of y_i is defined by:

$$L(y_i|\theta) = \prod_{i=1}^{n} f(y_i|\theta) \tag{4.1}$$

where θ is a vector of parameters and $f(.|.)$ is a probability density function or a probability mass function. Clearly, in a spatial context where spatial units are expected to obey Tobler's law of geography, the assumption of the conditional independence of data is violated (Waters, 2016). Locations that are nearby each other have very similar values relative to those located far from one another. This implies that, within spatial analysis, spatial correlation is crucial and must be catered for during analysis. It is accounted for in the prior probability distribution level, but not in the likelihood function.

Applications with areal unit data are also commonplace. For instance, we may look at annual lung cancer rates by county for a given state over a number of years to judge the effectiveness of a cancer control program. Or we might consider daily asthma hospitalization rates by zip code, over a period of several months. From a methodological point of view, the introduction of time into spatial modelling brings a substantial increase in the scope of our work, as we must make separate decisions regarding spatial correlation, temporal correlation, and how space and time interact in our data. Such modelling will also carry an obvious associated increase in notational and computational complexity (Banerjee, Bradley, & Gelfand, 2014)

4.2.1 Measures of spatial association

For this section the study assume that our spatial process has a mean, say μ (**s**) = E (Y (**s**)), associated with it and that the variance of Y (**s**) exists for all **s** $\in D$. The process Y (**s**) is said to be Gaussian if, for any $n \geq 1$ and any set of sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, **Y** = ($Y$ ($\mathbf{s}_1$), . . ., Y ($\mathbf{s}_n$))^T has a multivariate normal distribution. The process is said to be strictly stationary (sometimes strong stationarity) if, for any given $n \geq 1$, any set of n sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and any $\mathbf{h} \in \Re$, the distribution

of (Y ($\mathbf{s}_1$) , . . . , Y ($\mathbf{s}_n$)) is the same as that of (Y ($\mathbf{s}_1$ + $\mathbf{h}$) , . . . , Y ($\mathbf{s}_n$ + $\mathbf{h}$)). Here D is envisioned as $\Re^r$ as well.

Two standard statistics that are used to measure strength of spatial association among areal units are Moran's *I* and Geary's *C* (Banerjee et al., 2014). This study will make use of the Moran's *I* to measure for spatial association among the health EAs in Erongo region. These are spatial analogues of statistics for measuring association in time series, the lagged autocorrelation coefficient and the Durbin-Watson statistic, respectively. They can also be seen to be areal unit analogues of the empirical estimates for the correlation function and the variogram, respectively. For point-referenced data, the empirical covariance function:

$$\hat{C}(t_k) = \frac{1}{N_k} \sum_{(s_i, s_j) \in N(t_k)} (Y(s_t) - \bar{Y})(Y(s_j - \bar{Y}) \tag{4.2}$$

Where $N(t_k) = \{(s_i, s_j) : \| s_i - s_j \| \in I_k\}$ for $k = 1, \ldots, K, I_k$ indexes the *kth* bin, and there are $N(t_k)$ pairs of points falling in this bin Equation (4.3) is a spatial generalization of a lagged autocorrelation in time series analysis and semivariogram:

$$\bar{\gamma}(d) = \frac{1}{2N(d)} \sum_{(s_i, s_j) \in N(d)} [(Y(s_i) - Y(s_j -)]^2 \tag{4.3}$$

where *N(d)* is the set of pairs of points such that $\|s_i - s_j\| = d,$ and $|N(d)|$ is the number of pairs in this set. This two respectively, provide customary nonparametric estimates of these measures of association.

Moran's *I* takes the form:

$$I = \frac{n \sum_i \sum_j w_i (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} w_{ij})(\sum_i (Y_i - \bar{Y})^2} \tag{4.4}$$

I is not strictly supported on the interval [−1, 1]. It is evidently a ratio of quadratic forms in **Y**, which provides the idea for obtaining approximate first and second moments through the delta method (Banerjee et al., 2014). Moran shows under the null model where the Y_i are i.i.d., *I* is asymptotically normally distributed with mean $-1/(n-1)$ and a rather unattractive variance of the form:

$$\mathrm{Var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2} \tag{4.5}$$

In (4.5) $S_0 = \sum_{i \neq j} w_{ij}$, $S_1 = \frac{1}{2}\sum_{i \neq j}(w_{ij} + w_{ji})^2$, and $S_2 = \sum_k(\sum_j w_{kj} + \sum_i w_{kj})^2$. This study recommends the use of Moran's *I* as an exploratory measure of spatial association, rather than as a "test of spatial significance".

The Moran's I statistic is a single global measure that tests for spatial association of a phenomenon (Haining, 2014). The Moran's I, like the Pearson's correlation coefficient, assume values between -1 and +1. A value of +1 indicates strong positive autocorrelation (clustering); a value of -1 indicates strong negative autocorrelation (pattern), while a value of 0 indicates a random distribution of death rates. There is also presence of spatial heterogeneity, which is due to locational effect. Overall parameter estimates for entire region may not describe the process at any given location. Spatial statistical methods have been used in wide variety of research areas: epidemiology, forestry, ecology, urban planning, and so many other research areas (Karki, 2017).

4.2.2 Spatial survival model:

In survival analysis the time to endpoint (*T*) is the random variable. Assume that a sample of individuals have associated endpoint times. Denote this sample of times as $\{t_i\}\ i = 1, 2, \ldots, m$. For now, assume these are all observed exactly. In addition to an endpoint time a geo-reference

is also available. The geo-reference could be an address location, in which case it is denoted as s_i, or a contextual spatial effect, denoted as w_i. The contextual effect is simply a factor that may have spatial correlation so that areas close together have similar risk. In addition to these ingredients, each observation unit can have covariates associated and, for the *i*th unit/person these are denoted by the vector $\mathbf{x}_i$. These covariates could be individual or contextual/ecological.

To illustrate, let t_{ij} be the time to death or censoring for subject j in EA $i, j = 1, \dots, n_i$, $i = 1, \dots, I$. Let $\mathbf{x}_{ij}$ be a vector of individual-specific covariates. The usual assumption of proportional hazards $h(t_{ij}; \mathbf{x}_{ij})$ enables models of the form:

$$h(t_{ij}; \mathbf{x}_{ij}) = h_0(t_{ij}) \exp(\beta^T \boldsymbol{x}_{ij}) \tag{4.6}$$

where h_0 is the baseline hazard, which is affected only multiplicatively by the exponential term involving the covariates. In the frailty setting, model above is extended to:

$$\begin{aligned} h(t_{ij}; \mathbf{x}_{ij}) &= h_0(t_{ij}) \omega_i \exp(\beta^T \mathbf{x}_{ij}) \\ &= h_0(t_{ij}) \beta^T \mathbf{x}_{ij} + W_i \end{aligned} \tag{4.7}$$

where $W_i \equiv \log \omega_i$ is the stratum-specific frailty term, designed to capture differences among the strata. Typically, a simple i.i.d. specification for the W_i is assumed.

$$W_i \underset{\sim}{iid} N(0, \sigma^2) \tag{4.8}$$

With the advent of Markov Chain Monte Carlo (MCMC) computational methods, the Bayesian approach to fitting hierarchical frailty models such as these has become increasingly popular (Banerjee et al, 2014). Perhaps the simplest approach is to assume a parametric form for the baseline hazard h_0. While a variety of choices (gamma, lognormal, etc.) have been explored, we will adopt the Weibull, which seems to represent a good tradeoff between simplicity and flexibility. This then produces;

$$h(t_{ij}; x_{ij}) = \rho t_{ij}^{\rho - 1} \exp(\beta^T \boldsymbol{x}_{ij} + W_i) \tag{4.9}$$

Now, placing prior distributions on ρ, β, and σ^2 completes the Bayesian model specification.

Banerjee, Carlin, & Gelfand (2015) proposed a relaxation of the Weibull model to allow a semi-parametric formulation whereby, with the patients *j*, ($j = 1, \dots, n_i$) in the *i* EAs"

$$h(t_{ij}|\mathrm{x}_{ij}) = h_{0i}(t_{ij}) \exp\{x'_{(ij)}\beta + \omega_i\} \tag{4.10}$$

where h_{0i} denotes the county-specific baseline hazard. For an individual with censoring indicator Y_{ij} (0 if alive, and 1 if dead) the likelihood contribution is then:

$$h(t_{ij}|\mathrm{x}_{ij})^{\gamma ij} \exp\{-H_{0i}(t_{ij})\, exp\{x'_{(ij)}\beta + \omega_i\}\} \tag{4.11}$$

Where $-H_{0i}(t_{ij}) = \int_o^{t_{ij}} h_{0i}(u)\, du$ is a county-specific cumulative baseline hazard, and the covariates are assumed to be not time-dependent. The baseline hazard appears in this likelihood and so must be estimated. Different approaches have been proposed for estimation of this baseline. One approach assumes a gamma process which is a function of a parametric cumulative hazard (Collett, 2015). Another is the use of beta mixtures (Banerjee et al., 2014). A related spatial model was developed by Bastos & Gamerman (2006) whereby they allowed time-dependent covariates which are fixed within small time periods and a CAR spatial frailty. They assume no separate baseline risk however.

The Cox proportional hazards model has been applied in a spatial context by Jerrett, et al., (2013). In the context of population and mortality survival, the authors posited that the partial likelihood could be used without recourse to the estimation of the baseline. For ordered uncensored times $\{t_{(1),\dots\dots,}t_{(m)}\}$, then the partial likelihood is given by:

$$\prod_{\boldsymbol{i}} [\exp\{\mathbf{x}'_{(i)}\beta\} / \sum_{j \in R_i} \exp\{\mathbf{x}'_{(j)}\beta\}] \tag{4.12}$$

where R_i is the set of those individuals at risk just before the *i*th event time. This can be used for the estimation of regression parameters. This is less parametric than models that include

baseline components. Note that spatial effects can be included again as contextual effects by extending the specification of the intensity term $\exp\{\mathbf{x}'_{(j)}\beta\}$ to include the random effects:

$$\exp\{\mathbf{x}'_{(j)}\beta\} + \omega_i + \nu_i\} \tag{4.13}$$

Where ω_i and ν_i are random contextual effects which could be at an aggregate level such as EAs, district and constituency. In addition, these effects could also be purely individual (in the sense of frailty rather than context).

4.2.3 Spatial temporal survival model

In this section we use semiparametric (Cox) hierarchical Bayesian frailty model in (Banerjee et al, 2014) to capture for spatiotemporal heterogeneity in survival data. We then use these models to describe the pattern of death in Erongo Region while accounting for important covariates, spatially correlated differences in the hazards among the Enumeration Areas (EA), and possible space-time interactions.

We apply the framework in the preceding section to incorporate temporal dependence. Here we have t_{ijk} as the response (time to death) for the *j*th subject residing in the *i*th district who died in the *k*th year, while the individual-specific vector of covariates is now denoted by $\mathbf{x}_{ijk}$, for $i = 1, 2, ..., I$, $k = 1, . . ., K$, and $j = 1, 2, ..., n_{ik}$. We note that "time" is now being used in two ways. The measurement or response is a survival time, but these responses are themselves observed at different areal units and different times (years). Furthermore, the spatial random effects W_i in the preceding section are now modified to W_{ik}, to represent spatiotemporal frailties corresponding to the *i*th district for the *k*th diagnosis year. Our spatial frailty specification in (4.1) now becomes:

$$h(t_{ijk}; \mathbf{x}_{ijk}) = h_{0i}(t_{ijk}) \exp(\beta^T \boldsymbol{x}_{ijk} + W_{ik}) \tag{4.14}$$

Our CAR prior would now have conditional representation $W_{ik}| \quad W_{(i\prime\neq i)k} \quad \sim$

$N\left(\overline{W}_{ik}, 1/(\lambda_k m_i)\right)$.

Note that we can account for temporal correlation in the frailties by assuming that the λ_k are themselves identically distributed from a common hyperprior. A gamma prior (usually vague but proper) is often selected here, since this is particularly convenient for MCMC implementation. A flat prior for β is typically chosen, since this still admits a proper posterior distribution. Adaptive rejection Metropolis-Hastings sampling are usually required to update the $\mathbf{W}_k$ and β parameters in a hybrid Gibbs sampler.

We remark that it would certainly be possible to include both spatial and nonspatial frailties. This would mean supplementing our spatial frailties W_{ik} with a collection of nonspatial frailties, say, $V_{ik} \overset{iid}{\sim} N(0.1/\tau k)$. We summarize our full hierarchical model as follows:

$$L(\beta, \mathbf{W}; \mathbf{t}, \mathbf{x}, \gamma) \propto \prod_{k=1}^{K} \prod_{i=1}^{I} \prod_{j=1}^{n_{ik}} \left\{h_{0i}(t_{ijk}; \mathrm{x}_{ijk})\right\}^{\gamma \mathrm{ijk}} \times \exp\left\{(-H_{0i})\exp(\beta^T \mathrm{x}_{ijk} + W_{ik} + V_{ik})\right\}, \qquad (4.15)$$

$$\text{where} \quad p(\mathrm{W}_k|\lambda_k) \sim CAR(\lambda_k) \quad p(\mathrm{V}_k|\tau_k) \sim N_I\,(0, \tau_k \mathbf{I})$$

$$\text{and } \lambda_k \sim G(a, b)\ ,\ \tau_k \sim G(c, d) \text{ for } k = 1, 2, \ldots, K\ .$$

In the sequel we adopt the beta mixture approach to model the baseline hazard functions $H_{0i}(t_{ijk})$ nonparametrically (Banerjee et al., 2015).

4.3 Model GOF measure using DIC

Goodness-of-fit criteria vary depending on the properties of the criteria and the nature of the model. In conventional generalized linear modelling with fixed effects, the deviance is an important measure (Lawson, 2013). Usually this measure of model adequacy compares a fitted

model to a saturated model. It is based on the difference between the log likelihood of the data under either model:

$$D = -2\left[l\left(y|\hat{\theta}_{fit}\right) - l\left(y|\hat{\theta}_{sat}\right)\right] \quad (4.16)$$

The saturated model has a single parameter per observation. Often a relative measure of fit is used so that deviances are compared and the change in deviance between model 1 and model 2 is used:

$$\Delta D = -2\left[l\left(y|\hat{\theta}_1\right) - l\left(y|\hat{\theta}_2\right)\right] \quad (4.17)$$

Hence the saturated likelihood cancels in this relative comparison. The deviance is used in goodness-of-fit measures in Bayesian modelling, but usually without reference to a saturated model.

One disadvantage of using the deviance directly is that it does not allow for the degree of parameterization in the model: a model can be made to more closely approximate data by increasing the number of parameters (Lawson, 2013). Hence attempts have been made to penalize model complexity. The Deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) has been proposed by Spiegelhalter et al., (2002) and is widely used in Bayesian modelling. This is defined as:

$$DIC = 2E_{\theta|y}(D) - D(\left[E_{\theta|y}(\theta)\right] \quad (4.18)$$

where *D*(.) is the deviance of the model and *y* is the observed data. Note that the DIC is based on a comparison of the average deviance ($\overline{D} = -2\sum_{g=1}^{G} l(y|\theta^g)/G$ and the deviance of the posterior expected parameter estimates, $\hat{\theta}$ say ($\widehat{D}\left(\hat{\theta}\right) = -2l(y|\hat{\theta})$. For any sample parameter value θ^g the deviance as $\widehat{pD} = \overline{D} - \widehat{D}(\hat{\theta})$ and the $DIC = \overline{D} + \widehat{pD} = 2\overline{D} - \widehat{D}(\hat{\theta})$.

Unfortunately in some situations the $\widehat{pD}$ can be negative (as it can happen that $\widehat{D}(\hat{\boldsymbol{\theta}}) > \widehat{D}$). Instability in *pD* can lead to problems in the use of this DIC. For example, mixture models, or more simply, models with multiple modes can 'trick' the *pD* estimate because the

overdispersion in such models (when the components are not correctly estimated) leads to $\widehat{D}(\widehat{\boldsymbol{\theta}}) > \overline{D}$. However, it is also true that inappropriate choice of hyper-parameters for variances of parameters in hierarchical models can lead to inflation also, as can nonlinear transformations (such as changing from a Gaussian model to a log normal model). In such cases it is sometimes safer to compute the effective number of parameters from the posterior variance of the deviance. Gelman et al., (2013) propose the estimator:

$$\widetilde{pD} = \frac{1}{2}\frac{1}{G-1}\sum_{g=1}^{G}(\widehat{D}(\theta^g) - \overline{D})^2 \tag{4.19}$$

This value can also be computed from sample output from a chain. An alternative estimator of the variance is direct available from output:

$$Var(D) = \frac{1}{G-1}\sum_{g=1}^{G}(\widehat{D}|(\theta^g) - \overline{D})^2 = 2\widetilde{pD} \tag{4.20}$$

Hence a DIC based on this last variance estimate is just $DIC = \overline{D} + \widehat{var}(D)$. Note that the expected predictive deviance (EPD: D_{pr}) is an alternative measure of model adequacy and it is based on the out-of-sample predictive ability of the fitted model. The quantity can also be approximately estimated as $\widehat{D}_{pr} = 2\overline{D} - \widehat{D}(\widehat{\boldsymbol{\theta}})$.

4.4 Modelling spatial patterns of TB and HIV related deaths

4.4.1 Spatial autocorrelation

Spatial autocorrelation plays an important role in geographical analysis, Global Moran's I is the most common measures of the spatial autocorrelation which determine the overall strength of spatial dependence (Ntirampeba, Kazembe, & Neema, 2017). This statistic is a useful measure of the overall clustering. The global Moran's statistic was 0.17 (p-value<0.001) with a variance of 0.001. Thus, the positive significant Moran's I value indicates that values in neighbouring Enumeration Areas tend to cluster. To detect local spatial patterns, we have used local Moran's I. It enhances to identify clusters, which are observations with very similar neighbours and hotspots, which are characterised by observations with very different neighbours. Figure 16 shows the map of local Moran's I statistics. From this figure, it can be noted that Enumeration Areas around Karibib and Usakos districts have elevated positive Moran's I values. These results have inspired the spatial analysis undertaken in the subsequent sections in the study.

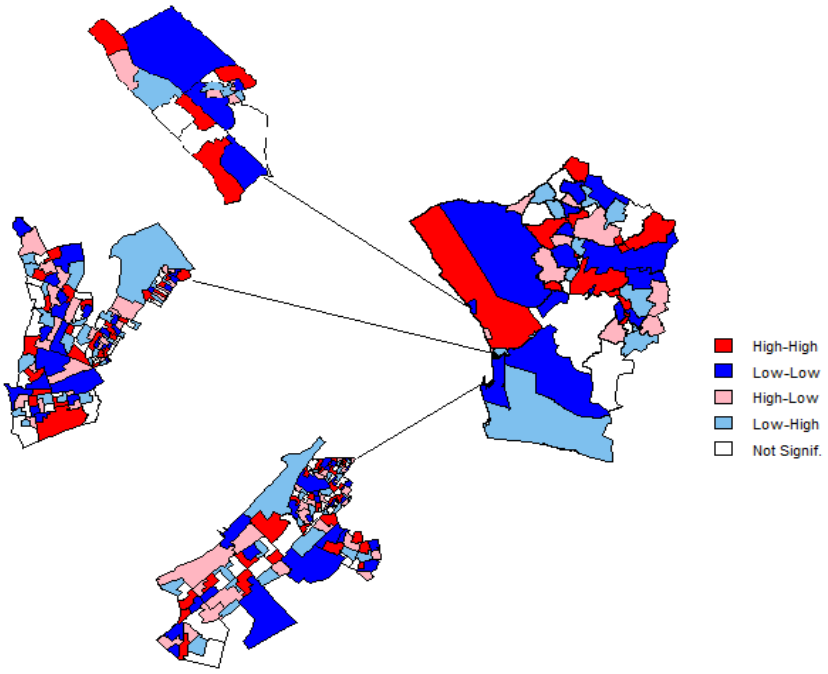


Figure 16: LISA Clustered map for TB and HIV death cases

4.4.2 Frailty and spatial effects

Models referred to in this section are those under spatial survival modelling under section 4.3.2. The study implemented a simple nonhierarchical ("no-frailty") model, spatial ("frailty") model (4.6) and random effect model with spatially-correlated term (w_i) added (4.7), and finally the last model is a convolution model with the uncorrelated random effect (v_i) and w_i added to the basic model (4.13). Table 10 below compares our four models in terms of the criteria discussed in subsection 4.4.

Table 10: Comparison of four spatial models for seizure data: basic model and models 1 through 3

Model	$\overline{\boldsymbol{D}}$	**pD**	**DIC**
Basic model (Null)	3294.61	15.19	3324.99
Spatial (1)	3152.38	71.80	3295.98
Random (2)	3184.60	43.25	3269.11
Spatial + Random (3)	3169.37	50.20	3269.76

The models fitted were as follows: The first line 1 displays the overall goodness-of-fit results for the basic model with no spatial random effects. Under this model, the DIC was found to be 3394.61, with effective number of parameter pD = 15.19. The results of fitting these different models are found on Table 10. It is interesting to note that all models including spatially referenced random effects (Models 1, 2, 3) have a considerably lower DIC than the basic model. The best fitting model by the DIC criterion is model 2 with the uncorrelated random effect that has a DIC = 3269.11 and effective number of parameter pD = 43.25. Although the convolution model yields a slightly higher DIC, it has lower DIC and pD than a model with only the correlated spatial effect.

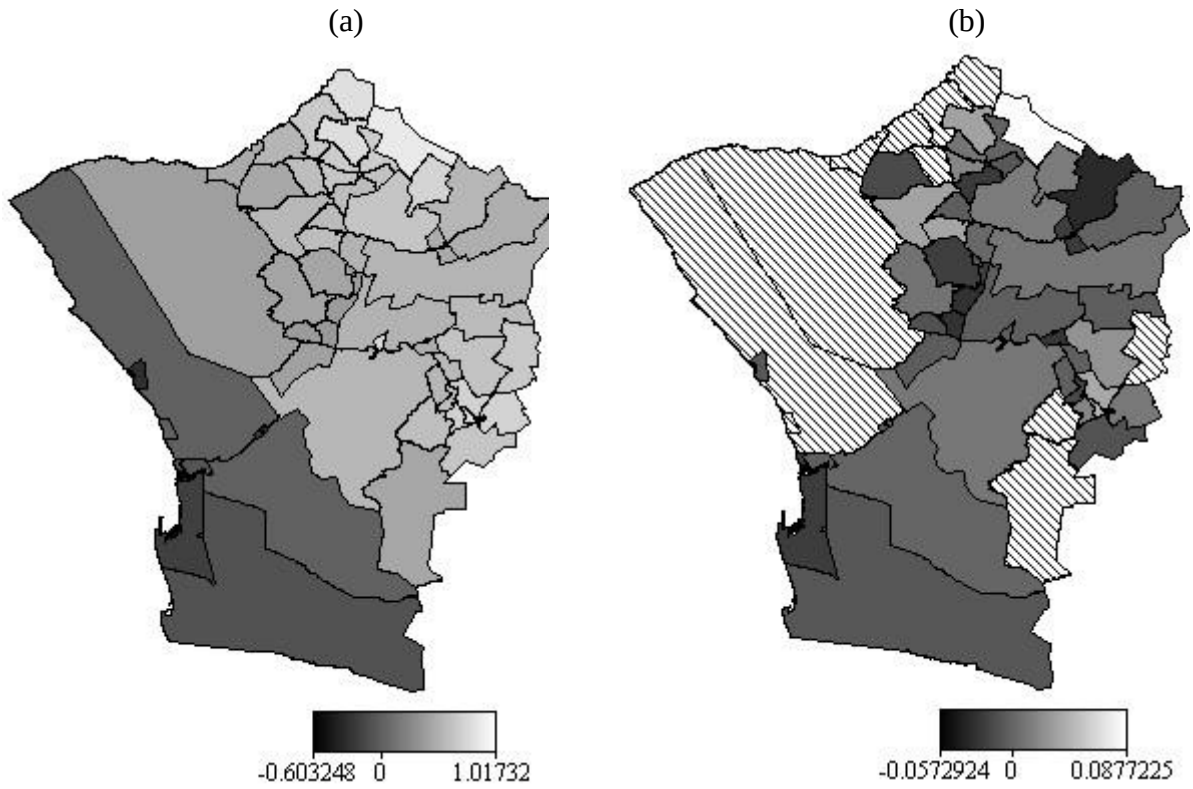


Figure 17: (a) Posterior mean of the structured spatial effect for each area and (b) posterior mean of the unstructured spatial effect. Areas with no or insufficient data are marked with diagonal solid lines

It is reassuring that the uncorrelated noise model has best fit as we did not present any overt spatial structure in the data. However, the fact that any spatially-referenced model (whether correlated or not) leads to much reduced DIC is an important consideration. The reason for the closeness of models 1 and 2 is largely because the spatially-correlated term (w_i) will often mimic small scale variation found. A further advantage of the graphs is, that they allow to visualise the estimation results for the uncorrelated spatial effects. Figure 17 displays the results of the structured (a) and unstructured posterior mean spatial effect (b). Map further, illustrate the structured model and unstructured has posterior mean ranging from -0.60 to 1.02 and -0.06 to 0.09 respectively. Enumeration Ares with no or insufficient data are marked with diagonal solid lines.

4.5 Modelling spatial temporal patterns of mortality of TB and HIV

Models referred to in this section are those under spatial temporal survival modelling under section 4.3.3. The first three models are simple separable models with no spatial-temporal interaction: a model with a random spatial and temporal term and an autoregressive temporal effect (model 1); a model with random spatial, (model 2); and a model with spatial and uncorrelated random effect. Table 11 displays the DIC results for models 1 through 6.

Table 11: Space-time models for TB and HIV related data set

Model	$\overline{\boldsymbol{D}}$	**pD**	**DIC**
Model 1	3673.59	45.71	3765.02
Model 2	3182.61	44.67	3271.94
Model 3	3673.59	45.71	3765.02
Model 4	3671.60	46.69	3764.98
Model 5	3182.61	44.67	3271.94
Model 6	3163.98	53.76	3271.50

Model 3 has spatial and uncorrelated random effect, models 4 has an interaction of temporal and spatial effect and the last involve temporal, spatial and also adding an uncorrelated random effect. It is clear that Models 1 and 2 while parsimonious are far from the best models. It can be also observed clearly on Table 8 the Model 7 that has an interaction of temporal and spatial component yield the lowest DIC of 3271.51 and effective number of parameter pD of 53.76. Of course these results depend on prior specifications and in any particular applications, sensitivity to prior specification should be examined (Lawson, 2018). For the model with lowest DIC, various posterior summaries are available. Table 12 displays the summary of posterior mean, standard deviation and the three posterior quantiles of q2.5%, q50% and q97.5% for the main effect of the model.

Table 12: Posterior summary statistics: fixed effects for model 6

Covariates	**Mean**	**Std. Dev.**	**2.5%**	**50%**	**97.5%**
Intercept	-4.52	0.91	-6.35	-4.48	-2.85
Sex	0.11	0.11	-0.11	0.10	0.32
Age category	0.42	0.12	0.21	0.41	0.62
CD4 category	-0.47	0.13	-0.74	-0.47	-0.23
Marital status	0.25	0.14	-0.01	0.25	0.53
Facility level	0.33	0.13	0.08	0.32	0.60
District	0.05	0.11	-0.18	0.05	0.26
Who stage	0.15	0.05	0.06	0.16	0.25
Function	-0.76	0.07	-0.87	-0.74	-0.61

In the tables, all of the predictors are significant at the 0.05 level. Overall, the results are broadly similar to those parametric analysis of (Banerjee et al., 2015). Since the reference group for the sex variable is female, we see that men have slightly higher hazard of death compared to women, respectively; the posterior median hazard rate increases by a factor of $e^{0.104} = 1.11$ for the latter group. Higher age at diagnosis also increases the hazard, but a larger number of primaries actually leads to a lower hazard, presumably due to other competing risk from one of the competing risk event discussed in section 3.4.3. Figure 18 to Figure 23 present the exceedance of the posterior mean of the structured spatial temporal effect for the lowest DIC model fitted for the data (model 6). Again, like in section 4.4.2, Enumeration Areas with no or insufficient data are marked with diagonal solid lines.

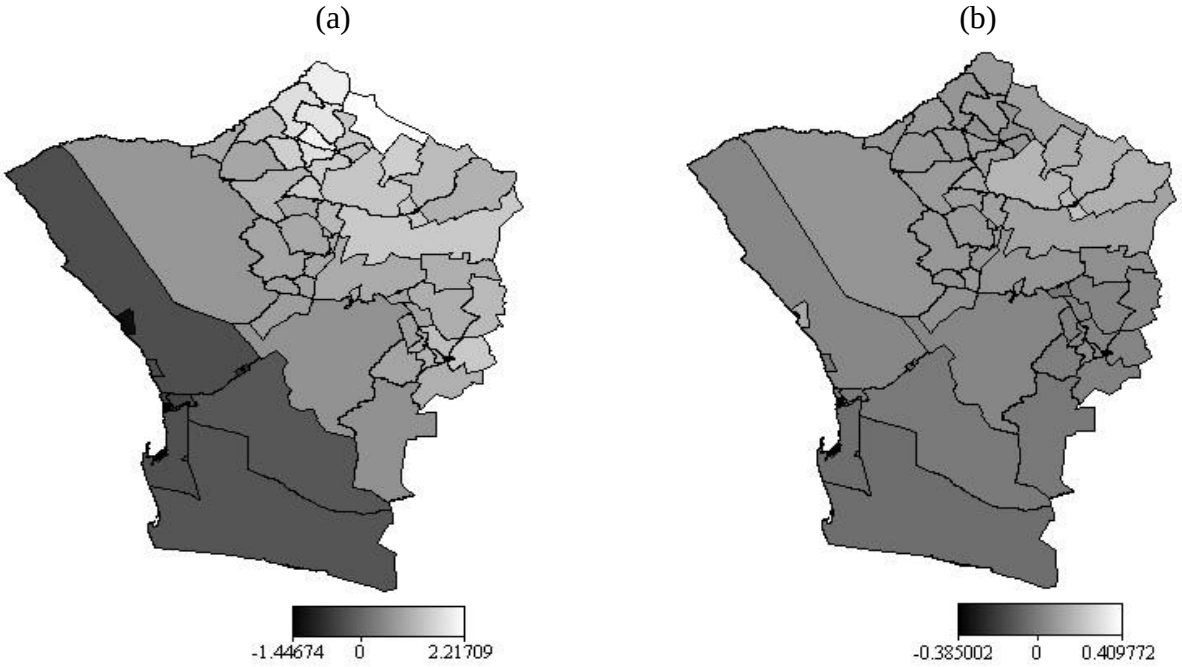


Figure 18: (a) Posterior means of the temporal variability in risk for 2007 and (b) posterior mean of the temporal variability in risk for 2008

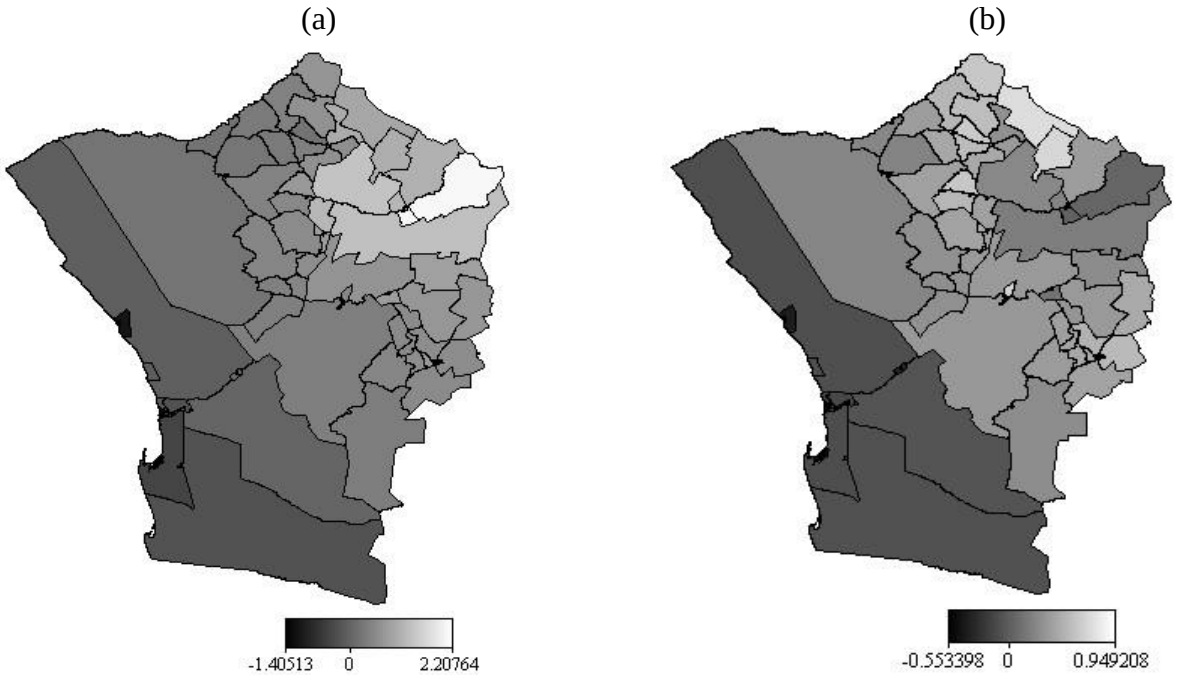


Figure 19: (a) Posterior mean of the temporal variability in risk for 2009 and 2010 (b) posterior mean of the temporal variability in risk for 2010

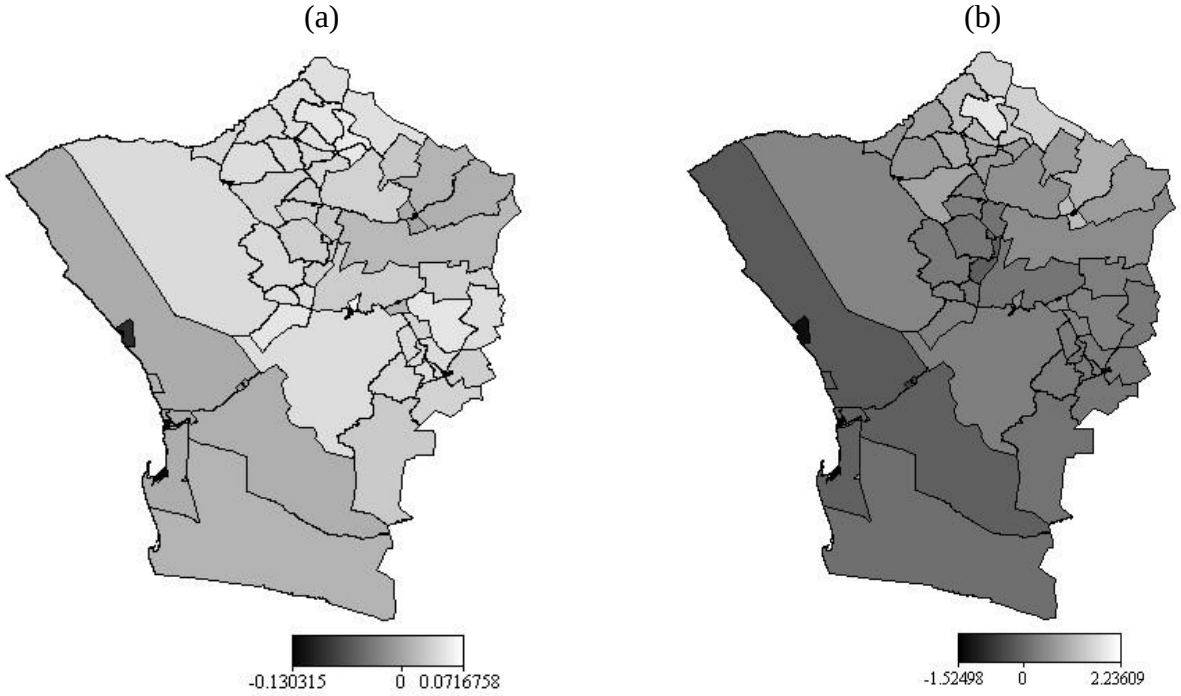


Figure 20: (a) Posterior mean of the temporal variability in risk for 2011 and (b) posterior mean of the temporal variability in risk for 2012

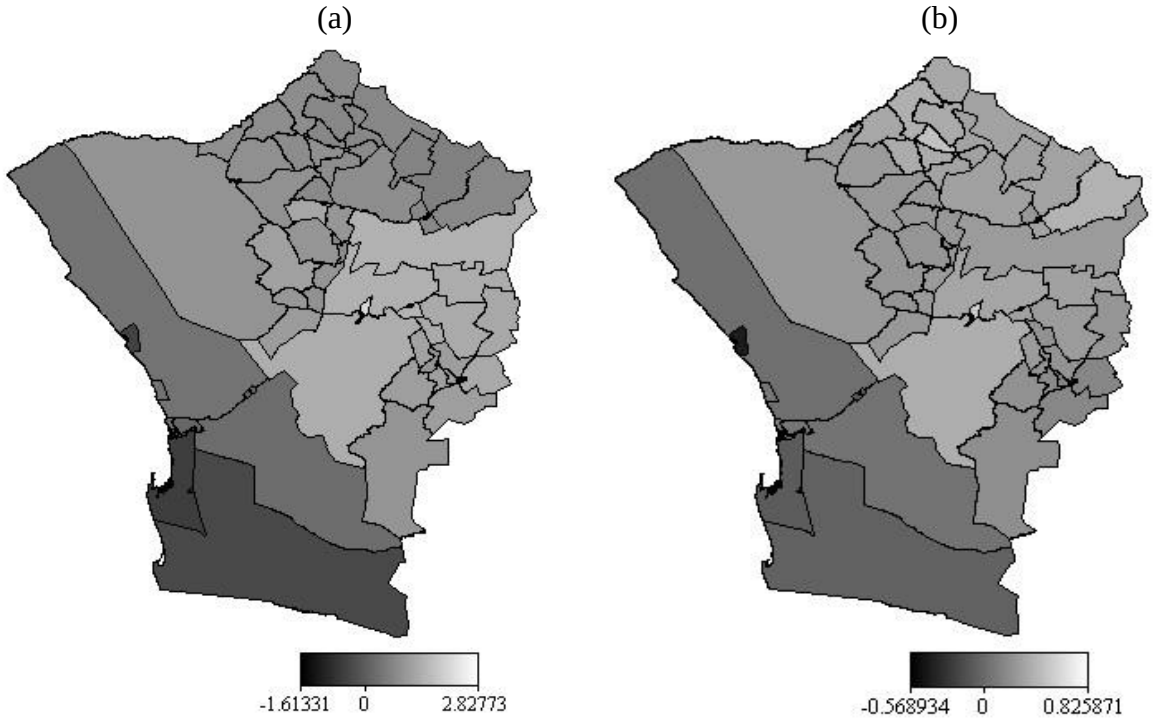


Figure 21: (a) Posterior mean of the temporal variability in risk for 2013 and (b) posterior mean of the temporal variability in risk for 2014

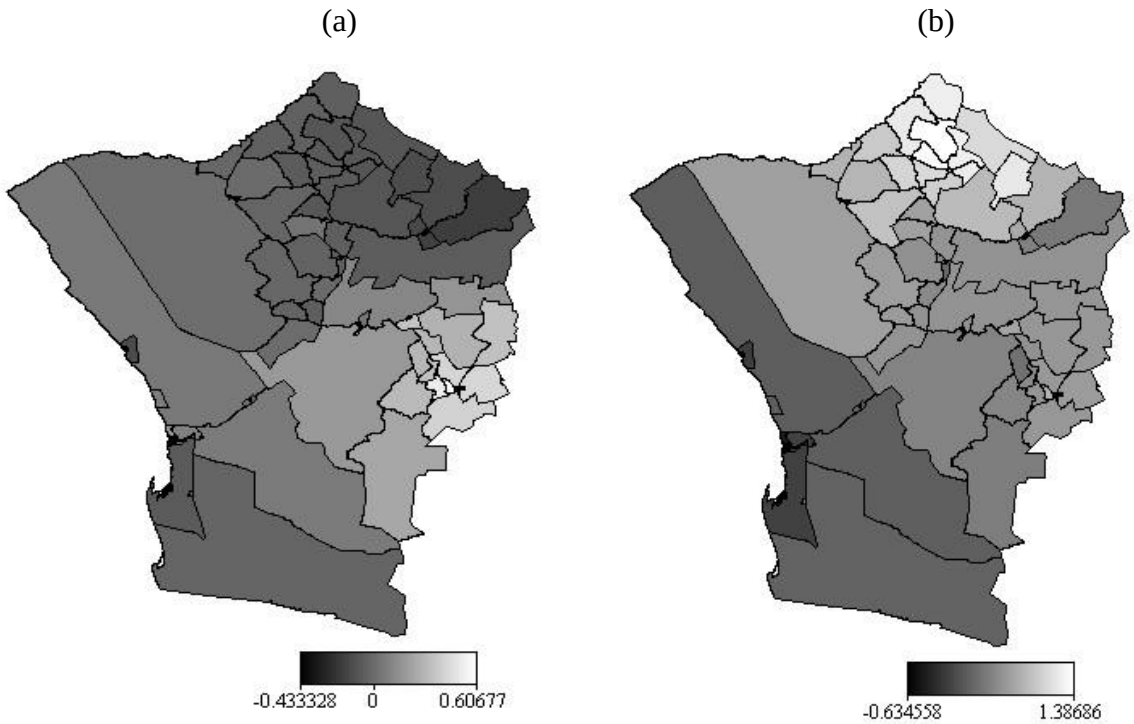


Figure 22: (a) Posterior mean of the temporal variability in risk for 2015 and (b) posterior mean of the temporal variability in risk for 2016

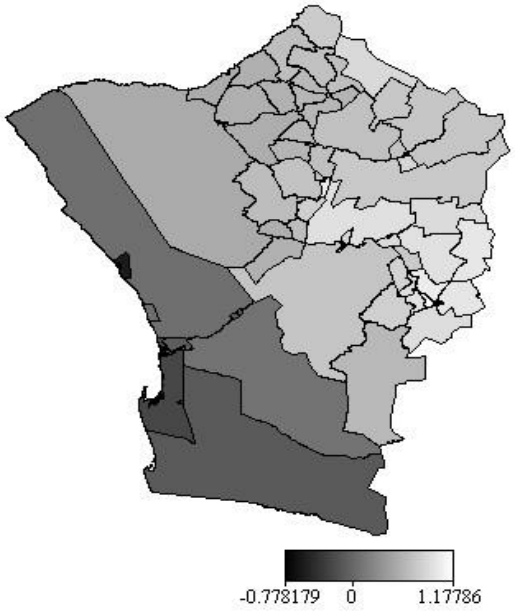


Figure 23: Posterior mean of the temporal variability in risk for 2017

Figures 18 to 23, shows spatial temporal variability in risk over the years on the map of Erongo Region, the colours ranged from light grey to dark grey with the extreme negative random effects corresponding to extreme dark grey and the extreme positive random effects corresponding to extreme light grey. Three different numbers were used to distinguish significant observed random effects. Dark grey denoted by (-1) indicated significant negative random effects, (0) indicated non-significant random effects, and light grey denoted by (1) represented significant positive random effects. Figure 16 illustrates the structured model with a posterior mean ranging from -1.45 to 2.24 for the year 2007 and -0.39 and 0.41 for the year 2008.

4.6 Discussion

The study shows that areas around Swakopmund, Walvisbay, Usakos, Karibib and Otjimbingue are a hotspot and this has been shown on Figure 16, by LISA. The positive significant Moran's I value indicates that values in neighbouring Enumeration Areas tend to cluster and neighbouring EAs have the same characteristics. County-level direct estimates of less common mortality outcomes are often highly unstable. Many prior studies on county-level variations in less common causes of mortality outcomes have relied on estimates aggregated over time or larger geographic areas (Khana, Lauren, Holly, & Margaret, 2018). However, this type of aggregation precludes the examination of detailed temporal and spatial trends. To overcome these limitations, this study uses hierarchical Bayesian methods to generate robust model-based estimates of regional-level to examine the spatial and temporal variations across a span of 14 years.

We found spatial and spatial-temporal clusters and variations in the distribution of mortality of TB and HIV in the western and east districts of the study area. These clusters were stable

over the years with the exception of a few location differences. This could be explained by a high transmission over many years due to the existence of disproportionate high-risk factors, and a varying program performance.

In our finding, most cluster locations were identified in urban areas, semi-urban areas with a high population density, as well as neighbouring areas close to towns and areas near road networks, which connect major towns. Various studies have reported that poor socioeconomic conditions such as social inequality, low income, poverty, poor housing conditions, overcrowding and social unrest could all be risk factors for the high burden and variations of disease occurrence (Corbett, et al., 2009) (Couceiro, Santana, & Nunes, 2011) (Gomez-Barroso, Rodriguez-Valin, Ramis, & Cano, 2013) (Maciel, et al., 2010) (Oren, Koepsell, Leroux, & Mayer, 2012) (Wong, Yadav, Nishikiori, & Eang, 2013). In addition, patient care factors (Randremanana, Richard, Rakotomanana, Sabatier, & Bicout, 2010) and poor access to health care and TB and HIV control services could also contribute to a high rate of the disease (Ng, Wen, Wang, & Fang, 2012) since infectious cases may remain undiagnosed and may not acquire treatment, which could consequently contribute to the transmission dynamics of the disease.

Furthermore, most urban towns where the clusters identified were the capitals of the districts, had market places, had public transportation routes and were the hub for different socio-economic activities. The better access to road and movement using public transportation in a crowded and poorly ventilated environment may assist in facilitating contact with infectious cases, which could favour a transmission of the disease in the locations where clusters were detected (Feske, Teeter, Musser, & Graviss, 2011) (Edelson & Phypers, 2011) . Therefore, the

inclusion of these factors in cluster analyses in the future may help to improve the understanding of their effect on the clusters of the diseases in the study area.

Improving TB and HIV control efforts could help reduce the transmission and change the geographic distribution of TB. In our study, despite different intervention programs aimed at reducing disease transmission and improving case detection over many years, the unusually high rates of the diseases persisted in the same places, with the most likely spatial clusters showing a stable pattern in the preceding years during the study. This could explain, at least in part, that the interventions may not be properly focused on influencing the disease epidemiology or could be due to a low case finding and poor treatment of infectious cases, which may indicate the continued transmission of the two disease.

Chapter 5

Conclusions and Recommendations

5.1 Conclusions

The study found spatial temporal clusters and spatial variations of TB and HIV related mortality in Erongo Region. As a result, TB and HIV in the study area did not uniformly occur in different geographic settings, and revealed a non-random distribution. The findings can be used to guide TB control programs to help devise effective TB and HIV control strategies for the geographic areas characterized by the highest cases of TB and HIV in the Region. Further investigations based on individual level locations are needed to identify the presence of localized spatial clustering, especially on the MoHSS health district and causes for unusually high rates in those areas by incorporating socioeconomic factors, type of TB strain, other opportunistic infections and access to health services so as to improve our understanding of the possible causes for unusually high disease burden in the region.

The results showed some form of clustering which indicated that one EA may be surrounded by other EAs with similar characteristics suggesting that the closer a EA is to another, the more similar the EAs are. This confirms the Tobler's first law of geography which states that everything is related to everything else, but near things are more related than distant things (Miller, 2004). Also, the hot spot analysis showed spatial association of the EAs with hot spot areas in red and cold spots in blue which further indicated the existence of sub-areas that developed TB and HIV co-infections differently over space.

The spatial pattern of the disease has been seen to mirror the spatial pattern of the population at risk which included the region with high level of poverty and overcrowding. These identified areas of high risk will enable the Division of Leprosy, Tuberculosis and HIV Disease to

propose innovative interventions to minimize the risk of disease at both individual and population level. Also, it will help the division in the effective deployment of resources. Keeping in mind the difficulties in diagnosis of HIV among TB patients, the cases reported might not be 100% efficient and hence a need for tools to accurately and timely diagnose the presence of the disease or the infection and probably a mortality study be conducted to find out other causes of death TB and HIV patients.

The model fit was examined via DIC comparisons. Amongst the models fitted, the contribution of different space and time components was examined by the subsequent reduction in DIC values and the effective number of parameters to estimate. The best-fitting model captured the spatial autocorrelation and the time dependence structure of the data and was further improved by using time varying covariates accounting for the extra variability that was not captured by the main time and EA effects. The best fitting model was found to have the lowest DIC with small number of effective parameters to estimate as compared to the model without time varying covariates. However, the temporal random effect was found to be an autoregressive process of order 1 which dampens out after a certain period of time.

Several limitations affected the study. First, the study used routine secondary hospital data. Routine hospital data is inherently poor particularly in resource limited context like Namibia. For example, routine hospital data is usually incomplete and sometimes inconsistent, which may bias the results of the analyses. It is therefore important to interpret studies that use routine hospital data within the context of the limitations.

Second, the study would have been stronger if other characteristics were included in the modelling. It was however not possible to include all characteristics that have been found to be

significantly associated with mortality of TB and HIV in similar studies due to poor quality of data. Literature has indicated that distance from the health facility to patients' place of residence has been found to be an important predictor of one of our competing event. Routine hospital data does not include distance in Namibia. ART and TB study dataset subsequently did not include this variable.

Finally, most survival analysis methods assume that censoring is noninformative, but censoring might be informative. This might have introduced bias into our study as outcome of patient is unknown due to defaulter or Lost to Follow Up.

Generally, this study has demonstrated the usefulness of spatial analysis in describing the geographical distribution of TB and HIV mortality in the different EAs in Erongo Region

5.2 Recommendations

It will be very important for the Division of Leprosy, TB and HIV to continue giving a special focus on the co-infection of TB and HIV. This will ensure that the future epidemics will start to follow a decreasing trend as the (MoHSS, 2016a) and (MoHSS, 2016c) has shown a consistent prevalence trend over the years. Mortality of TB and HIV in this study was significant and that some areas were found to be hot spots. Therefore, there is a need to have targeted intervention among these Areas to ensure that Namibia strives to attain its firth National Development Plan Goal (NDP5) which is intended to reduce mortality of TB among HIV patients. There is also because of the complexity in the diagnosis of tuberculosis among HIV patients, there is a probability that the number of cases reported could be underestimated and hence there is need for refinement of existing tools and development and testing of new tools are urgently required to improve early diagnosis and treatment of TB among HIV patients.

The finding of the research also proposes a sensitization of the ART nurses, DTLCs and health care workers at the facility so as to ensure a turnaround in the way we view tuberculosis among HIV patients and vice versa, hence optimize the testing of HIV in all TB patients and TB among all HIV patients. It is possible to identify a number of areas for further research which will help to inform areas of policy interest. These include further studies on spatial survival of TB and HIV in Namibia and also determine the risk factors associated with TB and HIV. Also, more studies need to be done on other infectious diseases e.g. malaria so as to be able to identify the hot spots and their risk factors in Namibia.

The factors studied in this study were age, weight, sex, CD4 category, marital status, WHO clinical stage, function, facility level and district. However, the study did not fully understand the other factors related mortality of TB and HIV, for example types of TB, distance to health facility, socio economic status and viral suppression among TB and HIV patient. Depends on availability of data, future studies on mortality of TB and HIV may include these factors in modelling. Further studies aimed at determining the hot spots areas of the infections among the general population in Namibia will be of value in determining the population prevalence, as it will identify region that are hot spot of mortality of TB and HIV, so that proper public health measures could be instituted.

References

Albuquerque, M. d., Coimbra, I., Batista, J. d., Maruza, M., Ximenes, R. A., Lacerda, H. R., . . . Rodrigues5, L. C. (2014). Empirical treatment for TB in HIV: lessons from a cohort study of people living with HIV treated in Recife, Brazil. *BMC Public Health, 14(289)*, 1471-2458.

Allison, P. D. (2014). *Event history and survival analysis: Regression for longitudinal event data.* California: (Vol. 46). SAGE publications.

Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation, 133(6)*, 601-609.

Bakoyannis, G., & Touloumi, G. (2012). Practical methods for competing risks data: a review. *Statistical methods in medical research, 21(3)*, 257-272.

Balcha, T. T., Skogmar, S., Sturegard, E., Bjorkman, P., & Winqvist, N. (2015). Outcome of tuberculosis treatment in HIV-positive adults diagnosed through active versus passive case-finding. *Glob Health Action, 8(27048)*, doi:10.3402/gha.v8.

Banerjee, S., Bradley, C. P., & Gelfand, A. E. (2014). Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: CRC Press.

Bastos, L. S., & Gamerman, D. (2006). Dynamic survival models with spatial frailty. *Lifetime data analysis, 12(4)*, 441-460.

Belayneh, M., Giday, K., & Lemma, H. (2015). Treatment outcome of human immunodeficiency virus and tuberculosis co-infected patients in public hospitals of eastern and southern zone of Tigray region, Ethiopia. *Brazilian Journal of Infectious Diseases, 19(1)*, 47-51.

Boyles, T. H., Wilkinson, L. S., Leisegang, R., & Maartens, G. (2011). Factors influencing retention in care after starting antiretroviral therapy in a rural South African programme. *PLoS ONE, 6(5)*, e19201.

Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications.* Chapman and Hall/CRC.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective.* Chapman and Hall/CRC.

Catala, L., Orcau, Á., de Olalla, G., Millet, J. P., Rodríguez-Mondragón, A., & The TB-HIV Working Group. (2011). Survival of a large cohort of HIV-infected tuberculosis patients in the era of highly active antiretroviral treatment. *The International Journal of Tuberculosis and Lung Disease, 15(2)*, 263-269.

CDC. (2017, May 30). *Centers for Disease Control and Prevention*. Retrieved from https://www.cdc.gov: https://www.cdc.gov/hiv/basics/livingwithhiv/opportunisticinfections.html

Churchyard, G., & Lawn, S. D. (2009). Lawn, S. D., & Churchyard, G. (2009). Epidemiology of HIV-associated tuberculosis Running Head: Epidemiology of TB/HIV. *Current Opinion in HIV and AIDS*, 325.

Collett, D. (2015). *Modelling survival data in medical research.* New York: Chapman and Hall/CRC.

Corbett, E. L., Bandason, T., Cheung , Y. B., Makamure, B., Dauya, E., & Munyati, S. S. (2009). Prevalent infectious tuberculosis tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 13(10)*, 1231–1237.

Couceiro, L., Santana, P., & Nunes, C. (2011). Pulmonary tuberculosis and risk factors in Portugal: a spatial analysis. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 15(11)*, 1445–1454.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal*, 220.

Cox, D. R. (2018). *Analysis of survival data.* Routledge.

Dignam, J. J., Zhang, Q., & Kocherginsky, M. N. (2012). Dignam, J. J., Zhang, Q., & Kocherginsky, M. N. (2012). The use and interpretation of competing risks regression models. *Clinical Cancer Research*, clincanres-2097.

Du Toit, E., Van Schalkwyk, C., Dunbar, R., Jennings, K., Yang, B., Coetzee, D., & Beyers, N. (2014). Missed opportunities for retention in pre-ART care in Cape Town, South Africa. *PloS one, 9(5)*, e96867.

Edelson, P. J., & Phypers, M. (2011). TB transmission on public transportation: a review of published studies and recommendations for contact tracing, 9(1. *Travel medicine and infectious disease*, 27–31.

Feske, M. L., Teeter, L. D., Musser, J. M., & Graviss, E. A. (2011). Giving TB wheels: Public transportation as a risk factor for tuberculosis transmission, . *Tuberculosis (Edinburgh, Scotland), 91(1)*, 16–23.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* Chapman and Hall/CRC.

Gesesew, H., Tsehayneh, B., Massa, D., Gebremedhin, A., Kahsay, H., & Mwanri, L. (2016). Predictors of mortality in a cohort of tuberculosis/HIV co-infected patients in Southwest Ethiopia. *Infectious Diseases of Poverty, 5(109)*, DOI 10.1186/s40249016-0202-1.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: KaplanMeier estimate. *International journal of Ayurveda research, 1(4)*, 274.

Gomez-Barroso, D., Rodriguez-Valin, E., Ramis, R., & Cano, R. (2013). Spatio-temporal analysis of tuberculosis in Spain, 2008–2010. *The international journal of tuberculosis*

and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 17(6), 745–751.

Grover, G., Swain, P. K., & Ravi, V. (2014). A Competing Risk Approach with Censoring to Estimate the Probability of Death of HIV/AIDS Patients on Antiretroviral Therapy in the Presence of Covariates. *Statistics Research Letters (SRL), 3*, 7-16.

Haining, R. (2014). Spatial Statistics and the Analysis of Health Data. In M. L. Anthony Gattrell, *In GIS and Health* (pp. 47-66). Boca Raton: CRC Press.

Hazra, A., & Gogtay, N. (2017). Biostatistics Series Module 9: Survival Analysis. *Indian Journal of Dermatology, 62(3)*, 251–257.

James, I. (2014). *Accelerated Failure-time Models.* Wiley StatsRef: Statistics Reference Online.

Jerrett, M., Burnett, R. T., Beckerman, B. S., Turner, M. C., Krewski, D., Thurston, G., . . . Pope III, A. C. (2013). Spatial analysis of air pollution and mortality in California. *American journal of respiratory and critical care medicine, 188(5)*, 593-599.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association, 53(282)*, 457-481.

Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Andrew , G. L., . . . Karim, Q. A. (2011). Integration of antiretroviral therapy with tuberculosis treatment. *New England Journal of Medicine, 365(16)*, 1492-1501.

Karki, L. (2017). *Spatial and Spatiotemporal Modeling of Epidemiological Data.* South Dakota: Open PRAIRIE.

Kazembe, L., Clarke, A., & Kandala, N. B. (2012). Childhood mortality in sub-Saharan Africa: cross-sectional insight into small-scale geographical inequalities from Census data. *BMJ open, 2(5)*, e001421.

Khana, D., Lauren, R. M., Holly, H., & Margaret, W. (2018). A Bayesian Spatial and Temporal Modeling Approach To Mapping Geographic Variation in Mortality Rates for Subnation Areas With R-INLA. *HHS Public Access, 16(1)*, 147–182.

Kleinbaum, D. G., & Klein , M. (2015). *Survival Analysis - A Self-Learning Text.* New York: Springer-Verlag.

Kleinbaum, D. G., & Klein, M. (2012). Kaplan-Meier Survival Curves and the Log-Rank Test. In D. G. Kleinbaum, & M. Klein, *In Survival analysis* (pp. 55-96). New York, NY: Springer.

Kleinbaum, D. M., & Klein, M. (2010). *Survival Analysis - A Self-Learning Text.* New York: Springer.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology.* New York: CRC Press.

Lawson, A. B. (2018). *Bayesian Disease: Hierarchical Modeling in Spatial Epidemiology.* Boca Raton: CRC Press.

Maciel, E. L., Pan, W., Dietze, R., Peres, R. L., Vinhas, S. A., & Ribeiro, F. K. (2010). Spatial patterns of pulmonary tuberculosis incidence and their relationship to socioeconomic status in Vitoria, Brazil. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 14(11)*, 1395–1402.

Manda, S. O., Masenyetse, L. J., Lancaster, J. L., & van der, M. L. (2013). Risk of Death among HIV Co-Infected Multidrug Resistant Tuberculosis Patients, Compared To Mortality in the General Population of South Africa. *Journal of AIDS & clinical research, 7*.

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia medica, 23(2)*, 143149.

Mee, P., Collinson, M. A., Madhavan, S., Root, E. D., Tollman, S. M., Byass, P., & Kahn, K. (2014). Evidence for localised HIV related micro–epidemics associated with the decentralised provision of antiretroviral treatment in rural South Africa: a spatio–temporal analysis of changing mortality patterns (2007–2010). *Journal of global health, 4(1)*, 010403.

Menges, W. (2018, Feruary 12). *The Namibian Newspaper*. Retrieved from The Namibian Newspaper Website: https://www.namibian.com.na/174350/archive-read/GlobalFund-cuts-deepen-Namibias-woes

Mfinanga, S. G., Kirenga, B. J., Chanda, D. M., Mutayoba, B., Mthiyane, T., Yimer, G., . . . Onyebujoh, P. (2014). Early versus delayed initiation of highly active antiretroviral therapy for HIV-positive adults with newly diagnosed pulmonary tuberculosis (TBHAART): a prospective, international, randomised, placebo-controlled trial. *The Lancet infectious diseases, 14(7)*, 563-571.

Miller, H. J. (2004). Tobler‟s First Law and Spatial Analysis. *Annals of the Association of American Geographers*, 284-289.

MoHSS. (2017). *National Staratergic Fframework for HIV and AIDS Response in Namibia.* Windhoek: MoHSS.

MoHSS. (2016a). *National Tuberculosis and Leprosy Programme.* Windhoek: MoHSS.

MoHSS. (2017b). *Third Medium Term Strategic Plan for Tuberculosis and Leprosy 2017/18 – 2021/22*. Windhoek: Ministry of Health and Social Services.

MoHSS. (2016c). *Surveillance Report of the 2016 National HIV Sentinel Survey.* Windhoek: MoHSS.

Narasimhan, P., Wood, J., MacIntyre, C. R., & Mathai, D. (2013). Risk Factors for Tuberculosis. *Pulmonary medicine, 2013*.

Ng, I. C., Wen, T. H., Wang, J. Y., & Fang, C. T. (2012). Spatial dependency of tuberculosis incidence in Taiwan. *PloS one, 7(11)*, e50740. doi: 10.1371/journal.pone.0050740.

NSA. (2016). *Namibia Household Income and Expenditure Survey (NHIES) 2015/2016 Report.* Windhoek: Namibia Statistics Agency.

Ntirampeba, D., Kazembe, L. N., & Neema, I. (2017). Modelling spatial patterns of misaligned disease data: An application on measles incidence in Namibia. *Clinical Epidemiology and Global Health, 5(4),* 190-195.

Ntirampeba, D., Neema, I., & Kazembe , L. (2017). Joint spatial modelling of disease risk using multiple sources: An application on HIV prevalence from antenatal sentinel and demographic and health surveys in Namibia. *Global Health Research and Policy, 2(1),* 22.

O'Donnell, M. R., Padayatchi, N., Kvasnovsky, C., Werner, L., Master, I., & Jr, R. C. (2013). Treatment outcomes for extensively drug-resistant tuberculosis and HIV co-infection. *Emerging infectious diseases, 19(3),* 416.

Odone, A., Amadasi, S., White, R. G., Cohen, T., Grant, A. D., & Houben, R. M. (2014). The Impact of Antiretroviral Therapy on Mortality in HIV Positive People during Tuberculosis Treatment: A Systematic Review and Meta-Analysis. *PloS one, 9(11),* e112017.

Oren, E., Koepsell, T., Leroux, B. G., & Mayer, J. (2012). Area-based socio-economic disadvantage and tuberculosis incidence. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 16(7),* 880–885.

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models (Doctoral dissertation).*

Qin, L., Knol, M. J., Corpeleijn, E., & Stolk, R. P. (2010). Does physical activity modify the risk of obesity for type 2 diabetes: a review of epidemiological data. *European journal of epidemiology, 25(1),* 5-12.

Randremanana, R. V., Richard, V., Rakotomanana, F., Sabatier, P., & Bicout, D. J. (2010). Bayesian mapping of pulmonary tuberculosis in Antananarivo, Madagascar. *BMC infectious diseases,* 10-21.

Rosen, S., Maskew, M., Fox, M. P., Nyoni, C., Mongwenyana, C., Malete, G., . . . Long, L. (2016). Initiating Antiretroviral Therapy for HIV at a Patient's First Clinic Visit: The RapIT Randomized Controlled Trial. *PLoS medicine, 13(5),* e1002015.

Sileshi, B., Deyessa, N., Girma, B., Melese, M., & Suarez, P. (2013). Predictors of mortality among TB-HIV Co-infected patients being treated for tuberculosis in Northwest Ethiopia: a retrospective cohort study. *BMC infectious diseases, 13(1),* 297.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4),* 583-639.

StataCorp, L. P. (2013). *Stata Survival Analysis and Epidemiological Tables: Reference Manual Release 13.* Texas: StataCorp LP, College Station.

Suthar, A. B., Lawn, S. D., del Amo, J., Getahun, H., Dye, C., Sculier, D., . . . Granich, R. M. (2012). Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. *PLoS medicine, 9(7),* e1001270.

Tabarsi, P., Saber-Tehrani, A. S., Baghaei, P., Padyab, M., Mansouri, D., Amiri, M., . . . Altice, F. L. (2009). Early initiation of antiretroviral therapy results in decreased morbidity and mortality among patients with TB and HIV. *Journal of the International AIDS Society, 12(1)*, 14.

Tadesse, K., Haile, F., & Hiruy, N. (2014). Predictors of Mortality among Patients Enrolled on Antiretroviral Therapy in Aksum Hospital, Northern Ethiopia: A Retrospective Cohort Study. *PloS one, 9(1)*, e87392.

Tenthani, L., Haas, A. D., Tweya, H., Jahn, A., van Oosterhout, J. J., Chimbwandira, F., . . . Keiser, O. (2014). Retention in care under universal antiretroviral therapy for HIV infected pregnant and breastfeeding women ("Option B+") in Malawi. *AIDS (London, England), 28(4)*, 589.

Therneau, T. M., & Grambasch, P. M. (2014). *Modeling survival data: extending the Cox model.* New York: Springer Science & Business Media.

Therneau, T. M., & Grambsch, P. M. (2013). *Modeling survival data: extending the Cox model.* Berlin: Springer Science & Business Media.

Tsiatis, A. A. (1981). A large sample study of cox's regression model . *The Annals of Statistics, 9(1)*, 93-108.

Twabi, H. S., & Mukaka, M. (2018). Modelling Length of Hospital Stay for Tuberculosis Treated in-Patients at Queen Elizabeth Central Hospital: a Competing Risk Perspective. *Open Access Biostatistics & Bioinformatics, 1(2)*, OABB.000507.

UNAIDS. (2017, November 3). *UNAIDS*. Retrieved from http://www.unaids.org:

http://www.unaids.org/en/topic/tuberculosis

Uthman, O. A., Yahaya, I., Ashfaq, K., & Uthman, M. B. (2009). A trend analysis and sub-regional distribution in number of people living with HIV and dying with TB in Africa, 1991 to 2006. *International journal of health geographics, 8(1)*, 65.

Vijay, S., Kumar, P., Chauhan, L. S., Rao, S. V., & Vaidyanathan, P. (2011). Treatment Outcome and Mortality at One and Half Year Follow-Up of HIV Infected TB Patients Under TB Control Programme in a District of South India. *PLoS One, 6(7)*, e21008.

Waters, N. (2016). Tobler's first law of geography. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1-13.

WHO. (2018, January 30). *World Helath Organization*. Retrieved from World Helath

Organization : http://www.who.int/mediacentre/factsheets/fs104/en/

WHO. (2016a). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach.* Geneva: World Health Organization.

WHO. (2018b, January 30). *World Helath Organization*. Retrieved from World Helath Organization Web site: http://www.who.int/mediacentre/factsheets/fs104/en/ Wienke, A. (2010). *Frailty Models in Survival Analysis.* CRC Press.

Wong, M. K., Yadav, R. P., Nishikiori, N., & Eang, M. T. (2013). The association between household poverty rates and tuberculosis case notification rates in Cambodia, 2010, 4(1). *Western Pacific surveillance and response journal: WPSAR*, 25–33.

Zulu, L. C., Kalipeni, E., & Johannes, E. (2014). Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi. *BMC infectious diseases, 14(1)*, 285.

Appendices

Appendix 1: STATA codes for Survival analysis (Chapter 3)

```
destring patientd, gen(patientid1)

encode agegroup, gen(agegroup1)

tabulate stilloncare

stset Time, failure(Status==1) tabstat age,
stat(N mean sd q) by(stilloncare) tabstat
cd4cat, stat(N mean sd q) by(stilloncare)
tabstat eight, stat(N mean sd q) by(stilloncare)

stset time_y, failure(status==1)

stcox age, nohr
stcox cd4, nohr
stcox weight, nohr

tabulate sex stilloncare, column tabulate
maritalstatus stilloncare, column
tabulate facilitylevel stilloncare, column
tabulate district stilloncare, column
tabulate whostage stilloncare, column
tabulate function stilloncare, column
tabulate agegroup stilloncare, column

recode CD4 0/250=1 250/max=2, gen(CD4Cat) label
define CD4Cat 1 "Less than 200" 2 "More than 200"
label values CD4Cat CD4Cat
tabulate cd4cat stilloncare, column
```

```
sts test CD4Cat, logrank

tabulate BMICategory StillOnCare, column

sts test sex, logrank sts test
maritalstatus, logrank sts
test facilitylevel, logrank sts
test district, logrank sts test
whostage, logrank sts test
function, logrank sts test
agegroup, logrank
sts test cd4cat, logrank

label define CD4Category 1 "Less than 250" 2 "More than 250"
label values CD4Category CD4Category

ltable _t _d if _st==1,
interval(1.5,2,2.5,3,3.5,4,4.5,5,5.5,6,6.5,7,7.5,8,8.8,9,9.5,10,10.5,11,11.5,12,12.5,13,13.5,14,
14.5,15)

stcox weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat, nohr
tvc(weight sex agegroup facilitylevel district whostage function cd4cat) texp(ln(_t))

quietly stcox weight sex agegroup facilitylevel district whostage function
cd4cat, schoenfeld(sch*) scaledsch(sca*) stphtest, detail
stphtest, plot(weight) msym(oh)
stphtest, plot(sex) msym(oh)
stphtest, plot(agegroup) msym(oh)
stphtest, plot(facilitylevel) msym(oh)
stphtest, plot(district) msym(oh)
stphtest, plot(whostage) msym(oh)
stphtest, plot(function) msym(oh)
stphtest, plot(whostage) msym(oh)
stphtest, plot(cd4cat) msym(oh)

stphplot, by(treat) plot1(msym(oh)) plot2(msym(th))
stphplot, by(site) plot1(msym(oh)) plot2(msym(th))

drop sch1-sch5 sca1-sca5

 quietly stcox weight sex agegroup facilitylevel district whostage function cd4cat, nohr
mgale(mg)
predict cs, csnell

stset cs, failure(_d)
sts generate H = na
```

```
line H cs cs, sort xlab(0 1 to 3) ylab(0 1 to 3)
drop mg

stcox weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat, hr
strata(function)

stcox weight stcox
i.sex stcox
i.agegroup stcox
i.facilitylevel
stcox i.district
stcox i.whostage
stcox i.function
stcox i.cd4cat
stcox i.function

test 2.district 3.district 4.district

stset time_y, id(patientid) failure(stilloncare=2)

stcrr weight sex cd4cat facilitylevel district whostage function stilloncare,

compete(stilloncare==3 5) noshow nolog stcurve,
cif at1(stilloncare = 3) at2(stilloncare = 5)

stcrr weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat,
compete(stilloncare==3) noshow nolog

stcrreg weight sex agegroup facilitylevel district whostage function cd4cat ,

compete(stilloncare == 3) tvc( weight sex agegroup facilitylevel district whostage function
cd4cat) noshr

stcrr weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat,
compete(stilloncare==5) noshow nolog

stcrreg weight sex agegroup facilitylevel district whostage function cd4cat ,

compete(stilloncare == 5) tvc( weight sex agegroup facilitylevel district whostage function
cd4cat) noshr

stcrr weight sex cd4cat facilitylevel district whostage function, compete(stilloncare==5)
noshow nolog

stcrr weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat,
compete(stilloncare==3) strata(ambulatory) noshow nolog

stcrreg weight i.sex i.agegroup i.facilitylevel i.district i.whostage i.function i.cd4cat,
compete(stilloncare==3) strata(function)
```

Appendix 2: R Codes for Spatial Analysis

```
setwd("C:/Users/hst-su/Desktop/MasterData")

library(sp)
library(rgeos)
library(ggplot2)
library(ggmap)
library(dplyr)
library(raster)
library(tmap)
library(leaflet)
library(spatstat)
library(spdep)
library(rgdal)
library(spdep)
library(rgdal)
library(rgeos)
library(latticeExtra)
library(RColorBrewer)
library(gridExtra)
library(Matrix)
library(lattice)

MyData <- read.csv(file="REGIONAL_TB_HIV.csv", header=TRUE, sep=",")

edit(mydata)
install.packages("pec")
library(prodlim)
library(pec)
library(survival)

install.packages("mapview")

cox2 <-
coxph(Surv(Time_Y,Status)~Weight+Sex+AgeGroup+FacilityLevel+District+WHOStage+F
unction+CD4Cat,data=mydata)
library(riskRegression)

CSC(Hist(Time_Y,StillOnCare)~Weight+Sex+AgeGroup+FacilityLevel+District+WHOStag
e+CD4Cat+strata(Function),data=mydata)

# Clear the window for the spatial analysis

#Set working directory
```

```
setwd("C:/Users/hst-su/Desktop/MasterData")

# Load in data
DeathData <- read.csv(file="TB_HIV_DEATH.csv", header=TRUE, sep=",")

# Load in shapefile

Erongomap <- readOGR(dsn = ".", layer= "Erongo")
plot(Erongomap, col = "gray")

# first, we have to isolate the coordinates of death of patients and let R know that these are
spatial points

coords <- SpatialPoints(DeathData[,c("Xlong", "Ylat")])

# Now Plot the location of the cordinates of death on the
map plot(Erongomap, col = "gray", axes = FALSE)
plot(coords, pch = 21, bg = "red", cox = 1.0 , add=TRUE)
legend("topleft", title = "", pch = 21, pt.bg = "red", bty =
"n")

# Create neighbourhood structure, adjacency matrix and Weight
Matrix queen.nb = poly2nb(Erongomap) summary(queen.nb)

queen.listw=nb2listw(queen.nb) #convert nb to listw type
listw=queen.listw

# plot neighbourhood
plot(Erongomap, border=gray(.5))

plot(queen.nb, coordinates(Erongomap), add=TRUE)

# Read in data

data1 <- read.csv(file="EA_Deaths.csv", header=TRUE, sep=",")

# Merge the two data files

data2 <- merge(Erongomap, data1, by='ID2')

# Computing moran's I in R for spatial Data

mi <- moran.test(data1$no_death, listw = nb2listw(queen.nb))
mi

# Plot Moran I scatter plot, Moran I (local), and probability of most significant Moran I

(Chapter 4, Figure 16)
```

```
moran.plot(data1$no_death, listw = queen.listw, xlab="Number of death", ylab="Spatially lagged death cases")

locm <- localmoran(data1$no_death, listw = nb2listw(queen.nb))
summary(locm)

data1$no_death <- scale(data1$no_death)
data1$lag_death <- lag.listw(listw, data1$no_death)

summary(data1$no_death)

summary(data1$lag_death)

data2$quad_sig <- NA

data2@data[(data2$no_death >= 0 & data1$lag_death >= 0) & (locm[, 5] <= 0.05),

"quad_sig"] <- 1

data2@data[(data2$no_death <= 0 & data1$lag_death <= 0) & (locm[, 5] <= 0.05),

"quad_sig"] <- 2

data2@data[(data2$no_death >= 0 & data1$lag_death <= 0) & (locm[, 5] <= 0.05),

"quad_sig"] <- 3

data2@data[(data2$no_death >= 0 & data1$lag_death <= 0) & (locm[, 5] <= 0.05),

"quad_sig"] <- 4

data2@data[(data2$no_death <= 0 & data1$lag_death >= 0) & (locm[, 5] <= 0.05),
"quad_sig"] <- 5

breaks <- seq(1, 5, 1)

labels <- c("High-High", "Low-Low", "High-Low", "Low-High", "Not Signif.")
np <- findInterval(data2$quad_sig, breaks)

colors <- c("red", "blue", "lightpink", "skyblue2", "white")

plot(data2, col = colors[np])

legend("bottomright", legend = labels, fill = colors, bty = "n")
```

Appendix 3: BayesX Codes for Spatial temporal modelling (Chapter 4)

```
defaultpath = c:\book

%Dataseb object

dataset tbhiv

%Read in dataset into dataset object in object called tbhiv
tbhiv.infile using C:\book\tbhiv.txt

%Describe the dataset and view the dataset
tbhiv.describe

%loading in the map object
map m

m.infile using C:\book\erongo.csv

m.reorder

%Plot the map object and view the map
m.describe

m.outfile, replace graph using c:\book\erongosort.bnd

bayesreg b

b.outfile=c:\book\b

b.regress status = timey(baseline) + agecat + weigt(psplinerw2) + cd4cat + maritalstatus +
faclev+ district + whostage + function, iterations=12000 burnin=2000 step=10 family=cox
predict using tbhiv

bayesreg b1

b.outfile=c:\book\b1 b1.regress status = timey(baseline) + agecat + weigt(psplinerw2) +
cd4cat + maritalstatus + faclev+ district + whostage + function + eacod(spatial,map=m),
iterations=12000 burnin=2000 step=10 family=cox predict using tbhiv

bayesreg b2 b2.outfile=c:\book\b2 b2.regress status = timey(baseline) + agecat +
weigt(psplinerw2) + cd4cat + maritalstatus + faclev+ district + whostage + function +
eacod(random), iterations=12000 burnin=2000 step=10 family=cox predict using tbhiv

bayesreg b3
b3.outfile=c:\book\b3

b3.regress status = timey(baseline) + agecat + weigt(psplinerw2) + cd4cat + maritalstatus +
faclev + district + whostage + function + eacod(spatial,map=m) + eacod(random),
iterations=12000 burnin=2000 step=10 family=cox predict using tbhiv
```

```
dataset res graph g res.infile using
c:\book\b3_f_eacod_spatial.res g.drawmap
pmean eacod, map=m using res

dataset res1
res1.infile using c:\book\b3_f_eacod_random.res
graph g1

g1.drawmap pmean eacod, map=m using res1
```

Spatial temporal codes

```
bayesreg r

r.outfile=c:\temp\r

r.regress status = timey(baseline) + eacod(spatial,map=m), iterations=12000 burnin=2000
step=10 family=cox predict using tbhiv

bayesreg r1 r1.outfile=c:\temp\r1 r1.regress status = timey(baseline) + sex +
agecat + weigt(psplinerw2) + cd4cat + maritalstatus + faclev+ district +
whostage + function + eacod(spatial,map=m), iterations=12000 burnin=2000
step=10 family=cox predict using tbhiv

bayesreg r2
r2.outfile=c:\temp\r2

r2.regress status = timey(baseline) + timey * eacod(spatial,map=m), iterations=12000
burnin=2000 step=10 family=cox predict using tbhiv

bayesreg r3
r3.outfile=c:\temp\r3

r3.regress status = timey(baseline) + timey * eacod(spatial,map=m) + eacod(random),
iterations=12000 burnin=2000 step=10 family=cox predict using tbhiv

bayesreg r4 r4.outfile=c:\temp\r4 r4.regress status = timey(baseline) + sex + agecat +
weigt(psplinerw2) + cd4cat + maritalstatus + faclev+ district + whostage + function +
timey*eacod(spatial,map=m), iterations=12000 burnin=2000 step=10 family=cox
predict using tbhiv

bayesreg r5 r5.outfile=c:\temp\r5 r5.regress status = timey(baseline) + sex + agecat +
weigt(psplinerw2) + cd4cat + maritalstatus + faclev+ district + whostage + function +
timey*eacod(spatial,map=m) + eacod(random), iterations=12000 burnin=2000
step=10 family=cox predict using tbhiv
 # Spatial and temporal variability
%Dataseb object
```

```
dataset tbhiv2017

%Read in dataset into dataset object in object called tbhiv
tbhiv2017.infile using c:\temp\tbhiv2017.txt

%loading in the map object
map m

m.infile using c:\temp\erongo.csv

m.reorder

bayesreg r6
r6.outfile=c:\temp\r6

r6.regress status = timey(baseline) + sex + agecat + weigt(psplinerw2) + cd4cat +
maritalstatus + faclev+ district + whostage + function + eacod(spatial,map=m),
iterations=12000 burnin=2000 step=10 family=cox predict using tbhiv2017 dataset res
res.infile using c:\temp\r6_f_eacod_spatial.res
graph g

g.drawmap pmean eacod, map=m using res

dataset res1

res1.infile using c:\temp\r6_f_eacod_random.res
graph g1

g1.drawmap pmean eacod, map=m using res1
```