ISSN: 2410-1397

APPLICATION OF COX PROPORTIONAL HAZARDS MODEL IN CASE OF TUBERCULOSIS PATIENTS IN KISUMU COUNTY, KENYA

MASTER OF SCIENCE IN SOCIAL STATISTICS

MIRIAM TABITHA ABONGO

Supervisor:

Dr. Nelson Owuor

Supervisor:

Dr. Elias Obudho

Research Reports in Mathematics

Number 34, 2015

School of Mathematics

College of Biological and Physical Sciences

UNIVERSITY OF NAIROBI

DECLARATION

This research project is an original work and has never been presented for examination at any other learning institution in Kenya or elsewhere.

STUDENT

MIRIAM TABITHA ABONGO

SIGN: ---------- **DATE: ----------**

This project has been submitted for examination with the approval of the following as the University supervisors.

SUPERVISOR

DR. NELSON OWUOR

SIGN: ---------- **DATE: ----------**

SUPERVISOR

DR. ELIAS OBUDHO

SIGN: ---------- **DATE: ----------**

ACKNOWLEDGEMENTS

My most wholehearted gratitude to the almighty God for his favor, grace and love that has seen me thorough my entire Masters course and the completion of this research work; without God I would not have made it. My sincere gratitude to my research supervisors; Dr. Nelson Owuor and Dr. Elius Obudho for their guidance and technical support during the entire course of the project. I am indebted to whole heartedly appreciate the support from my family; Jared and Shammah thank you very much for your selfless support throughout my entire course, may God richly bless you.

DEDICATION

I would like to dedicate this piece of work to my son Shammah and to my husband Dr. J. Abongo.

TABLE OF CONTENT

DECLARATION ii
ACKNOWLEDGEMENTS iii
DEDICATION iv
TABLE OF CONTENT v
LIST OF TABLES vii
LIST OF FIGURES viii
ABSTRACT ix
CHAPTER ONE: INTRODUCTION 1
1.1 Back ground Information 1
1.2 Objectives of the study 2
1.3 Problem statement 3
1.4 Justification of the study 4
CHAPTER TWO: LITERATURE REVIEW 6
2.1 What is tuberculosis (TB)? 6
2.2 Tuberculosis transmission 6
2.3 Clinical forms of Tuberculosis 7
2.4 Dangers of Tuberculosis 8
2.5 Tuberculosis treatment 8
2.6 DOTS (Directly Observed Treatment, Short-Course) 9
2.7 Tuberculosis patient classification 10
2.8 Review of previous studies 11
CHAPTER THREE: METHODOLOGY 15
3.1 Study area 15
3.2 Data and Sampling design 16
3.3 Ethical issues 17
3.4. Study variables 17
3.5 Survival analysis 18
3.6 Quantities in survival analysis 19
3.7 Comparing survival functions of two groups of TB patients 20

3.7.1 Kaplan-Meier product limit estimate 20
3.7.2 The Nelson-Aalen estimator 22
3.8 Using Log rank test to compare survival functions of two groups of TB patients (Male and Female) 24
3.9 Application of Cox Proportional Hazards regression model in identification of significant predictors of loss to follow-up among Tuberculosis patients 27
3.9.1 The proportionality of hazards assumption 29
3.9.2 Estimation of the Cox proportional hazards model parameters 31
3.9.3 Testing the significance of the parameters 33
3.9.3b Variable selection 34
3.10 Overall goodness of the model 35
3.11 Interpretation of the results 36
CHAPTER FOUR: DATA ANALYSIS AND RESULTS 37
4.1 Summary Statistics 37
4.2 Time to Tuberculosis treatment interruption 38
4.3 Log rank and Wilcoxon test of no significance on the Tuberculosis patients factors 40
4.4 Results of the proportional hazards Cox regression model of TB patients 43
With all variables 43
4.5 Proportional Hazards assumption test 44
4.6 Final Model with the significant variables (age, weight, patient type) 46
CHAPTER 5: CONCLUSION AND THE RECOMMENDATIONS 47
5.1 Conclusion 47
5.2 Recommendations 47
REFERENCES 48

LIST OF TABLES

Table 1: Tuberculosis patient classification, (WHO, 2013) 10

Table 2: Tuberculosis cases for top 10 counties in Kenya - 2013 17

Table 3: Kaplan-Meir Estimator table 22

Table 4:Nelson-Aalen Estimator table 23

Table 5:Log rank test 2*2 contingency table 25

Table 6:Tuberculosis Patient characteristics..... 38

Table 7:Summary statistics of status of TB patients and months of follow-up time..... 39

Table 8:Log rank and Wilcoxon test of no significance..... 40

Table 9: Results of the proportional hazards Cox regression model of TB patients (All variables 44

Table 10: Results of proportionality assumption test containing time dependent covariates. 45

Table 11:Linear hypothesis test results 45

Table 12:Final Model with significant covariates..... 46

LIST OF FIGURES

Figure 1: Kisumu County Map 16

Figure 2:Comparison of Kaplan-Meier survival curves by age category 41

Figure 3: Comparison of Kaplan-Meier survival curves by type of patient category 42

Figure 4: Comparison of survival functions by Gender 43

ABSTRACT

Tuberculosis (TB) remains a major global health problem. In 2013, World Health Organization estimated 8.6 million people to have developed TB and 1.3 million died from the disease including 320 000 deaths among HIV-positive people. Kenya is ranked 15th up from 13th position among the 22 high TB burden countries worldwide which contributes 80% of the global TB. The main objective of this study was to describe the pattern of time of treatment interruption in TB patients in Kisumu County Kenya. The specific objectives were to compare Survival functions for two groups of Tuberculosis patients and identify significant predictors of loss to follow-up (LTFU) among the TB patients. This was a retrospective cohort study based on TB patients that were registered in the unit TB registers. Kisumu County was randomly sampled for the study from the top ten high TB burden counties in Kenya; Secondary data was collected from documents of all TB cases registered from January to December 2013 in the health facilities. This study focused on time of loss to follow-up for TB patients enrolled on treatment in the health facilities. Kaplan-Meir estimator and Cox proportional hazard model have been used for the analysis and model building. Logrank tests and Wilcoxon tests were used for comparison of survival data. Cox model was able to provide the estimates covariates and their effect on loss to follow-up of Tuberculosis patients after adjustment for other explanatory variables. From the 1,275 patients in this study a total of 107(8.4%) interrupted treatment and 1,168 (91.6%) were censored. The median time of treatment interruption was 2.167 months (65 days).This analysis shows that most of the Tuberculosis patients are lost to follow up during the intensive phase of treatment which is within the first two months. The Log rank and Wilcoxon test were not significant in LTFU survival experience between the various categories for covariates HIV status, weight, gender and type of TB at alpha significance level of 5% ($p<0.05$). We find significant differences in survival experience of the patients in different categories of age; patients with more than 35 years of age had better survival rates compared to those who were younger. Results of the proportional hazards Cox regression analysis of TB patients revealed that the covariates age, weight and Tuberculosis patient type were significant factors associated with treatment interruption for TB patients. There is need to strengthen follow up of patients on TB treatment especially during the intensive phase and until completion of treatment.

CHAPTER ONE: INTRODUCTION

This chapter gives the background information on Tuberculosis, the objectives of the study, the problem statement and justification of the study.

1.1 Back ground Information

Tuberculosis (TB) remains a major global health problem. In 2012, an estimated 8.6 million people developed TB and 1.3 million died from the disease including 320 000 deaths among HIV-positive people. (World Health Organization, 2013). According to the WHO, 2012 annual TB report, there were 8.7 million range (8.3-9.0 million) incident cases in 2011 including 1.1M cases among PLHIVs. Over the same year 6.2 million notified cases of TB against an estimated 8.7 million cases, representing a case detection rate (CDR) of 66%.

The World Health Organization (WHO) estimates that the average incidence of tuberculosis in African countries to be more than doubled between 1990 and 2005 from 149 to 343 per 100,000 population. (Richard *et .al*, 2008).

Kenya is ranked 15th up from 13th position among the 22 high TB burden countries worldwide which contributes 80% of the global TB (World Health Organization, 2013) .Tuberculosis remains a major cause of morbidity and mortality in Kenya. It affects all age groups, but has its greatest toll in the most productive age group of 15 to 44 years. The major factor responsible for the large TB disease burden in Kenya is the concurrent HIV epidemic. Other factors that have contributed to this large TB disease burden include poverty and social deprivation that have led to mushrooming of peri-urban slums, congestion and limited access to general health services. Recently, there have been increasing concerns about the emergence of drug resistant TB, a threat

that would pose major challenges in the fight against TB in resource limited countries like Kenya (MOH, 2013).

The Government of Kenya (GOK) is implementing all the six components of STOP TB strategy which include: Pursue high-quality DOTS expansion and enhancement, Address TB-HIV, MDR-TB, and the needs of poor and vulnerable populations, Contribute to health system strengthening based on primary health care, Engage all care providers, Empower people with TB, and communities through partnership and finally Enable and promote research. During the year 2013, a total of 89,760 patients were notified in the country.

In Kisumu the total cases notified in the county in 2013 were 3,362 giving a case notification rate for TB at 317 per 100,000 which is above the national average of 217 per 100,000. The high burden of tuberculosis in Kisumu County has mainly been attributed to the high HIV prevalence (MOH, Annual Report 2013).

1.2 Objectives of the study

The main objective of this study was to describe the pattern of time of treatment interruption in TB patients in Kisumu County.

The specific objectives are:

1. To compare Survival functions for two groups of Tuberculosis patients
2. Identify significant predictors of loss to follow-up among the TB patients
3. Make recommendations on appropriate strategies for reducing loss to follow-up and propose further studies on predictors of loss to follow-up.

1.3 Problem statement

Tuberculosis (TB) is second only to HIV/AIDS as the greatest killer worldwide due to a single infectious agent. In 2012, 8.6 million people fell ill with TB and 1.3 million died from TB and over 95% of TB deaths occur in low- and middle-income countries, and it is among the top three causes of death for women aged 15 to 44. In 2012, an estimated 530 000 children became ill with TB and 74 000 HIV-negative children died of TB. TB is a leading killer of people living with HIV causing one fifth of all deaths.(World Health Organization, 2012) .

Successful treatment of TB involves taking anti-TB drugs for 2 months during intensive phase and 4-6 months in the continuation phase but many find it difficult to complete their course of treatment and this serves as a major constraint to eradicating the disease (Cuneo, 1989). WHO defines defaulters as Interruption of treatment for two consecutive months? Defaulting from treatment means patients remain infectious for longer and are more likely to relapse and die. Default from treatment can also lead to development of MDR TB that is expensive to treat and can develop into X-DR T.B which is fatal. It costs approximately 2 million kshs to treat one patient of MDR T.B and will overstretch the already stretched health services in Kenya thus there is need for the country to understand and address factors contributing to default and treatment interruption especially in high T.B burden areas like western Kenya due to high prevalence of also HIV and AIDS in the region.

The ability of African health care systems to respond to, manage, and contain the growing number of cases of tuberculosis is constrained by limitations of funding, facilities, personnel, drug supplies, and laboratory capacity. Although the President's Emergency Plan for AIDS Relief (PEPFAR) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria have donated

large sums of money to help address Africa's health problems, most of the money has been earmarked for HIV, with a lesser focus on tuberculosis. (Richard et. al, 2008)

Loss to follow-up is one of the unfavorable outcomes for TB patients and represents an important challenge for the control programs (Ministry of Health, 2013). There are various causes and risk factors for LTFU. These include gender, alcoholism, treatment interruption, poor knowledge of tuberculosis, irregular treatment and socioeconomic status. Other factors related to the disease, patients and service providers have also been identified as reasons for no completion of treatment yet it has remained difficult to predict non-adherence.

During the year 2013, a total of 89,760 patients were notified (all forms of tuberculosis) in the Kenya. In Kisumu County the total cases notified in the county in 2013 were 3362 giving a case notification rate for TB is at 317 per 100,000 which are above the national average of 217 per 100,000.This study focuses on a key factor in eradication of Tuberculosis, the time of treatment LTFU.

1.4 Justification of the study

Up to half of the people with tuberculosis do not complete their treatment (WHO, 2011). Strategies to improve adherence to diagnostic and treatment regimens are therefore important. One barrier to TB control is default, defined as treatment interruption for at least two consecutive months (Global Tuberculosis Control, WHO, 2011).

Poor treatment adherence increasing the risk of drug resistance, treatment failures, relapses, deaths and prolonged infectiousness remains a hurdle to the success of tuberculosis programmes (Snider Jr. DE (1982). A major contributor to both treatment failure and the rise of multidrug-

resistant TB is inadequate and incomplete treatment (Sharma & Mohan 2006).While structural factors such as interruptions in drug supply play a role, patient treatment interruption or drop-out from TB treatment is one of the most important reasons for not completing treatment (Borgdorff *et al*. 2002).

This study will contribute information on the likely time to treatment LTFU and therefore support programs in planning appropriate control interventions around the time when TB patients interrupt treatment to reduce the LTFU rate. This project will help in better management of TB patients by understanding what possess as a barrier in treatment compliance and measures will be put in place to manage those challenges to reduce the level of TB transmission and treatment failure and resistance. The study will also contribute to fill up the gap of knowledge on TB LTFU which can be replicated elsewhere.

CHAPTER TWO: LITERATURE REVIEW

2.1 What is tuberculosis (TB)?

Tuberculosis (TB) is an infectious disease caused by bacteria whose scientific name is Mycobacterium tuberculosis. It was first isolated in 1882 by a German physician named Robert Koch who received the Nobel Prize for this discovery. TB most commonly affects the lungs but also can involve almost any organ of the body (Haydel, 2003).

Tuberculosis is curable and preventable and is spread from person to person through the air. When people with lung TB cough, sneeze or spit, they propel the TB germs into the air. A person needs to inhale only a few of these germs to become infected. About one-third of the world's population has latent TB, which means people have been infected by TB bacteria but are not (yet) ill with disease and cannot transmit the disease (World Health Organization, 2014)

A definite case of TB or one in which a health worker (clinician or other medical practitioner) has diagnosed TB and has decided to treat the patient with a full course of TB treatment. Any person given treatment for TB should be recorded as a case. (World Health Organization, 2011)

2.2 Tuberculosis transmission

Tuberculosis is transmitted from an infected person to a susceptible person in airborne particles, called droplet nuclei. These are 1–5 microns in diameter. These infectious droplet nuclei are tiny water droplets with the bacteria that are released when persons who have pulmonary or laryngeal tuberculosis cough, sneeze, laugh, shout etc. These tiny droplet nuclei remain suspended in the air for up to several hours. Tuberculosis bacteria, however are transmitted through the air, not by surface contact. This means touching cannot spread the infection unless it is breathed in. (Dannenberg, 1982)

Tuberculosis is caused by a bacterium called Mycobacterium tuberculosis. The bacteria usually attack the lungs, but TB bacteria can attack any part of the body such as the kidney, spine, and brain. If not treated properly, TB disease can be fatal. Not everyone infected with TB bacteria becomes sick. As a result, two TB-related conditions exist: latent TB infection and TB disease. Both latent TB infection and TB disease can be treated (Centers for Disease Control and Prevention, Tuberculosis, 2012).

If not diagnosed and treated promptly, tuberculosis may be spread via an airborne route to family and community members. Untreated, someone with active tuberculosis will infect an estimated 10 to 15 people per year (Joint United Nations Programme on HIV/AIDS, 2012)

2.3 Clinical forms of Tuberculosis

TB occurs in two clinical forms; Tuberculosis in the lungs which is referred to as Pulmonary TB (PTB) and Tuberculosis outside the lungs known as the Extra pulmonary TB (EPTB). Pulmonary TB is further classified on the basis of smear microscopy into; Smear positive PTB and Smear Negative PTB(. (World Health Organization, 2013)

Smear positive PTB is a patient with at least one initial sputum smear examination positive for acid fast bacilli.

Smear negative PTB refers to at least two sputum specimens at the start of treatment are negative for AFB, Radiologic abnormalities consistent with active pulmonary tuberculosis, No response to a course of broad spectrum antibiotics (excluding anti-TB drugs and fluoroquinolones and aminoglycosides especially in HIV negative patients. PTB is the most common form of TB accounting for 80% of the total cases and is of public health concern as it serves as the source of infection (Ministry Of Health , 2013).

Extra-pulmonary TB affects other parts of the body outside the lung tissue. Body organs affected by EPTB include: lymph nodes, bones, spine, kidneys, liver, bladder, skin, eyes, and gastrointestinal system. EPTB can occur in any organ of the body (except nails, hair and teeth) (World Health Organization, 2013)

2.4 Dangers of Tuberculosis

Tuberculosis (TB) is a bacterial disease that almost invariably infects the lungs, but can also spread to almost any organ system to cause disease and symptoms; this is the reason why screening and treatment are pursued vigorously, so much so that legal mandates have been passed with different requirements depending on a person's TB status. (Saad, 2012)

2.5 Tuberculosis treatment

Directly observed therapy short-course (DOTS) is the most cost effective TB treatment currently. It consists of 6-8 months regimen; an intensive 2 months phase, and continuation phase of 4-6 months. First-line anti-TB drugs include Isoniazid (INH), Rifampicin (RIF), streptomycin (STR), Ethabutol (EMB), Pyrazinamide (PZA) and Thiazetazone (TAZ) (ATS-CDC, 1994). For the re-treatment of TB and the treatment of known or suspected drug resistant TB cases 2^{nd} line anti – TB drugs are used. This includes para-aminosalycylic acid (PAS), ethionaminide, kanamycin, cepreomycin viomycin cycserine and flouroquinolones (Johnson *et al*., 1999).

The following WHO recommended regimes continue to be used in Kenya: 2RHZE/4RH for new cases with smear-positive PTB, smear negative PTB and extra-pulmonary TB; 2SRHZE/1RHZE/5RHE (re-treatment regimen) for smear positive relapse cases, recurrent negative PTB/EPTB cases, failures and treatment interruption and2RHZ/4RH for new cases of

smear positive or negative PTB or EPTB who are younger than 15 years. (World Health Organization, 2013).

Regardless of the treatment regimen, one morning sputum specimen should collected for follow-up at the end of the intensive phase of treatment to determine whether the patient can proceed to the continuation phase if the smear is negative or, if the smear is positive, continue the intensive phase. Another sputum specimen must be taken during the continuation phase to check patient evolution and to detect possible treatment failure, and another upon completion of chemotherapy to verify cure (World Health Organization, 2014)

There are three TB treatment outcomes: cure, treatment failure (TF), or relapse (R). A relapse is a smear positive TB patient who has previously been treated and declared cured, the patient gets better while on treatment and only to get worse again on stopping treatment after completion of treatment course. Treatment failure is a patient with a positive smear at the end of five months despite being on anti TB treatment (World Health Organization, 2013)

2.6 DOTS (Directly Observed Treatment, Short-Course)

Directly observed treatment, short-course), is the name given to the tuberculosis control strategy recommended by the World Health Organization (WHO/TB/97.220. 1997) .According to WHO, "The most cost-effective way to stop the spread of TB in communities with a high incidence is by curing it. The best curative method for TB is known as DOTS"Tuberculosis. (World Health Organization, 2014).

DOTS have five main components:

- Government commitment (including political will at all levels, and establishment of a centralized and prioritized system of TB monitoring, recording and training).

- Case detection by sputum smears microscopy.
- Standardized treatment regimen directly of six to eight months observed by a healthcare worker or community health worker for at least the first two months.
- A regular, uninterrupted drug supply.
- A standardized recording and reporting system that allows assessment of treatment results.

2.7 Tuberculosis patient classification

Type of patient	Case definition
New	A patient who has never had treatment for TB, or has been on anti-TB treatment for less than four weeks.
Relapse	A patient who has been declared cured or treatment completed for any form of TB in the past, but who reports back to the health service and is found to be sputum smear-positive or culture positive.
Treatment after previous treatment failure	A patient who, while on treatment remained sputum smear-positive or became sputum smear-positive at the end of the five months or more, after commencing treatment.
Treatment after treatment interruption (did not complete previous treatment)	A patient who had previously been recorded as treatment interruption from treatment and returns to the health service with smear-positive sputum.
Transfer in	A patient who is transferred from another district to continue treatment.
Other	A patient who does not fit into any of the above categories.
Chronic case	A patient who is still sputum smear-positive at the completion of a re-treatment regimen.

Table 1: Tuberculosis patient classification, (WHO, 2013)

2.8 Review of previous studies

A study on treatment default and duration associated with treatment interruption among new patients with tuberculosis was done in six regions of Russia. (Jakubowiak, Bogorodskaya, Borisov, Danilova, & Kourbatova, 2009).The objectives were to determine the frequency and length of treatment interruptions among new pulmonary tuberculosis (TB) patients and to evaluate the duration of interruption associated with treatment interruption in the tuberculosis services of six Russian regions. This was a retrospective study of all adult patients with new pulmonary TB enrolled for treatment from April 1 to September 30, 2003. Data from patients with treatment outcomes of treatment interruption (n=84), failure (n=130), death (n=113), and success (n=1444) were analyzed. This study revealed that 63% of patients had treatment interruption and 36% of those successfully treated had interruptions of treatment during the intensive phase and 45% of those with a successful outcome had interrupted treatment during the continuation phase. The length of treatment interruptions was 1-125 days during the intensive phase and 1-127 days during the continuation phase among patients with outcomes other than treatment interruption.

According to (Afutu *et al*, 2012) in his study on high initial treatment interruption in patients with smear-positive pulmonary tuberculosis in Ridge Hospital in Accra, Ghana. He identified that there were 84 laboratory confirmed TB cases in 2009, of whom 32 (38%) were initial defaulters. Cure and treatment interruption rates based on this cohort were 54% and 43% respectively, compared with rates of 87% and 8% when using the cohort based on 52 patients registered for treatment. This study showed that programme performance may be poor when patients in laboratory registers are used as the cohort to evaluate treatment outcomes a gap that

the study on application of Cox Proportional Hazards Model in Case of Tuberculosis Patients will address.

The Prevalence and predictors of treatment interruption with tuberculosis treatment study in Sri Lanka by Pinidiyapathirage J *et.al* identified that out of the 892 patients recruited, 770 were new cases and 122 were relapses. The 95% confidence interval for treatment interruption rates reported were between 8.3%-12.6% among new cases and between 22.7% to 38.1% among the retreatment cases. Ninety percent of new cases and 94% of retreatment cases were sputum positive for acid-fast bacilli at diagnosis. This study did not focus on the period within which majority of the patients were lost to follow up a gap that this research seeks to fill.

A study on determinants of treatment interruption from Pulmonary Tuberculosis treatment in Kuwait ; Out of the 954 patients who were treated for tuberculosis in the study period, 110 (11.5%) refused treatment or failed to attend the clinic for more than 2consecutive months, Majority of the treatment defaulters were men (76.4%) compared to the women treatment defaulters . There was no significant deference in age or marital status between treatment defaulters and no treatment defaulters. Approximately 56% of patients who treatment defaulted did so within the first 2 months of treatment. Multiple logistic regression analyses were performed to determine the risk factors associated with treatment interruption among those with pulmonary tuberculosis. The important risk factors associated with treatment interruption included male sex.(Zhang, Gaafer, & El Bayoumy, 2014)

A study in Kenya showed that out of 945 treatment defaulters, 22.7% and 20.4% abandoned treatment within first and second months (intensive phase) of treatment respectively. Among 120 treatment defaulters interviewed, 16.7% attributed their treatment interruption to ignorance,

12.5% (15) to traveling away from treatment site, 11.7% (14) to feeling better and 10.8% (13) to side effects. On multivariate analysis, inadequate knowledge on herbal medication use, low income, alcohol abuse previous treatment interruption, co-infection with Human immune-deficient Virus (HIV) and male gender were independently associated with treatment interruption.(Muture *et al*., 2011)

A study in Morocco for patients with smear- or culture-positive pulmonary tuberculosis presenting for retreatment were identified using clinic registries; to identify factors that had put these individuals at risk for failure, treatment interruption, or early relapse in the first place, initial treatment records were also abstracted. A total 291 patients presenting for retreatment were included; treatment interruption from retreatment was most frequent among patients with initial treatment interruption (57%). Independent risk factors for failure, treatment interruption, or early relapse after initial treatment included male gender and positive sputum smear after 3 months of treatment as well as higher weight at treatment initiation was protective. Male sex, substance use, missed doses, and hospitalization appeared to be risk factors for treatment interruption, but subgroup analyses were limited by small numbers. This study did not focus on the period within which majority of the patients were lost to follow up a gap that this research seeks to fill. (Dooley *et al*., 2011)

A retrospective cohort study based on hospital registers was done in Yaounde, Cameroon. Socio demographic and clinical predictors of treatment discontinuation were investigated with the use of Cox regressions models. Of the 1688 included patients, 337 (20%) treatment defaulters from treatment, 86 (5.1%) died, treatment failed in 6 (0.4%) and 104 (6.2%) were transferred. Median duration to treatment discontinuation was 90 days and 62% of treatment discontinuation occurred during the continuation phase. Hospitalization during the intensive phase (adjusted HR 0.69;

95% CI 0.54 to 0.89) and non-consenting for HIV screening (1.65;1.24 to 2.21) were the main determinants of default from treatment in multivariable analysis. (Pefura Yone *et al*.,2011).

CHAPTER THREE: METHODOLOGY

3.1 Study area

Kisumu County has a total population of 1,059,733 with proportion of males at 49%, urban population stands at 52% and the area is populated at 508 persons per sq. km (SAREM, 2013).

The key health indicators are high burden of HIV at 18.7% HIV prevalence far much higher than the national prevalence at 5.6% (KAIS, 2012). The doctor to population ratio of 1:15000, nurse to population ratio of 1:1433, and health facility to population is 1: 6374, the life expectancy is 40.4 years, Human Development Index (HDI) of 0.4939 and literacy level of 80.3%. The main economic activities are fishing and agriculture (MOH, 2013).

The total cases notified in the county in 2013 were 3,362 giving a case notification rate for TB is at 317 per 100,000 which is above the national average of 217 per 100,000. The TB/HIV co-infection rate is 67% far above the national rate at 37% and this reflects the prevailing high HIV burden in the county (MOH, 2013).

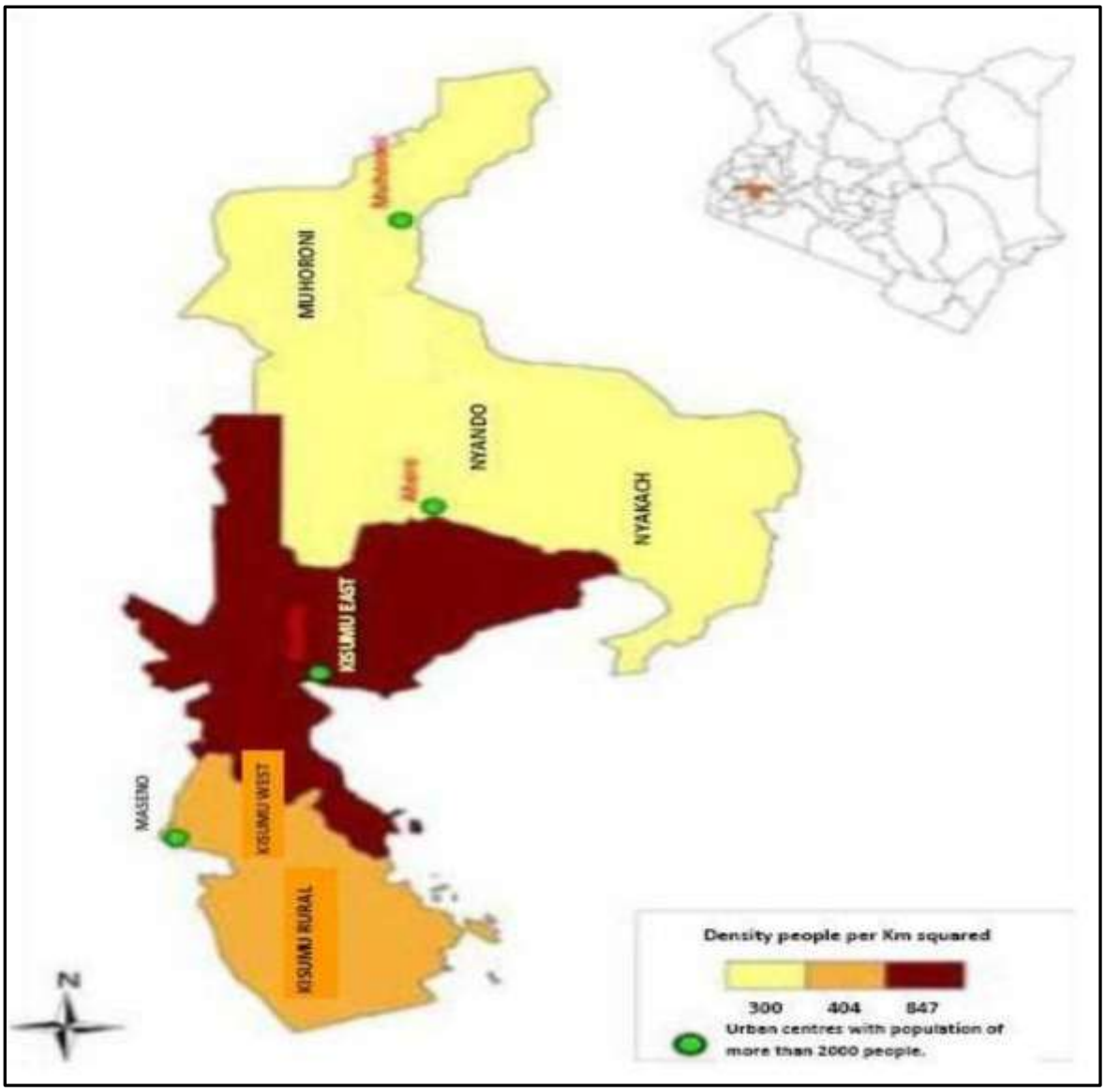

Figure 1: Kisumu County Map

3.2 Data and Sampling design

This was a retrospective cohort study based on TB patients that were registered in the unit TB registers. Kisumu county was randomly sampled for the study from the top ten high TB burden counties in Kenya (table 2) which are Nairobi South, Nairobi North, Mombasa, Nakuru, Kiambu, Homabay, Meru, Kisumu,Migori and Kakamega .Secondary data was collected from documents

of all TB cases registered from January to December 20013 in the health facilities in Kisumu County.

No	Name Of county	TB cases
1	Nairobi	14,401
2	Mombasa	4,726
3	Nakuru	4,130
4	Kiambu	4,010
5	Homabay	3,528
6	Kisumu,	3,362
7	Meru	3,358
8	Migori	2,579
9	Kakamega	2,535
10	Machakos	2,346

Table 2: Tuberculosis cases for top 10 counties in Kenya - 2013

3.3 Ethical issues

The study was retrospective and did not involve any experimental procedures on patients. However, research and ethical clearance to conduct the study was sought and obtained from Nairobi University and a research permit was granted by National Tuberculosis and Lung Disease Unit. TB is strongly associated with HIV/AIDS leading to stigmatization and is difficult to discuss it in public and so confidentiality and anonymity was assured by excluding the names and addresses of patients from the database shared.

3.4. Study variables

Predictor Variables (Covariates) are independent variables also referred as explanatory variables or covariates in the study which included patient gender, age, TB Patients category, type of TB,

initial weight of patients, and HIV status. This study evaluated the effect of the covariates on the time to treatment interruption.

3.5 Survival analysis

The term "survival data" has been used in a broad sense for data involving time to the occurrence of a certain event. This event may be death, the appearance of a tumor, the development of some disease, recurrence of a disease, conception, cessation of smoking, and so forth (Persson, I. 2002)

Truncation is a condition which screens certain subjects so that the investigator will not be aware of their existence, only individuals who meet some condition are observed. (Persson, I. 2002)

Censored data arises when an individual's life length only is known to occur in a certain period of time. Possible types of censoring are right censoring, where all that is known is that the individual is still alive at a given time, left censoring is when all that is known is that the individual experienced the event of interest prior to the start of the study, or interval censoring, where the only information is that the event occurs within some interval of time. One type of right censoring that is very common is Type I censoring, where the event is observed only if it occurs prior to some prespecified time; for example at the closing of a study. A second type of right censoring is Type II censoring in which the study continues until the failure of the first *r* individuals, where *r* is some predetermined integer. Experiments involving Type II censoring are often used in testing of equipment life. Most methods used in survival analysis are proper for right censored data. Truncation is a condition which screens certain subjects so that the investigator will not be aware of their existence, only individuals who meet some condition are observed. Types of truncation are (1) left truncation, where only individuals who survive a

certain time before the study starts are included and (2) right truncation, where only individuals who have experienced the event by a certain time are included in the study.

3.6 Quantities in survival analysis

i. Hazard function

$$\lambda(t) = \lim_{\Delta t \to 0} = \frac{P[t \le T \angle t + \Delta t \mid T \ge t]}{\Delta t} \qquad (3.1)$$

Thus $\lambda(t)\Delta t$ can be seen as the conditional probability that the event of interest occurs in the interval $[t, t+\Delta t)$, given that it has not occurred before time *t*. This function is particularly useful in determining the appropriate failure distributions utilizing qualitative information about the mechanism of failure (Klein, 1997).

ii. The survival function

$$\boldsymbol{S(t) = P(T > t)} \qquad (3.2)$$

A function describing the proportion of individuals surviving to or beyond a given time.

Notation:

$T \equiv$ survival time of a selected individual

$t \equiv$ a specific point in time.

$h(t) \equiv$ instantaneous failure rate among survivors at time *t* (*aka* hazard function

Equivalently, the **hazard function** can also be expressed as a survival function

$$\lambda(t) = \frac{d}{dt}\left(-\ln S(t)\right) \qquad (3.3)$$

4. Mean residual life quantity

For individuals of age *t*, this parameter measures their expected remaining lifetime and is for a continuous random variable defined as

$$mr1(t) = \frac{\int_t^{\infty} (x-t) f(x)dx}{S(t)} = \frac{\int_t^{\infty} S(x)dx}{S(t)} \qquad (3.5)$$

$\mu = mr(0)$, is the total area under the survival curve

$$\mu = E(t) = \int_0^{\alpha} Xf(x)dx = \int_0^{\alpha} S(x)dx \qquad (3.6)$$

The estimator of the mean lifetime, $\hat{\mu}$ is obtained by using the Kaplan-Meier

Estimators (3.4) as an estimator of the survival function $S(x)$ in (3.5). The estimator does not use the information on survival available in times larger than the largest survival time.

3.7 Comparing survival functions of two groups of TB patients

The objective of comparing the survival functions for two groups would be achieved by first estimating their respective survival functions.

The Kaplan-Meir estimator and the Nelson-Aalen estimator ware used to establish the survival functions to achieve this described objective

3.7.1 Kaplan-Meier product limit estimate

The standard estimator of the survival function is the Kaplan-Meier product limit estimate (Kaplan and Meier (1958).

The Kaplan-Meir estimator is the method for estimating the Survival function ***S(t)***

As demonstrated in equation 3.2;

Survival function; $S(t) = p(T > t)$ is the probability of survival beyond time t which means the time to event (T) occurring past time t.

Suppose we denote the rank-ordered failure times as $t_1 < t_2 \ldots < t_k$

Let;

r_j, be the number of participants at risk of event of interest (loss to follow-up) at time t_j.

d_j, be the number of events of interest (loss to follow-up) observed at time tj.

The probability of surviving past time t given survival to that time is estimated by

$$\frac{(r_j - d_j)}{r_j}$$

Therefore for the overall probability of surviving beyond any time t is estimated by the equation;

$$\hat{s}(t) = \prod_{t_1 \le t} \frac{(r_j - d_j)}{r_j} \qquad (3.7)$$

Equation 3.7 is called the **Kaplan-Meir estimator** where $j = 1,2 \ldots k$

One needs to construct a table to be able to obtain the Kaplan-Meir estimator, this is shown in table (3) with the survival function estimate.

j	t_j	d_j	c_j	r_j	$(r_j - d_j)$	$\frac{(r_j - d_j)}{r_j}$	$\hat{s}(t)$
0	t_0	d_0	c_0	r_0	$(r_0 - d_0)$	$\frac{(r_0 - d_0)}{r_0}$	$\frac{(r_0 - d_0)}{r_0}$
1	t_1	d_1	c_1	r_1	$(r_1 - d_1)$	$\frac{(r_1 - d_1)}{r_1}$	$\frac{(r_0 - d_0)}{r_0} * \frac{(r_1 - d_1)}{r_1}$
.							
.							
.							
k	t_k	d_k	c_k	r_k	$(r_k - d_k)$	$\frac{(r_k - d_k)}{r_k}$	$\frac{(r_0 - d_0)}{r_0} * \frac{(r_1 - d_1)}{r_1} * \ldots * \frac{(r_k - d_k)}{r_k}$

Table 3: Kaplan-Meir Estimator table

Intervals defined by censored observations are not considered, the formula (3.7) only uses the points at which the value of the estimator changes. Thus, the **Kaplan-Meier estimator is a step function** with jumps at the observed event times. The size of these jumps depends not only on the number of events observed at each event time t_j, but also on the pattern of the censored observations prior to *tj*.

3.7.2 The Nelson-Aalen estimator

This is the method for estimating the Cumulative hazard function $H(t)$

The relationship between the cumulative **hazard function** $H(t)$ **and the survival function *S*(*t*)** can be illustrated as ,

$$S(t) = \exp(\int_0^t h(u)du$$

Where $h(u)$ is the hazard function?

$S(t) = \exp(-H(t))$

$S(t) = \Pr(T > t)$ is the probability of survival beyond time t; this means the time to event (T) occurring after time t.

Let;

r_j, be the number of participants at risk of event of interest (loss to follow-up) at time tj .

d_j, be the number of events of interest (loss to follow-up) observed at time tj .

By description the estimate of the cumulative hazard function is given by

$$\hat{H}(t) = \sum_{tj \le t} \frac{d_j}{r_j} \qquad (3.8)$$

This equation is known as the **Nelson-Aalen estimator** where j=1,2…k

For purposes of obtaining the **Nelson-Aalen estimator** one needs to construct a table (4) as shown

j	t_j	d_j	c_j	r_j	$\frac{d_j}{r_j}$	$\hat{H}(t)$
0	t_0	d_0	c_0	r_0	$\frac{d_j}{r_j}$	0
1	t_1	d_1	d_1	r_1	$\frac{d_1}{r_1}$	$0+\frac{d_1}{r_1}$
.						
.						
.						
k	t_k	d_k	c_k	r_k		$0+\frac{d_1}{r_1}+\frac{d_2}{r_2}+\ldots\frac{d_k}{r_k}$

Table 4:Nelson-Aalen Estimator table

3.8 Using Log rank test to compare survival functions of two groups of TB patients (Male and Female)

Log rank test is a method for comparing the survival times of two or more groups of subjects. It involves the calculation of observed and expected frequencies of failures in separate time intervals. The relevant test statistic is a comparison of the observed number of events (loss to follow-up) occurring at each particular point with the number to be expected if the survival experience of the two groups is the same (Walter, 2008).

To compare whether the survival functions for male and female Tuberculosis patients are statistically significantly different one would test the hypothesis

$H_0 : S(m) = S(f)$

Versus

$H_1 : S(m) \neq S(f)$

Where;

S (m) is the survival distribution function for the male TB patients group

S (f) is the survival function for female TB patients group

As described earlier let;

r_j, be the number of participants at risk of event of interest (loss to follow-up) at time tj .

d_j, be the number of events of interest (loss to follow-up) observed at time tj .

Suppose the ordered distinct event (loss to follow-up) times are $t_1 < t_2 \cdots < t_k$

Therefore the distribution of individuals at time t_j will be a 2*2 contingency table as shown in table (5)

Group	**Number lost to follow-up**	**Number not lost**	**Number at risk**
Group 1(Male)	d_{1j}	$r_{1j} - d_{1j}$	r_{1j}
Group 1(Female)	d_{2j}	$r_{1j} - d_{1j}$	r_{2j}
Total	d_j	$r_j - d_j$	r_j

Table 5:Log rank test 2*2 contingency table

Under H_0, d_{1j} follows a hyper-geometric distribution given by

$$f(d_{1j}) = \frac{\binom{d_j}{d_{1j}}\binom{r_j - d_j}{r_{1j} - d_{1j}}}{\binom{r_j}{r_{1j}}}$$

Given by

$$\binom{d_j}{d_{1j}} = \frac{d_j!}{d_{1j}!(d_j - d_{1j})!}$$

The expectation and variance of the random variable d_{1j} is given by

$$\mathrm{E}(d_{1j}) = \frac{d_j r_{1j}}{r_j}$$

$$Var(d_{1j}) = \frac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}$$

Hence, **the sum of differences between the observed and expected numbers** of Group I individuals experiencing the event can be written as

$$U_L = \sum_{j=1}^{k} d_{1j} - E(d_{1j})$$

Under H_0 , the mean of U_L is

E(U_L)=0 and the

Variance is given as

$$Var(d_{1j}) = \frac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}$$

$$= V_L$$

When the number of lost to follow-up is not too small, it can be shown that under, H_0

U_L has approximately a normal distribution

$$\frac{U_L}{\sqrt{V_L}} \sim N(0,1) \text{ and}$$

$$\frac{U_L}{\sqrt{V_L}} \sim \chi^2(1)$$

Let the Log rank statistic be denoted by Z

$$Z = \frac{\sum_{tj} d_{1j} - \sum_{tj} \mathrm{E}(d_{1j})}{\sqrt{\sum_{tj} \mathrm{var}(d_{1j})}} \qquad (3.7)$$

From the probability distribution if Z~N(0,1), then $Z^2 \sim \chi^2(1)$

Z^2 is therefore known as the log rank statistic which follows a chi-square distribution with one degree of freedom .

$$Z^2 = \frac{\left\{\sum_{tj}^{k}\left(d_{lj-}\frac{d_{1j}r_{1j}}{r_{1j}}\right)\right\}^2}{\frac{r_{1j}r_{2j}d_j(r_j-d_j)}{r_j^2(r_j-1)}}$$

Thus testing the Hypothesis using the log rank test we obtained a critical value from the chi-square distribution table at α level of significance.

If the Test statistic (Z^2) is greater than the critical value then we reject $H_0 : S(m) = S(f)$ and conclude that survival distribution function for the two groups is significantly different.

3.9 Application of Cox Proportional Hazards regression model in identification of significant predictors of loss to follow-up among Tuberculosis patients

To achieve the objective of identifying the significant predictors of loss to follow-up among the Tuberculosis patients we had to conduct a **cox proportional hazards model** regression analysis. This was to test whether the regression coefficients in the model pertaining to each of the predictor variables are statistically significantly different from zero.

The hypotheses tested were;

$H_0 : (\beta_i) = 0$

Versus

$H_1 : (\beta_i) \neq 0$

The most common approach to model covariate effects on survival is the Cox proportional

hazards model by Cox (1972), which takes into account the effect of **censored** observations. (Klein,1997).

The cox proportional hazards regression model is the most popular approach to modeling the Predictors in order to estimate their effects on the time to event. It is a **semi-parametric** model because it takes no assumption of the probability distribution function for the time to event; the baseline hazard takes any form and the predictors enter the model through the linear predictor ni.

Cox proportional hazards regression model is based on the assumption that the **hazards are proportional.**

The cox proportional hazards regression model is based on the hazard function $h(t, X)$ at time t

Suppose we have $x_1, x_2 ..., x_p$ be the values taken by the covariates $X_1, X_2 ..., X_p$ with p as the total number of predictor variables in the model. The linear predictor ni is given by

$$n_i = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$$

The Cox Proportional Hazards regression model is expressed as;

$$h(t, X) = h_0(t) \exp\left(\sum_{i=1} \beta_i X_i \right) \qquad (3.7)$$

Where;

$h_0(t)$ Is the hazard baseline function

β_i is the ith regression parameter for the Cox proportional hazards model

X_i is the ith predictor variable

Where ;

$h_0(t) \geq 0$

$$\sum_{i=1}^{p} \beta_i X_i \geq 0$$

We recall; $i = 1,2..., p$ the number of predictors in the cox proportional hazards model

The quantity $h_0(t)$ is the baseline or underlying hazard function and corresponds to the probability of reaching an event when all the explanatory variables are zero. The baseline hazard function is analogous to the intercept in ordinary regression (since exp0 = 1). It is unknown (arbitrary) function (*the value of transformed hazard-rate when all predictors are 0*) giving the hazard function for the standard set of conditions **X = 0**.

Making special assumptions about $h_0(t)$ leads to parametric models; the exponential and Weibull distributions. But the advantage of Cox's model is the fact that such assumptions can be avoided. His approach is said to be *semi parametric*. (Klein, 2003)

The advantage of Cox Proportional Hazards Model is that it takes care of the censored observations aside from it being semi-parametric.

3.9.1 The proportionality of hazards assumption

The proportionality of the hazards is an important assumption of the cox proportional hazards model. The Cox model is often called **proportional hazards** model which means that the hazard ratio is constant over time. This also implies that the hazard for an individual observation is proportional to the hazard for any other individual observation.

Given two individuals with covariate values X and X*, the ratio of their hazard rates is obtained as follows;

Let,

$h(t, X) = h_0(t) \exp\left(\sum_{i=1} \beta_i X_i \right)$ be the hazard function for individual X

And

$h(t, X) = h_0(t) \exp\left(\sum_{i=1} \beta_i X^*_i \right)$ be the hazard function for individual X^*

The Hazard ration (HR) is obtained by

$$HR = \frac{h(t, X) = h_0(t) \exp\left(\sum_{i=1} \beta_i X_i \right)}{h(t, X^*) = h_0(t) \exp\left(\sum_{i=1} \beta_i X^*_i \right)} \tag{3.8}$$

$$\frac{h(t, X)}{h(t, X^*)} = \frac{h_o(t) \exp\left[\sum_{k=1}^{p} \beta_i X_i \right]}{h_o(t) \exp\left[\sum_{k=1}^{p} \beta_i X^*_i \right]} = \exp\left[\sum_{k=1}^{p} \beta_k (X_i^* - X_i \right] \tag{3.9}$$

For the proportional hazards assumption to be satisfied, the value for the individual X and X* should be a **constant** and so hazard rates are proportional.

The quantity (3.9) is called the relative risk (hazard ratio) of an individual with risk factor X having an event as compared to an individual with risk factor X *.

Since the Cox proportional hazards model relies on the hazards to be proportional, i.e. that the effect of a given covariate does not change over time, it is very important to verify that the covariates satisfy the assumption of proportionality. If this assumption is violated, the simple Cox model is invalid, and more sophisticated analyses are required. If the interest centers upon a

binary covariate, X_1, whose relative risk changes over time, one approach is to introduce a time-dependent covariate .(Persson, I. 2002)

Considering the several procedures proposed to check the assumption of proportional hazards, both numerical and graphical. This study used the plot of survival curves based on the Cox model and Kaplan-Meier estimates for groups. By plotting the cumulative observed survival times to establish the Cox model fitness, if the model fit is adequate then the points should follow a 45degree line starting from the origin of the plot.(Tolosie & Sharma, 2014)

The evaluation of the proportional hazards assumption can be done numerically or graphically, a great number of procedures have been proposed over the years. Some of the procedures require partitioning of failure time, some require categorization of covariates, some include a spline function, and some can be applied to the untransformed dataset.

None of the methods, neither numerical nor graphical, are today known to be better than the others in finding out whether the hazards are proportional or not. Some authors recommend using numerical tests (e.g. Hosmer and Lemeshow (1999)) and others recommend graphical procedures since they believe that the proportional hazards assumption only approximates the correct model for a covariate and that any formal test, based on a large enough sample, will reject the null hypothesis of proportionality (Klein ,1997)

3.9.2 Estimation of the Cox proportional hazards model parameters

The regression coefficients in this study were estimated using Maximum Likelihood statistical method (Collett, 2003).Parameter estimates in the Cox PH model are obtained by maximizing the partial likelihood as opposed to the likelihood. Cox and others have shown that this partial log-likelihood can be treated as an ordinary log-likelihood to derive valid (partial) Maximum

Likelihood Estimates (MLEs) of β. Therefore we can estimate hazard ratios and confidence intervals using maximum likelihood. (BIOST515, 2004)

Steps followed in obtaining the partial Likelihood

(i) Arrange the event times in order

Let $t_1 < t_2 < ... < t_{k.}$ represent the distinct ordered event times assuming there are no ties.

(ii) Obtain the risk sets

$$R(t_1), R(t_2), ..., R(t_k)$$

(iii)Obtain the probability of an individual experiencing the event of interest at time t_i given the risk set $R(t_i)$, $t_i \big/ R(t_i)$

This is given by

$$\frac{\exp(\beta X_i)}{\sum_{r \in R(t_i)} \exp(\beta X_r)}$$

(iii) Obtain the partial likelihood $L(\beta)$

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp\left[\sum_{k=1}^{p} \beta X_i\right]}{\sum_{j \in R(t_i)} \exp \sum_{k=1}^{p} \beta X_r} \qquad (3.10)$$

When we solve equation (3.10) we obtain $\exp(\beta)$ and then β

This is not likelihood in the traditional sense but it is treated as one, and inference is carried out by usual means. The partial maximum likelihood estimates are found by maximizing (3.10) or, equivalently, the logarithm of (3.10).

Steps to solving the equation (3.10)

(i) Obtain the partial likelihood $L(\beta)$

(ii) Obtain the $\log(\exp(\beta)$ (natural log)

(iii) We then obtain the first derivative of the $\log(\exp(\beta)$ with respect to β

$$\frac{\partial \log(L(\beta))}{\partial \beta}$$

With the derivative given as

$\frac{\partial \log(L(\beta))}{\partial \beta} = 0$ we can obtain $(\exp(\beta)$ and then β

Therefore we can estimate β_i the i^{th} regression parameter for the cox proportional hazards regression model

3.9.3 Testing the significance of the parameters

To test the significance of the parameters estimated using the cox proportional hazards regression model; there are several methods that can be used: Wald test, Partial likelihood ratio and score Test

Using the Wald Test to test the significance of the parameters

We needed to set and test the hypothesis;

$H_o : \beta_i = 0$

Versus

$H_0 : (\beta_i) \neq 0$

The Wald test statistic is denoted by W,

We obtain the Wald test statistic as follows;

$$W = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \qquad (3.11)$$

We note that $W \sim H_o \sim N(0,1)$

To carry out the hypothesis test to show that the coefficients are statistically significantly different from zero, we obtain the critical value from the normal distribution tables at α level of significance. If the test statistic (W) is greater than the critical value then we reject the $H_o : \beta_i = 0$ and conclude that the regression parameter is statistically significantly different from zero and therefore a significant predictor of Tuberculosis loss to follow up.

3.9.3b Variable selection

Several methods for model building have been described in literature with the options 'backward', 'forward', 'stepwise' and 'score' are provided for model building. If multi-level factors are to be considered in the model, the -2 Log Likelihood statistics may be used for variable selection. (Collet, 2004) suggests the following strategy:

Step 1

The first step is to fit models that contain each of the variables one at a time. The values of -2Log Likelihood for these models are then compared with that for the null model to determine which variables on their own significantly reduce the value of this statistic

Step 2

The variables which appear to be important from Step 1 are then fitted together. Only those whose exclusion from the model lead to a significant increase in the value of -2 Log Likelihood are retained in the model.

Step 3

Any variable that was not under consideration in Step 2 is now added to the model, one at a time Any variable that significantly reduces the value of -2 Log Likelihood is retained in the model

Step 4

A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of -2 Log Likelihood, and that no term not included significantly reduces the value of -2 Log Likelihood.

3.10 Overall goodness of the model

According to Arjas (1988) the overall goodness of fit of a Cox proportional hazards regression model can be assessed by plotting the cumulative observed versus the cumulative expected number of events for subjects with observed (not censored) survival times. If the model fit is adequate, then the points should follow a 45-degree line beginning at the origin. The proportionality assumption test containing all the time dependent and their **interaction with the log time** was carried out to test the proportionality hypothesis.

3.11 Interpretation of the results

Interpreting the Cox model involves examining the coefficients for each explanatory variable. The results in the final model are interpreted in terms of hazard ratios (HR). The coefficient of the categorical covariates is interpreted as the logarithm of the ratio of the hazard of loss to follow-up to the baseline (reference group) hazard. That is, they are interpreted by comparing the reference group with others. Similarly, the coefficient for a **continuous** explanatory variable indicates the estimated change in the logarithm of the hazard ratio for a unit increase in the value of the respective covariate when the remaining covariates in the model are controlled. A positive regression coefficient for an explanatory variable means that the hazard is higher and thus the prognosis worse. Conversely, a negative regression coefficient implies a better prognosis for patients with higher values of that variable (Walters, 1999)

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

4.1 Summary Statistics

From the 1,275 patients in this study a total of 107 (8.4%) interrupted treatment and 1,168 (91.6%) were censored. The gender proportions were 692 (54.3%) male and 583 (45.7%) female. Male patients showed higher percentage of treatment interruption (4.86%) compared to female (3.53%). This study agrees with findings of Dooley et al., 2011 in his research on risk factors for default found that the male gender was a risk factor for default.

The age group (25-44 years) showed the highest percentage (5.33%) compared treatment interruption proportions in other age groups. Comparing the weight of the patients enrolled; those who had weight (≥45) had higher percentage of (5.65%) compared to those with less than 45 kgs (2.35%). HIV negative patients had high proportions of treatment interruption of 2.90% (37) compared to HIV positive 2.35 % (30). In terms of the type of TB, pulmonary TB had higher risk of treatment interruption (7.29%) compared to extra pulmonary TB at 1.10%. New patients were more at risk (7.14%) compared to non-new patients (1.25%). The percentages of treatment interruption 5.88%, 2.12% and 0.39% occurred in public, private and prisons sector respectively as shown in table (6).

Patient Characteristics	Total	% of the total treatment interrupters	No and % from total treatment interrupters	Censored
Total=1275				
Gender				
Male	692 (54.3%)	4.86%	62 (57.9%)	630 (53.9%)
Female	583 (45.7%)	3.53%	45 (41.1%)	538 (46.1%)
Total	**1275**		**107**	**1168**
Age				
0-24	305	2.20%	28 (26.2%)	277 (23.7%)
25-44	697	5.33%	68 (63.6%)	629 (53.9%)
Over 45	273	0.86%	11 (10.3%)	262 (22.4%)
	1275		**107**	**1168**
Weight				
<45	375	2.35%	30 (29.4%)	345(30.4%)
≥45	861	5.65%	72 (70.6%)	789(69.6%)
			102	**1134**
Missing	39			
HIV status				
Negative	390	2.90%	37(34.6%)	353(30.2%)
Positive	885	2.35%	30(65.4 %)	815(69.8%)
Type of TB				
Extra pulmonary TB	257	1.10%	14(13.10)	243(20.8%)
Pulmonary	1018	7.29%	93(86.9%)	925(79.2%)
Type of Patient				
New	1165	7.14%	91(85.0%)	1074(92.0%)
Non new	110	1.25%	16(15.0%)	94(8.0%)
Sector				
Prisons	82	0.39%	5(4.7%)	27(2.3%)
Private	433	2.12%	27(25.2%)	406(34.8%)
Public	810	5.88%	75(70.1%)	735(62.9%)

Table 6:Tuberculosis Patient characteristics

4.2 Time to Tuberculosis treatment interruption

From the table (7) given we have a display of summary status of Tuberculosis patients, out of a total of 1,275 patients, 107 patients (8.4%) interrupted treatment and 1,168 (91.6%) were

censored. The median time of treatment interruption was 2.167 months (65 days) range from 1.33 and 3.7 months. The median follow-up time was 6.10 months (183 days) for the censored patients (range from 5.7 to 6.7 months).This analysis shows that most of the Tuberculosis patients are lost to follow up during the intensive phase of treatment which is within the first two months. This finding is similar to Yone et al (2011) who in his study on Incidence, time and determinants of tuberculosis treatment interruption in Yaounde, Cameroon in a retrospective hospital register-based cohort study found that at each time point during follow-up, the probability of treatment discontinuation was always lower in patients hospitalized during the intensive treatment phase than in those treated as outpatients during the intensive phase. Similarly, the probability of discontinuation was always lower in patients with known status for HIV infection than in those with unknown status.

Status of TB patients	**Frequency**	**Percent**
Censored	1168	91.6%
Treatment interrupters	107	8.4%

Status analysis	**Treatment interrupters**	**Censored**
Number of observations	107	1168
Mean	2.949	6.184
Standard deviation	2.5840	1.3280
Median	2.167	6.100
Lower quartile	1.333	5.700
Upper quartile	3.700	6.500
Minimum	0	
Maximum	13.7	

Table 7:Summary statistics of status of TB patients and months of follow-up time

4.3 Log rank and Wilcoxon test of no significance on the Tuberculosis patients factors

According to table (8), Log rank test and Wilcoxon test are not significant in treatment interruption survival experience between the various categories for covariates HIV status, weight, gender and type of TB at alpha significance level of 5% ($p<0.05$).

		Log rank test		Wilcoxon test	
Covariates	**Df**	**Chi-square**	**Pr>chi-square**	**Chi-square**	**Pr>chi-square**
Age	1	4.113	0.043	5.139	0.023
HIV status	1	1.191	0.275	1.306	0.253
Weight	1	0.440	0.507	0.243	0.622
Type of Patient	1	4.739	0.029	3.781	0.052
Gender	1	0.828	0.363	0.298	0.585
TB type	1	2.586	0.108	2.527	0.112

Table 8:Log rank and Wilcoxon test of no significance

We find significant differences in survival experience of the patients in different categories of age as shown in figure 2 where age was categorized as less than 35 years and greater than 35 years respectively. Patients with more than 35 years of age had better survival rates compared to those who were younger. This implies that an additional year in age increased the treatment interruption survival rate of Tuberculosis patient and those who were younger had higher treatment interruption.

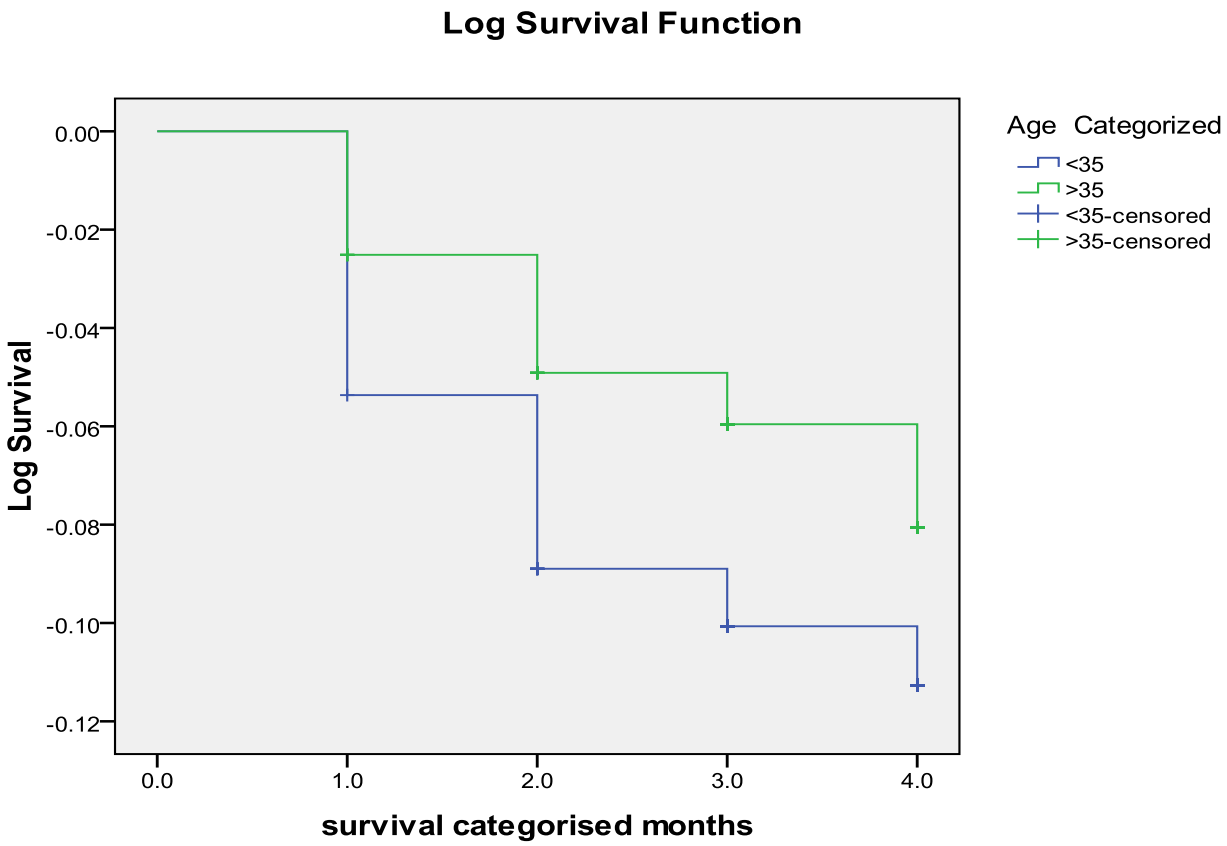

Figure 2:Comparison of Kaplan-Meier survival curves by age category

There were also significant differences in survival for the type of TB patients as shown in figure 3. The Tuberculosis patients had been categorized as new and not new patients. A new patient who is described as one who has never had treatment for TB, or has been on anti-TB treatment for less than four weeks had higher survival rate compared to non-new patients who had higher treatment interruption rate.

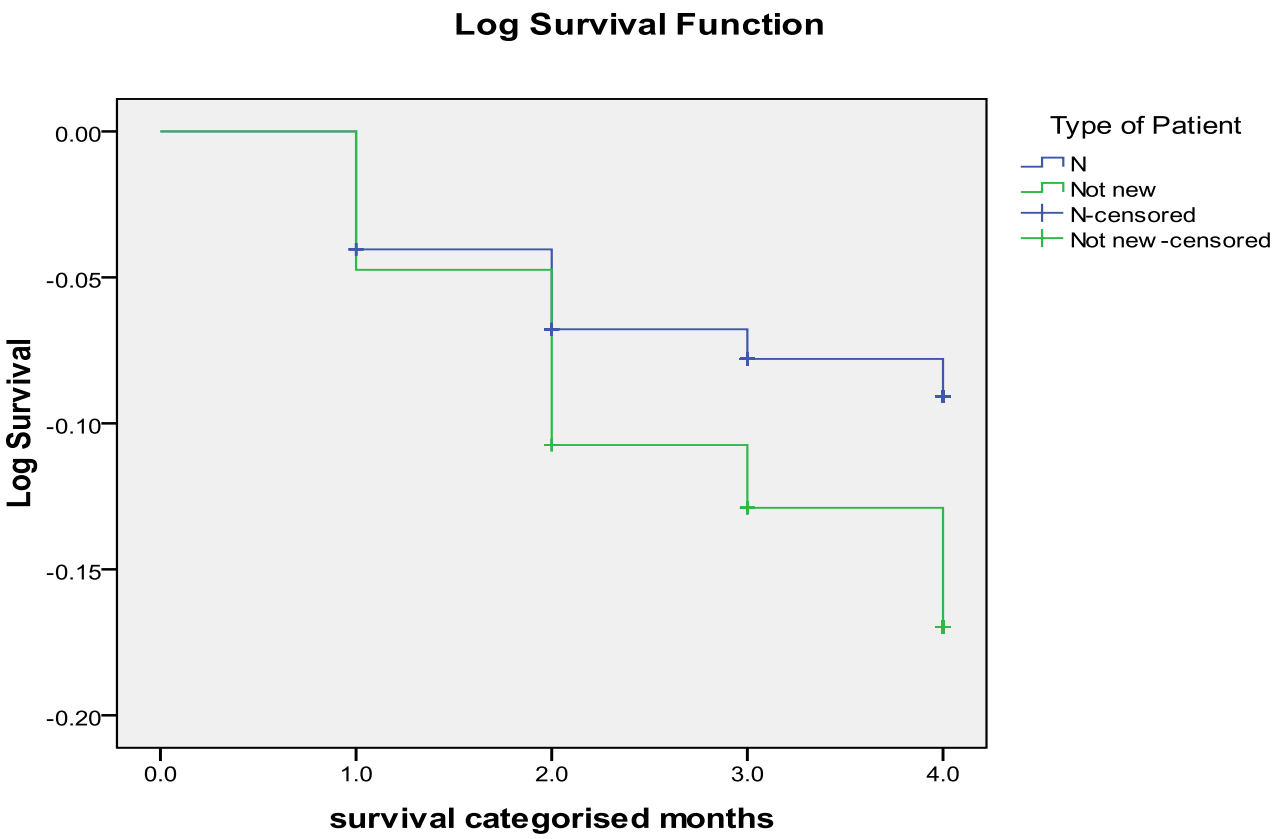

Figure 3: Comparison of Kaplan-Meier survival curves by type of patient category

The figure (4) below show the comparison for two groups of men and female, according to Kaplan-Meir Estimate there was difference in survival distribution function between male and female Tuberculosis.

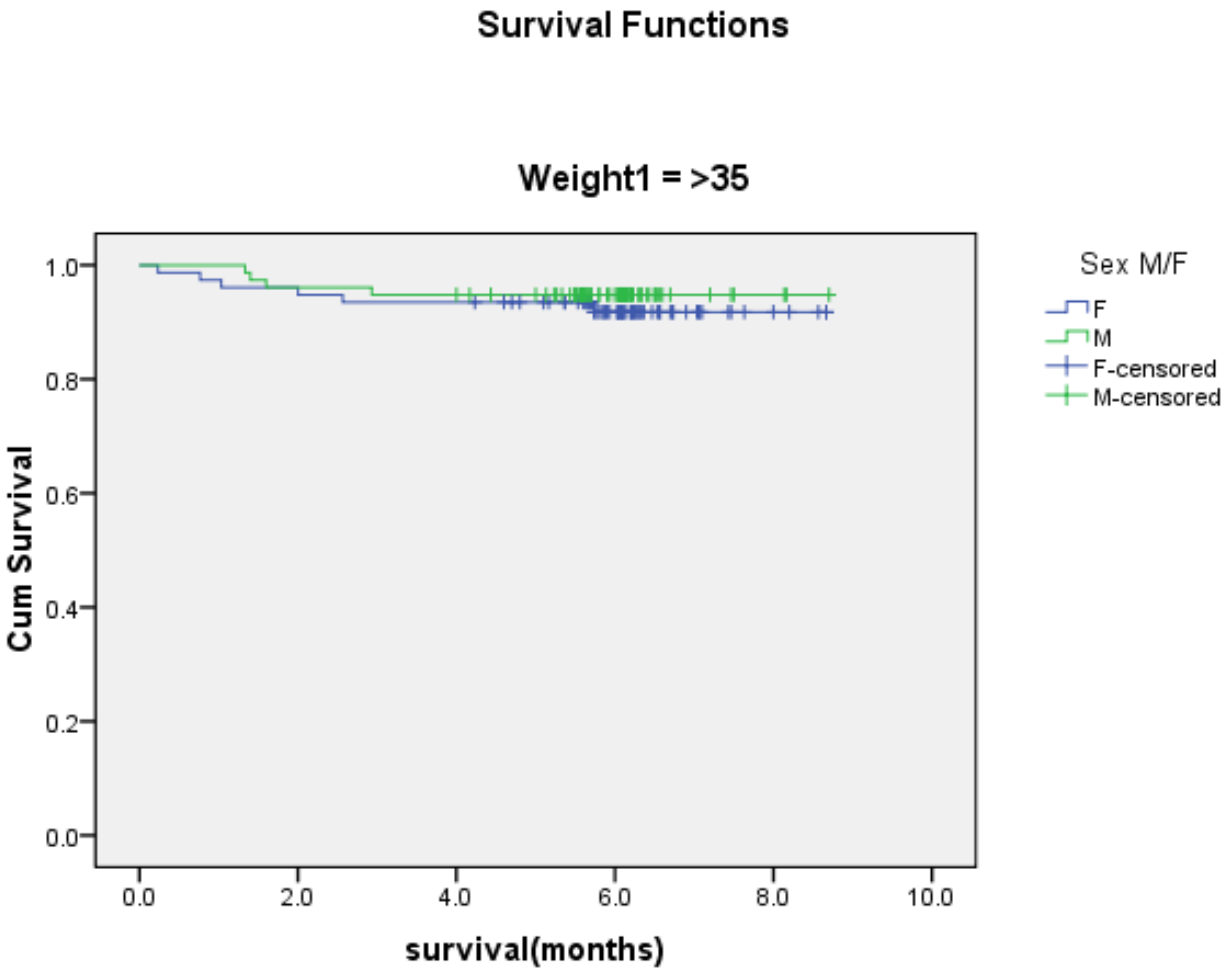

Figure 4: Comparison of survival functions by Gender

4.4 Results of the proportional hazards Cox regression model of TB patients

With all variables

From the six variable analyzed using Cox regression model based on the maximized log partial likelihood (-2LL) weight variable had the highest reduction in -2LL (β estimate) 1287.343). The difference from the weight variable is 156.444 which is statistically significant with p-value of (0.045) and therefore any improvement on the null model would be achieved by adding weight

variable. The reduction on adding age to the null model is 95.452 which is also significant (p value is less than 0.05).The results from the Wald chi-square test the variable weight, age and patient type are significant and thus can be important factors for inclusion in the model development as shown in the table (9).

Covariates	Standard error	-2LL	Wald	Df	P-value	Parameter estimate	Hazard ratio	95.0% CI	
								Lower	Upper
HIV Status	0.204	1352.167	1.089	1	0.097	0.196	1.216	0.823	1.891
Weight	0.006	1287.348	0.000	1	0.045	0.000	1.000	0.728	2.793
Age	0.007	1348.340	4.394	1	0.036	-0.014	0.986	0.416	0.971
Patient type	0.275	1441.048	4.809	1	0.028	-0.492	0.618	0.316	0.937
Type of TB	0.289	1349.876	1.426	1	0.232	-0.487	0.615	0.401	1.249
Gender	0.199	1353.007	0.839	1	0.360	-0.048	0.953	0.743	1.114
-2LL Null model value is 1443.792									

Table 9: Results of the proportional hazards Cox regression model of TB patients (All variables

4.5 Proportional Hazards assumption test

Test results from the time depended variables; age, weight and type of TB patient were not significant according to table (10). At 5% level of significance there is no sufficient evidence to reject the proportionality assumption of all the covariates; the Cox model was therefore fit and sufficient for the analysis of the Tuberculosis predicators and time to treatment interruption in this study.

Covariates	Standard error	Df	P- value	Parameter estimate	Chi-Square	Hazard ratio
Weight	0.02141	1	0.3963	0.00501	0.6405	1.0050
Age	0.73092	1	0.8395	0.76267	2.0218	2.1439
Patient type	1.23372	1	0.3531	0.9191	3.7103	2. 5067
Weight*log(time)	0.0049	1	0.2107	-0.01195	0.5783	0.98812
Age*log (time)	0.01488	1	0.4203	0.36468	2.6481	1.44005
Patient type*log(time)	0.37071	1	0.2583	0.14114	3.1645	1.15504

Table 10: Results of proportionality assumption test containing time dependent covariates.

According to table (11), the hypothesis test results indicate that there was no sufficient evidence to reject the null hypothesis that proportionality of the data used holds true at 5% level of significance. This implies that the Cox model used in this study was fit for analysis of the covariates used.

Linear Hypothesis test results			
Label	Wald Chi-square	DF	Pr>Chi-square
Proportionality test	3.1795	3	0.5196

Table 11:Linear hypothesis test results

4.6 Final Model with the significant variables (age, weight, patient type)

The table (12) shows the final model with the significant covariates. The study found that the variable age, weight and patient type were significant predators of treatment interruption for TB patients.

Covariates	Standard error	Df	P- value	Parameter estimate	Chi-Square	Hazard ratio	95.0% CI	
							Lower	Upper
Weight	0.00689	1	0.0004	0.01327	9.0405	1.0134	1.011	1.041
Age	0.0089	1	**0.0271**	-0.017	3.396	0.9831	0.416	0.971
Patient type	0.60125	1	**<0.0001**	2.11351	16.7103	8.2772	4.375	19.353
The -2LL value for the model is 1401.167								

Table 12:Final Model with significant covariates

CHAPTER 5: CONCLUSION AND THE RECOMMENDATIONS

This chapter gives the conclusion and the recommendations of the research on application of Cox proportional hazards model in case of tuberculosis patients in Kisumu County, Kenya.

5.1 Conclusion

From the 1,275 patients in this study, 107 patients (8.4%) had treatment interruption. The median time of loss to follow up was 2.167 months (65 days) range from 1.33 and 3.7. The rate of loss to follow-up was highest during initial two months, the intensive phase of treatment. Results of the proportional hazards Cox regression analysis of TB patients revealed that the covariates age, weight and Tuberculosis patient type were significant factors associated with treatment interruption for TB patients.

5.2 Recommendations

There is need to strengthen the follow up of patients on Tuberculosis treatment especially during the first two months of intensive phase as well as until completion of treatment . Targeted interventions to increase treatment adherence for patients at highest risk of treatment interruption are critical; necessary actions are required to improve on the patients' weight. Enhanced patient pre-treatment counseling and education about TB is recommended to reduce the high rate of loss to follow up. This study recommends the need more research on the social economic factors that may be contributing to loss to follow up of Tuberculosis patients on treatment.

REFERENCES

World Health Organization. (2013). *Global tuberculosis report 2013*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2014). *Global Fact Sheets*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2012). *Global tuberculosis report 2013*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2012). *Global tuberculosis report 2011*. Geneva, Switzerland: World Health Organization.

Ministry of Health, (2013). TB Treatment Guidelines, National Tuberculosis program:Nairobi:Kenya.

Afutu, F. K., Zachariah, R., Hinderaker, S. G., Ntoah-Boadi, H., Obeng, E. A., Bonsu, F. A., & Harries, A. D. (2012). *High initial treatment interruption in patients with smear-positive pulmonary tuberculosis at a regional hospital in Accra*, Ghana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *106*(8), 511–513. http://doi.org/10.1016/j.trstmh.2012.05.002

Clark-Curtiss JE, Haydel SE. *Molecular genetics of Mycobacterium tuberculosis pathogenesis*. Annual Review of Microbiology 2003;57:517-49.

Dannenberg AM. *Pathogenesis of Pulmonary Tuberculosis*. American Review of Respiratory Disease 1982;125(3):25-30.

D.R. Cox, *Regression models and life-tables (with discussion)*, Journal of the Royal Statistical Society, B 24 (1972), 187-220

Dooley, K. E., Lahlou, O., Knudsen, J., Elmessaoudi, M. D., Cherkaoui, I., & El Aouad, R. (2011). *Risk factors for tuberculosis treatment failure, treatment interruption, or relapse and outcomes of retreatment in Morocco*. *BMC Public Health*, *11*(1), 140.

Jakubowiak, W., Bogorodskaya, E., Borisov, S., Danilova, I., & Kourbatova, E. (2009). *Treatment interruptions and duration associated with treatment interruption among new patients with tuberculosis in six regions of Russi*a. *International Journal of Infectious Diseases*, *13*(3), 362–368. http://doi.org/10.1016/j.ijid.2008.07.015

Joint United Nations Programme on HIV/AIDS. (2012). *Global report UNAIDS report on the global AIDS epidemic: 2012.* [Geneva]: UNAIDS. Retrieved from http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf

Klein, J. P. (2003). *Survival analysis: techniques for censored and truncated data* (2nd ed). New York: Springer.

Muture, B. N., Keraka, M. N., Kimuu, P. K., Kabiru, E. W., Ombeka, V. O., & Oguya, F. (2011). *Factors associated with treatment interruption from treatment among tuberculosis*

patients in Nairobi province, Kenya: a case control study. *BMC Public Health*, *11*(1), 696.

Persson, I. S. (2002). *Essays on the assumption of proportional hazards in Cox regression*. Uppsala: Acta Univ. Upsaliensis.

Tolosie, K., & Sharma, M. K. (2014). *Application of Cox Proportional Hazards Model in Case of Tuberculosis Patients in Selected Addis Ababa Health Centres*, Ethiopia. *Tuberculosis Research and Treatment*, *2014*, 1–11. http://doi.org/10.1155/2014/536976

Walters, S. J. (1999). *What is a Cox model?*. Hayward Medical Communications. Retrieved from http://www.ida.liu.se/~kawah/Cox1.pdf

World Health Organization. (2013). *Global tuberculosis report 2013*. Geneva, Switzerland: World Health Organization.

Zhang, Q., Gaafer, M., & El Bayoumy, I. (2014). *Determinants of Treatment interruption from Pulmonary Tuberculosis Treatment in Kuwait*. *The Scientific World Journal*, *2014*, 1–6. http://doi.org/10.1155/2014/672825.