Modeling Time-to- Death of Women with Cervical Cancer: A Case Study at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia

BY:

Selamawit Endale

A Thesis to be submitted to School of Graduate Studies Jimma University College of Natural Science Department of Statistics in the Partial Fulfillment of the Requirement for the Degree of Master of Science in Biostatistics

June, 2016

Jimma Ethiopia

Modeling Time-to- Death of Women with Cervical Cancer: A Case Study at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia

BY:

Selamawit Endale

Advisor: Agatamudi Lakshmana Rao(PhD)

Co- Advisor: Tesfaye Gatachew (M.Sc.)

June, 2016
Jimma Ethiopia

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission if an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in her or his judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Selamawit Endale

Date: _____

Signature: _____

June, 2016

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we here by certify that we have read the thesis prepared by Selamawit Endale under our guidance, which is entitled **"Modeling Time-to-Death of Women with Cervical Cancer: A Case Study at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia"**, in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready

Agatamudi Lakshmana Rao(PhD)	_____	_____
Advisor	Signature	Date
Tasfaye Getachew(M.Sc.)	_____	_____
Co-Advisor	Signature	Date

As the members of the board of examiners of M.Sc. thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

ACKNOWLEDGMENT

First and foremost I would like to thank the almighty God with his mother St. Mary for the gift of health and strength throughout the process of carrying out this thesis and for seeing me through all. I wish to acknowledge the support of all who contributed in various ways to make this study success.

I want to express a sincere acknowledgement to my advisor Dr. Agatamudi Lakshmana Rao for his advice and guidance that contributed to the successful realization of the thesis.

My deep gratitude goes to my co-advisor, Mr. Tesfaye Getachew for his valuable suggestions and comments to the successful realization of this thesis. It is a great pleasure working with him. I can only hope that his cooperation will keep on going in the future.

I would like to acknowledge the department of Oncology staffs at the Tikur Anbessa Specialized Hospital to undertake this study with their cooperation and permission in using the data with special thanks for Dr. Wondimagegn Astatke, Dr. Mathios Assefa, Dr. Abreham Adamu, Mr. Zelalem, Mr. Dabasa, Mr. Namomsa, Mis Etagni and Miss Mulu for their willingness to help me.

I gratefully acknowledge my brother Mr. Abyot Endale for his continuous supportive comments and constant guidance throughout my thesis work and also I want to express a sincere acknowledgement to Miss Selamawit Sisay, Mr. Garemew Muleta, Mr. Belay Birlie and Mr. Dinberu Seyoum for those unflinching encouragement, advice and guidance starting from title selection. Finally, I am also very grateful to all my beloved parents and friends whose presence and support have always been important to me.

DEDICATION

I dedicate this work to my dear sisters and brothers for making me who I am today, for their support and for teaching me the value of education.

ABSTRACT

Background: Cervical Cancer is one of the leading causes of death in the world and represents a tremendous burden on patients, families and societies. It is estimated that over one million women worldwide currently have cervical cancer; most of them have not been diagnosed, or have no access to treatment that could cure them or prolong their lives. This study aimed at investigated the potential risk factors affecting death time of women with cervical cancer at Tikur Anbessa Specialized Hospital.

Method: Data were taken from patients' medical record card that enrolled during September 2011 to September 2015. Kaplan-Meier estimation method, Cox proportional hazard model and Accelerated failure time were used for the analysis with the help of R statistical package and STATA software's.

Results: The Lognormal Accelerated failure time model was preferred over Exponential, Weibull and Log-logistic models based on Akaike's Information Criterion evidence. The results implied that not giving birth up to the study ends were prolong the timing death of cervical cancer patients while age class 51-60, 61-70,>70, smoking cigarettes, patients with stage III & IV disease, family history of cervical cancer, history of abortion and living with HIV AIDS were significantly shorten survival time of cervical cancer patients.

Conclusion: Finally, the findings of this study implied that age, smoking cigarettes, region, stage, family history, abortion history, living with HIV AIDS and age at first birth were major factors related to survival time of cervical cancer patients. It is recommended policy maker, ministry of health and mass media to make interventions based on the risky groups for cervical cancer.

Key words: Survival Data Analysis, Proportional hazard, AFT, Cervical cancer, survival time

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ANC	Antenatal Care
CSA	Central Statistical Agency
DHS	Demographic and Health Survey
DNA	Deoxyrebose Nucleic Acid
ECA	Ethiopian Cancer Association
EDHS	Ethiopian Demographic and Health Survey
FDA	Food and Drug Administration
FIGO	International Federation of Obstetrics and Gynecology
HDI	Human Development Index
HIV	Human Immune Virus
HPV	Human Pappiloma Viruses
IARC	International Agency for Research on Cancer
NHS	National Health Service
PATH	Program for Appropriate Technology in Health
PH	Proportional Hazard
PLND	Pelvic Lymph Nodes
PO	proportional Odds
TASH	Tikur Anbessa Specialized Hospital
UK	United Kingdom
USA	United State of America
WHO	World Health Organization

1. INTRODUCTION

1.1. Background of the Study

Cancer is the name given to a collection of related diseases. In all types of cancer, some of the body's cells begin to divide without stopping and spread into surrounding tissues. Cancer can start almost anywhere in the human body, which is made up of trillions of cells. It is one of the leading causes of death in the world and represents a tremendous burden on patients, families and societies (WHO, 2009). Chronic diseases like cancer and cardiovascular disease are major causes of morbidity and mortality. In particular, cancer is one of the most common and deadly diseases worldwide. The burden of disease impinges on the lives of tens of millions annually (NCI, 2014). According to estimates from the International Agency for Research on Cancer (IARC), there were 12.7 million new cancer cases in 2015 worldwide, of which 5.6 million occurred in economically developed countries and 7.1 million in economically developing countries. The corresponding estimates for total cancer deaths in similar year were 7.6 million, 2.8 million in economically developed countries and 4.8 million in economically developing countries. There were an estimated 4.9 million new cases and 0.266 million global deaths from cervical cancer in the same year accounting for 7.5% of all female cancer deaths. (Rebecca S.*et al;* IARC; WHO, 2015). Between 2000 and 2020, the total number of cases of cancer in the developing world is predicted to increase by 73% and 29% in the developed world largely as a result of an increase in the number of old people (Parkin DM., 2000).

The American Cancer Society estimates that 12,900 diagnoses of cervical cancer in the end of 2015 in the USA (ACS, 2015). From this over 4,000 women in the USA die from cervical cancer each year. The National Health Service (NHS), UK, says that over 3,000 women are diagnosed with cervical cancer each year in the UK (NHS, 2015).

Cancer is an emerging public health problem in Africa. As estimate by International Agency for Research on Cancer (IARC), about 715,000 new cancer cases and 542,000 cancer deaths occurred in Africa in 2009. These numbers are projected to nearly double which mean 1.28 million new cancer cases and 970,000 cancer deaths by 2030 simply due to the aging and growth of the population, with the potential to be even higher because of

the adoption of behaviors associated with western lifestyles, such as smoking, unhealthy diet, and physical inactivity. Despite this growing burden, cancer continues to receive low public health priority in Africa, largely because of limited resources and other pressing public health problems, including communicable diseases such as AIDS/HIV infection, malaria, and tuberculosis. It may also be in part due to a lack of awareness about the magnitude of the current and future cancer burden among policy-makers, the general public, and international private or public health agencies interested in global health. Cervical cancer was the most frequently diagnosed cancer (61,510) and the leading cause of cancer death (40,600) in women in Africa in 2009 accounting for about 45% of the total new cancer cases and death (IARC ;WHO;Rebecca. S. *et al,* 2015).

The total cancer cases and death in Ethiopia estimated to be 19,659 and 64.12/100,000 respectively (World Bank, 2014). In 2011, it was estimated that 20.9 million women were at risk of developing cervical cancer in Ethiopia and the estimated annual number of CC cases and deaths was 4,648 and 3,235, respectively. It is projected that the number of new cervical cancer cases will almost double by 2025. Cervical cancer affects Ethiopia's most vulnerable women; poor, rural, HIV^{+} women and women with disabilities. The burden of cervical cancer, not only affects women, but also has a disastrous impact on their children, families, and communities (CSA, 2011).

Cervical cancer deaths in Ethiopia reached 4,595 or 0.76% of total deaths. The age adjusted death rate is 18.51 per 100,000 of population ranks Ethiopia (WHO, 2015).The age-adjusted incidence rate of cervical cancer in Ethiopia is 35.9 per 100,000 women. Despite this fact, very few women receive screening services. Although there is no national cancer registry, reports from retrospective reviews of biopsy results have shown that cervical cancer, followed by breast cancer, is the most prevalent cancer among women in the Ethiopia. HIV infection and cervical cancer are major public health problems facing women in Ethiopia. More than 500,000 women are estimated to be infected with HIV and at risk of developing cervical cancer. Though screening of precancerous cervical cancer lesion for HIV infected women has been started in limited centers in Ethiopia, data on the prevalence and factors associated with the lesion are lacking. Knowledge about prevalence and associated factors is needed to identify HIV-infected women who are more likely to

develop precancerous cervical lesion and to plan appropriate screening and treatment strategies (Wittet.S, Tsu, 2009).

As data from Tikur Anbessa Specialized Hospital Oncology Unit, more than 500 adult and pediatric cases with hematologic malignancies are seen in the hematology clinics every year. Many patients with cancer are also seen at the surgical, gastrointestinal and gynecology clinics. The most common adult cancers are cervical, breast, sarcomas, head and neck, and colorectal cancers, while leukemia, lymphoma, retinoblastoma and osteosarcoma constitute the bulk of pediatric cancers. The Hospital aspires to become a center of excellence in the diagnosis, treatment and care of patients with cancer. With the support of Ethiopia's governmental institutions, non-government organizations and international partners, it is hoping to develop a comprehensive cancer care program, including cancer registry, early detection, prevention, standard treatment and palliative care (ECA, 2010). As stated by study from September 2008 to September 2012 of 2,300 CC patients, 1,059 patients with standardized treatment were included. At the end of the study, 249 patients had died (Kantelhardt*et al*, 2014).

Survival models: Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event (Klembaum, 1996). Hence, survival analysis is also referred to as "time-to-event analysis", which is applied in a number of applied fields, such as medicine, public health, social science, and engineering. One of the oldest and most straightforward non-parametric methods for analyzing survival data is to compute the life table, which was proposed by Berkson and Gage (1950) for studying cancer survival. Kaplan and Meier (1958) obtained one important development in non-parametric methods. The non-parametric methods work well for homogeneous samples; they do not determine whether certain variables are related to the survival times.

This leads to the application of regression methods for analyzing survival data. The standard multiple linear regression models are not well suited to survival data for several reasons. First, survival times are non-normally distributed. Second, censored data result in missing values for the dependent variable (Klein, J. P., and Moeschberger, M. L, 1997). In survival analysis studies, characterizing the different survival distributions that correspond to different subgroups within a heterogeneous population, a descriptive

summary of such a comparison could consist of parametric or semi parametric methods (Hosmer&Lemeshow,1999; David G.; Kleinbaum; Mitchel K.,2005).

There are two major regression models used for survival data: proportional hazards model (Cox) as a semi parametric method (Cox. D, 1972) and accelerated failure time model or linear model representation in log time as a parametric model (Hosmer&Lemeshow, 1999). Cox proportional hazard function the model that assumes the underlying hazard rate is a function of the independent covariates, but no assumptions are made about the nature or shape of the baseline hazard function that is why Cox's model is referred to as a semi-parametric model for the hazard function (David G. Kleinbaum&Mitchel Klein,2005). However, Cox PH model has the restriction that proportional hazards assumption holds with time-fixed covariates; and it may not be appropriate in many situations and other modifications such as stratified Cox model or Cox model with time-dependent variables are required (Collett, D.,2003). The Accelerated Failure Time (AFT) model is another alternative method for the analysis of survival data. Many of the standard parametric models such as Weibull, Exponential, Lognormal and log logistic are accelerated failure time models (Klein, J. P., and Moeschberger, M. L, 1997). Although the Cox regression model is the most favorable employed technique in survival analysis, parametric models do have a number of benefits (Andersen P.et al, 1993).

1.2. Statement of Problems

Even though CC is fully preventable and curable, at low cost and at low risk for lack of early detection it is mostly affecting females in the current world. It is the third most common cancer in women worldwide however; the prognosis of advanced, recurrent or metastatic cervical cancer remains poor (Louie et al.; Ferlay *et al*, 2009).Cervical cancer is the leading cause of death among women and it is the most common cause of cancer-related morbidity and mortality. Current estimates indicate that every year 527,624 women are diagnosed with cervical cancer and 265,672 die from the disease Worldwide. The actual prevalence of cancer of the cervix is not known and many patients usually present very late to the health facilities (Parkin DM.*et al,* 2014).

Cervical cancer deaths in Ethiopia reached 4,595 or 0.76% of total deaths. The age adjusted death rate is 18.51 per 100,000 of population ranks Ethiopia (WHO, 2015).The age-adjusted incidence rate of cervical cancer in Ethiopia is 35.9 per 100,000 women

Many developed countries have reduced the incidence of the cancer and hence treatment burden through screening programmers. Unfortunately Ethiopia does not have a routine screening scheme like many other resourced-poor countries and consequently patients present at late stage which results into significantly high death related to cervical cancer (Bogalech.F,2015).

Although cervical cancer is a leading cause of cancer related morbidity and mortality among women in Ethiopia, there is lack of information regarding the perception of the community about the disease. Understanding the risk factors of women death due to cervical cancer is essential to inform public health policies and design strategies. In Ethiopia cervical cancer is the second most common cancer in women after breast and is one of the leading causes of cancer death among women (Abreham B., 2011).This is one of the challenging problems that the country needs to address. Many studies haven't been done regarding cervical cancer death in Ethiopia. Most of the previous studies on this area were considered on screening and recurrence of cervical cancer.

This study, therefore, has tried to fill the gaps in understanding the status of cervical cancer patients by identifying determinant risk factors of death due to cervical cancer in Ethiopia.

Generally, this study has attempted to answer the following basic research questions:

1. Which factors significantly affect the death rate of cervical cancer patients?
2. Is there difference in death rate of patients among regional states of Ethiopia?
3. How much is the mean survival time of cervical cancer patients?
4. Which model is the most appropriate for analyzing the predictors of death rate of cervical cancer patients?

1.3. Objectives of the study

1.3.1. General Objective

The General objective of the study is modeling time-to-death of women with cervical cancer in Tikur Anbessa Specialized Hospital.

1.3.2. Specific Objectives

1. To identify risk factors associated with mortality of women due to cervical cancer
2. To fit a statistical model that predicts the survival probability for women with cervical cancer based on the risk factors
3. To estimate the survival time and compare the survival curves of time-to-death among different levels of covariates

1.4. Significance of the study

This study can be used to assess the possible effect of different prevention and treatment programs and provide information about the covariates or risk factors of women death with cervical cancer. The results of this study also help in reducing the death of women by giving awareness for the society on the factors that increase the probability of women death due to cervical. Used as source of information for the government of Ethiopia, ministry of health, that enables policy makers to enhance the awareness of the society about factors that increase the probability of death due to cervical cancer which is protectable and curable if it is screened and treated in its earlier stage with appropriate treatment.

2. LITERATURE REVIEW

2.1. Overview of Cervical Cancer

Cervical cancer is one of the main causes of cancer-related death. It is the fourth most common cancer-related deaths in women, and the third most frequently diagnosed female cancer (Imam et al., 2009). As stated by World Health Organization (WHO) data nearly 2 million women in the world have a diagnosis of cervical cancer, 500,000 new cases are identified each year, and 274,000 deaths due to cervical cancer occur, and 80% of deaths due to cervical cancer occur in developing countries (WHO, 2009). Approximately 44% of cancer cases and 53% of cancer deaths occur in countries at a low or medium level of the Human Development Index (HDI) and international agency for research on cancer was estimate 1.547 million Cervix cancer ,1.893 million Lung cancer, 6.255 million Breast cancer, 3.544 million Colorectum cancer, 3.924 million Prostate cancer, 1.538 million Stomach cancer, 0.633 million Liver cancer, 0.464 million Oesophagus cancer, 1.320 million Bladder cancer and 11.427 Other cancer death in the world (WHO,2015, Moosa *et al.*, 2014).

According to data from the Ministry of Health in Turkey in 2010, cervical cancer is third most common cancer in female genital organs cancers with 1,951 cases and incidence rate of 4.5 per hundred thousand (TC. Saglık B., 2014). As risk factors of cervical cancer, human papillomavirus (HPV) infection, early sexual intercourse or marriage at an early age, multiple sexual partners, smoking, history of sexually transmitted diseases, long-term use of oral contraceptives, too many births, genetic predisposition and compromised immunity, history of infertility, poor hygiene, family history and low socioeconomic status are indicated (Kruiroongro J. *et al,* 2014).

Cervical cancer can be protected with an effective screening program, Papanicolaou (Pap) smear test is used is a cancer screening (Sogukpınar *et al.,* 2014; Karadag *et al.,* 2015). Pap smear results reduction of the risk of death from cervical cancer annually from 4/1000 to 5/10000 in women with early diagnosis (Ball and Madden, 2009). Cancer of the cervix is one of the most serious public health problems among Thai women (National Cancer Institute, 2014). It is now widely accepted that high risk types of human papillomavirus

(HPV), particularly HPV 16 and 18, play an important role in the genesis of cervical carcinoma. Most HPV infections in the cervix spontaneously resolve and few (not all) HPV infected females develop cervical cancer. Other risk factors such as exposure to certain carcinogens and host genetic predisposition likely have an influence on cervical carcinogenesis (Zur Hausen, 2004).

2.2. Risk factors for Death of Cervical Cancer

As other cancer types, also cervical cancer the onset death of the disease can be promoted by specific factors. Based on different literature some of main risk factors for the death of Cervix cancer are as follows:

2.2.1. Human Papiloma Viruses (HPV)

The main risk factor for the development of cervical cancer is human papilloma virus (HPV) infection, DNA of which has been found in almost all cases of invasive cervical cancer(Bosch et al, 2003). HPV is a sexually transmitted infection, making cervical cancer chronic diseases with an infectious etiology. At least 50% of sexually active men and women get HPV at some point in their lives. Most women with HPV infection will not develop cancer, and the infections usually resolve spontaneously however, around 3-10% of women with HPV develop persistent infections, and are at high risk of developing cervical cancer (Monsonego *et al.,* 2004).

HPV infection is an essential factor in the development of CIN and cervical cancer. When HPV acquisition is followed by HPV persistence instead of clearance, there is a high chance for progression to precancerous lesions and ultimately invasive lesions. HPV is a double stranded closed circular DNA virus with the capacity to incorporate in the human DNA). Disease with more than 80% of the population infected at some time in their life. HPV infection was also detected in the controls indicated that not all of the HPV-infected women developed cervical cancer or CIN. Among healthy women, an HPV infection was able to clear within 1-2 years and <1% of HPV-positive women would go on to develop cervical cancer (Nagpal *et al.,* 2012).

2.2.2. Age

Even if cancer starts from one cell, aging is another fundamental factor for the development of cancer. The study by identified that the age of onset of cervical cancer is higher in the age group of 41-60 accounting for 65% of the study population. The cervical

cancer patients with age above 50 years showed a 1.6 fold higher risk of failure for than patients with below 50 years of age with a statistical significance of (p=0.022). In assessing the prognostic factors for the death of cervical cancer records of patients age 60 and older with invasive cervical cancer treated with radiation with or without concomitant chemotherapy. According to survival predictors of outcomes analyzed using Kaplan Meier and Cox regression analysis. From the result 85 women with invasive cervical carcinoma were identified (with median age=68, median age at first intercourse=18, median of number of sexual partners=2, median number of pregnancy=7) the stage distribution of the women was IA2-IB2-35%, IIA,IIB-34%, IIIA-IIIB- 28%, IVA-3%.After a median follow up of 17 months (range 3-44 months), (32%) women died. Factors associated with death in the univariate analysis were smoking (P=0.015), radiation therapy dose (P=0.010), stage I-II Vs III-IV (P=0.024) (Kumari *et al*., 2013).Yancik.R **(**2012) and Wright JD. (2014) also concluded that increases in age have been related to poorer survival time of CC which may be attributable to treatment differences in cervical (older women were found to be less likely to receive more aggressive therapies).

2.2.3. Stage of the Disease

According to the international Federation of Gynecology & Obstetrics (FIGO) staging of cervical cancer is based on careful clinical examination and the results of specific radiologic studies and procedures. These should be performed and the stage should be assigned before any definitive therapy is administered. The clinical stage should never be changed on the basis of subsequent findings. When it is doubtful to which stage a particular case should be allotted, the case should be assigned to the earlier stage, at stage I cervical cancer can be treated with surgery, radiation, or chemotherapy. With stage II cervical cancer, the cancer has spread beyond the cervix and the uterus. Stage III cervical cancer has spread to the lower pelvic wall, and if the tumor grows large enough to block the uterus, the tubes that connect the kidneys and the 12 bladder, the kidney may not work properly. Chemotherapy and radiation are the treatment of choice for stage III cervical cancer. Stage IV is the most advanced stage of cervical cancer. In Stage IV cervical cancer, the cancer has spread to the nearby bladder or rectum more distant organs and the cancer is in its metastasis stage. Like stage III, stage IV treatment involves chemotherapy and radiation. Early stages of the disease are associated with a favorable prognosis; five-year survival rates for stage I disease are higher than 90%. Women diagnosed with more advanced disease, however, experience a considerably worse prognosis and less than 10%

survive stage IV disease (FIGO, 2009). Overall survival after 5-years for stage I, II & III was 81%, 33% and 0% respectively. Better survival was observed in patients with early stages of the diseases (Tigist Y., 2014).

The highest incidence of cervical cancer was reported in developing countries including the sub-Saharan Africa .In sub-Saharan Africa, the majority of cancers (over 80%) are detected in late stages, predominantly due to lack of information about cervical cancer and prevention services. This high incidence is attributed to the unawareness of the disease and inadequacy of screening programs in less developed countries. Cervical cancer is a key reproductive health problem for women particularly in the developing countries where screening services are lacking or inaccessible for the majority (Bingham *et al.,* 2013).

2.2.4. Smoking

Tobacco smoke contains more than 4,000 chemical substances including some carcinogens. Among these, polycyclic aromatic hydrocarbons (PAHs) and volatile Nitrosamines are considered to be the main carcinogens (Moosa *et al.*, 2014). Smoking exposure-a well-documented environmental factor-is a leading cause of many types of cancer such as cervical cancers, lung, esophageal, gastric, bladder and liver. Risk of cervical cancer associated with tobacco smoking has been found in many studies .Among HPV positive women, an increased risk of cervical cancer was demonstrated among smokers than non-smokers. Women who smoke are about twice as likely to develop cervical cancer as women who do not smoke (Yetimalar *et al.*, 2011).

2.2.5. Initial type of treatment

The initial type of treatment the women took is another significant risk factor for the death of cervical cancer. The risk of death following radical trachelectomy surgery is increased when the tumor size exceeds 2 cm. The histological type of tumor does not appear to be a significant risk factor for death except in the case of rare aggressive tumor with a propensity for rapid and widespread metastasis. These tumors are clearly associated with a higher risk of death and recurrence.

In assessing the risk, certain combinations of pathologic risk factors are thought to represent sufficient risk for death (Manjunath *et al,* 2010). Justified that the use of postoperative pelvic radiotherapy, though this has never been shown to improve overall survival. By compared adjuvant radiotherapy with no adjuvant radiotherapy meta-analysis

of these two treatments indicated as they had no significant difference in survival at five years between women who received radiation and those who received no further treatment (HR = 0.8, 95% CI : 0.3-2.4).

However, women who received radiation had a significantly lower risk of disease progression at five years (HR = 0.6, 95% CI 0.4- 0.9). The risk of serious adverse events was consistently higher if women received radiotherapy rather than no further treatment. Advanced stage at presentation (HR=3.0, P=0.001,) and a short disease-free interval after primary treatment (<8 months, HR=3.4, P=0.003) were determined as independent prognostic factors with a significant negative influence on progression-free survival and overall survival from initial diagnoses, respectively (Massad LS, 2008). Beside, from the study that compared the use of radiotherapy with no radiotherapy in women with early cervical cancer who had radical hysterectomy and pelvic lymph nodes (PLND) and who was at high risk of death. These trials showed that adjuvant radiotherapy after radical surgery significantly decreases local death rates, but provides only weak evidence that it might improve overall survival. When we combined the findings from these two trials, we found that, on average, the risk of death within five years among women who received radiotherapy was between 40% and 90% of the risk among women who did not (RR = 0.6, 95% CI 0.4 to 0.9) (Monsonego *et al*., 2004).

A study by (Erlap *et al*., 2006) identified that surgery and use of chemotherapy were each significantly associated with improved survival. Those with a longer disease-free interval (8 months vs. 8 months) from initial diagnosis to death and response to chemotherapy had a tendency for longer survival duration. Multivariate analysis revealed that progressive response to chemotherapy (HR=4.6, P=0.002,) and death within the previously Irradiated field (HR=2.7, P=0.04) were significant independent prognostic factors. Furthermore, advanced stage at presentation (HR=3.0, P=0.001) and a short disease-free interval after primary treatment (<8 months, HR=3.4, P=0.003) were determined. A study by (Wang SJ*et al.*,2010) that intended retrospectively to compare the treatment outcome of chemo radiation with that of radiation on risk factors for death which then was calculated by the Kaplan-Meier method and compared by the log rank test. The survival rate of the chemo radiation group was significantly higher than that of the radiation group (P<0.01). A study by (Yetimalar *et al.,* 2010) that intended to determine the survival rate of cervical cancer women and factors influencing survival, from the log rank test and the multivariate analysis were used. From the result of the study of age ≤50 and>50 and stage categorized

as I-II and III-IV were analyzed. The Log Rank test revealed that age and stage of the disease is not significant (P=0.77) and (P=0.177) respectively and from the multivariate Cox regression analysis the significance for the two categorical variables was (P=0.374) and (P=0.503) respectively.

2.2.6 Family History

Women who have a mother or sister diagnosed with cervical cancer have a greater risk of developing this cancer than women without a family history. (Bellinger. D, 2013; IARC, 2014).

2.2.7 Many Sexual Partners and Becoming Sexually Active at a Young Age

Women have had many sexual partners have a higher risk of becoming infected with HPV, which raises their risk of developing cervical cancer. There is also a link between becoming sexually active at a young age and a higher risk of cervical cancer. If a woman develops cervical cancer it does not mean she had several sexual partners, or became sexually active earlier than most other females. It is just a risk factor. Women who only ever had one sexual partner can develop cervical cancer (Brinton. L., *et al.,* 2002; Chelimo C, *et al.,* 2013).

2.2.8 Weakened Immune System:

People with weakened immune systems, such as those with HIV/AIDS, or transplant recipients taking immunosuppressive medications have a higher risk of developing cervical cancer (Abel G. et al,2013), (Massad LS ,2008).

2.2.9 Giving Birth at a Young Age

Women who gave birth before the age of 16 are significantly more likely to develop cervical cancer compared to women who had their first baby when they were aged 25 or over. Several Pregnancies: Women who have had at least three children in separate pregnancies are more likely to develop cervical cancer compared to women who never had children(Brinton L. *et al.,* 2002), (Chelimo C, *et al.,* 2013),(Ahmed Ib. *et al.,* 2011), (WHO, 2015).

2.2.10 Distance from the Hospital

Distance from the hospital most likely influence the patient's decision not to return to the hospital for continued care although there was no significant association between

defaulting and having a personal income probably because most women in this environment rely on their husbands or parents for payment of hospital (Sule ST; Shehu MS, 2007).The numbers of cervical cancer cases were not related to population size of the regions in Ethiopia. The number of cases in other region of the country depends on their distance from Addis Ababa; the farther the region, the fewer number of cases (Abate.S , 2015).

2.2.11 Abortion History

The rates of cervical, ovarian, breast and liver cancer are greatly increased by abortion. Some studies suggest that one abortion increases the chances of cervical cancer by 130 percent, while two increase the chances by 392 percent. The increases in cancer risks for women are due to the violent disturbance of the normal hormonal balances of pregnancy, as well as frequent cervical damage (R. Somers, 2001).

2.3. Perceived Severity of Cervical Cancer

Most women know that cervical cancer is a serious disease and studies on the perceived severity of cervical cancer have not been carried out in many developing countries. A survey on the perceived severity of cervical cancer among adult females in Quebec found that 57% of women were afraid of developing cervical cancer sometime in their life, and 93% thought developing cervical cancer has serious consequences (NC IPress O., 2012). Cervical cancer related anxiety and perceived seriousness did not vary by age group or level of education. Studies conducted among college women reported that, 98% of college women felt that cervical cancer is a very serious condition and half of them think that it is not a treatable disease (Ferlay J.,*et al.*,2009).Similarly Agency for Health Care Policy and Research(2009)found that 92% of women believed that cervical cancer is the second most serious type cancer a woman can have (first being breast cancer) and most women who develop cervical cancer certainly die from it.

2.4. Over View Survival Models

Proportional hazards modeling is the most frequently used type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (T. Ibrahim, 2009), affective disorders childhood family breakdown interruptions in conversation (Dress, 1986), and employee turnover and in medical areas for identification of important covariates that have as significant impact on the response of the interested variables.

Cox semi-parametric method has been extensively used for modeling survival data. Actually, in the sciences, researchers lean to use Cox semi-parametric method instead of parametric methods to analyze survival data due to fewer assumptions in the use of Cox semi-parametric model. In some circumstances, however, parametric methods can provide more accurate estimates. Parametric models such as Weibull are accelerated failure time models. Weibull allows more flexibility than the Cox semi-parametric model because the associated hazard rate is not constant with respect to time (Zali MR. *et al.,* 2011). The Cox Proportional Hazard model is the most popular technique to analyze the effects of covariates on survival time but under certain circumstances parametric models may offer advantages over Cox's model. The results of data analysis using parametric models are, similar to the Cox regression. Although the Hazard Ratio in Cox and parametric models are approximately similar but Weibull and Exponential are the most favorable for survival analysis of the data (Datwyler, *et al.,* 2011).

Moreover, characterizing the different survival distributions that correspond to different subgroups within a heterogeneous population is the objective of many studies. In a review of survival analyses in cancer journals, it was found that only 5 % of all studies used the Cox model with respect to checking the underlying assumptions. If this assumption does not hold, the Cox model can lead to unreliable conclusions. Therefore, the parametric models such as Lognormal, Weibull, Exponential and log logistic are the common options. These models provide the interpretation based on a specific distribution for duration times without need to proportional hazard assumptions (Pourhoseingholi *et al.,* 2011).

According to the retrospective study in women's records from Hospital in Malaysia, Factors that were considered in the analysis were ethnicity (non-Malay, Malay), lymph node involvement (negative, positive), metastasis (with metastasis, without metastasis), histology (squamous cell carcinoma, adeno cell carcinoma), primary treatment (surgery, radiotherapy and chemotherapy), age at diagnosis (<40, 40 – 59, ≥ 60) and stage (stage I & II, stage III & IV). The suitability of the Weibull model for the data was assessed using a plot of the log of the negative log of the estimated survivor function against log time or log-cumulative hazard plot.

Univariate analysis was conducted using the simple Weibull regression analysis to identify the possible prognostic factors individually. Significant factors from the univariate analysis were further analyzed by the Weibull multivariate analysis to model the prognostic factors. Model selection procedure was based on the forward variable selection method with statistical significance set at A = 0.10. The proportionality of the hazards was assessed using the test based on Schoenfeld .As the proportional hazard assumption was not satisfied, a stratified model was used instead and in measuring the goodness of fit of the model, deviance residuals were used. Survival data between groups were compared with the Log-rank test for univariate analysis and Cox regression analysis for multivariate analysis. The 90% confidence interval (CI) was calculated for the risk ratios for each of the significant prognosis (Chidiebere.M.I, 2009)**.**

Accelerated failure time models are an important alternative to the Cox proportional hazards model even though they have been rarely considered in the medical literature. (Chapman et al., 1992) applied four parametric survival models (Exponential, Weibull, Log logistic, and Log normal) to the effects of prognostic factors on breast cancer survival and concluded that the lognormal model provided the best fit to the data. (Royston, 2001) demonstrated the practical value of the lognormal AFT model in the analysis of survival times of breast and ovarian cancer patients. More recently, an AFT model has been implemented to analysis of the time to AIDS onset in the Women's Interagency HIV Study (Komarek *et al.,* 2004).

3. METHODOLOGY

3.1. Study Area

Tikur Anbessa Specialized Hospital is the biggest open Public referral Hospital in Ethiopia, with 800 beds, 201 doctors, 379 nurses, 115 other health professionals dedicated to providing health care services and 950 permanent and contract administrative staff to support the hospital activities. The hospital gives service to population of Addis Ababa city and its soundings majorly, but patients come to the hospital from all over Ethiopia. The hospital gives service as both inpatient and ambulatory follow up. The hospital's oncology unit is one of which give service to both inpatient and ambulatory patients department which has 19 beds. It is staffed with eleven Specialists, twenty two general practitioners, eight bachelor's degree nurses and fifteen diploma nurses (DHS, 2011).

3.1.1. Study Population

This is a retrospective study aims to determine the death pattern of CC based on hospital registry in TASH, Oncology Center. The population of this study was all CC patients who had been registered at TASH starting from 2003-2007 E.C. All the data were carefully reviewed from the registration log book and patients' registration card; if any inadequate information was countered it was checked from the file and excluded from analysis if proven to be inadequate. The cards are prepared by Federal Ministry of Health to be uniformly used by clinicians to early identify and document clinical and laboratory variables. Thus, the data were collected from patient follow up records based on the variable in the study. Data analyses were employed with the help of R statistical package and STATA software's.

3.1.2. Inclusion and Exclusion Criteria

Inclusion Criteria

All patients' registered with full information including in the registration log book or in the patients' identification card were considered to be eligible for the study. And also the patients should take cancer treatment at least one time in the Hospital.

Exclusion criteria

Patients with insufficient information about one of the vital variables either in the registration book or in the card were not eligible. Also the patients' lost from the study without starting any cervical cancer treatment was not included.

3.1.3. Data Collection Procedure

Ethical permission was obtained from the Oncology Department of the TASH. The hospital based data collected by trained enumerator and principal investigator. So this study incorporates secondary data. Data collection process was carried out in the time interval of 01-08- 2008 to 21-08- 2008 E.C.

3.2. Variables in the Study

3.2.1. Response Variable

The response variable for this study is time to death of the women with cervical cancer. In survival analysis, the outcome of interest (death in this study) is the duration of time until event (death) occurs. The status variable is coded as 0 for censored and 1 for death.

3.2.2. Explanatory Variable

The predictor variables which are assumed to influence the survival of cervical cancer patients are given in the table below:

Table 3.1: Description of independent variables used in the analysis

Age	**Region**	**Cycles of chemotherapy**
1=20-30	1=Addis abeba	1=No chemo
2=31-40	2=Oromiya	2=First cycle
3=41-50	3=Amhara	3=Second cycle
4=51-60	4=Tigray	4=Third cycle
5=61-70	5=SNNPR	5=Fourth cycle
6= >70	6=Others	
Number of children	**Age at marriage**	**Treatment taken**
1=No child	1=≤15	1=Surgery
2=1-4	2=16-20	2=Chemotherapy
3=5-8	3=21-25	3=Radiotherapy
4=9-12		4=Combination of two or
5=Above12		three
Stage	**Aim of radiotherapy**	**Sexual partner**
1=I	1=no RT	1=None
2=II	2=Palliative	2=One
3=III	3=Radical	3=Few
4=IV		4=Multiple
Age at first birth	**Smoking Status**	**Recurrence**
1=≤15	0=No	0=No
2=16-20	1=Yes	1=Yes
3=21-25		
4=>25		
5=Not give birth		
Family History	**Family planning**	**Abortion**
0=No	0=No	0=No
1=Yes	1=Yes	1=Yes
HIV status		
0=No		
1=Yes		

3.3. Methods of Data Analysis

3.3.1. Survival Data Analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. By time, mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs. By event, we mean death, disease incidence, relapse from remission, recovery or any designated experience of interest that may happen to an individual. The problem of analyzing time-to-event data arises in several applied fields such as medicine, biology, public health, epidemiology, demography and etc. (Aalen *et a.l*, 2008).

The use of survival analysis, as opposed to the use of other statistical method, is most important when some subjects are lost to follow up or when the period of observation is finite certain patients may not experience the event of interest over the study period. In this latter case one cannot have complete information for such individuals. These incomplete observations are referred to as being censored. Most survival analyses consider a key analytical problem of censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly. The censoring used in this paper is right censoring. Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study. If the event of interest is not occurred before the final day of the study such types of censoring is called right censoring.

In reality right censoring can occur due to the following reasons:

1. Death from unrelated causes
2. Loss of follow-up
3. Termination of study

The prospect of censoring complicates research design and complicates statistical analysis. Thus, censoring creates some unusual problem in the analysis of data because such data cannot be handled properly by standard statistical methods. Researchers will use different techniques to respond to the complication due to censoring unsatisfactorily. New developments in statistical theory accompanied by new development in statistical computing have changed how researchers can study such data.

The response variable in survival analysis is survival time and is no longer limited to only time to death. It is a non-negative random variable used loosely for the time period from a starting time point to the occurrence of any event. In this study context, survival times are the length of time of death after starting follow up in TASH.

3.3.2. Survival Functions

The survivor function is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time. Thus, it gives the probability that an individual surviving beyond a specified time. Moreover, the distribution of survival time is characterized by three functions:

i. The survivorship function
ii. The probability density function and
iii. The hazard function.

Let T be a random variable associated with the survival times, t be the specified value of the random variable T and f (t) be the underlying probability density function of the survival time T.

The cumulative distribution function F(t),which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given by;

$$F(t) = P(T < t) = \int_0^t f(u)du \ , t \geq 0 \quad (1)$$

The survivor function S(t), is given by;

$$S(t) = P(T \geq t) = 1 - F(t), t \geq 0 \quad (2)$$

From equations (1) and (2) the relationship between F(t) and S(t) can be derived as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\big(1 - S(t)\big) = \frac{-d}{dt}S(t) \geq 0 \quad (3)$$

3.3.3. Hazard Function

The hazard function h(t) the instantaneous potential for failing at time t, given that the individual has survived up to time t. In contrast to the survivor function, which focuses on failing, the hazard function focuses on not failing, that is, on the event occurring. (David et. al., 2005).

The hazard function $h(t) \geq 0$ is given as;

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{an individual fails in the time interval (t, t+\Delta t) given survives until time t\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P(t \le T \le t+\Delta t \mid T \ge t)}{\Delta t}$$

By applying the theory of conditional probability and the relationship in equation (3), the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln S(t) \quad (4)$$

The corresponding cumulative hazard function H(t), is defined as:

$$H(t) = \int_0^t h(u)du = -\ln S(t) \quad (5)$$

Then;

$$S(t) = \exp(-H(t)) \text{ and } f(t) = h(t)S(t) \quad (6)$$

3.4. Non-parametric Survival Methods

Nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution. Survival data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of survival time. Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e., if the estimated survival functions for two groups are approximately parallel (do not cross).

3.4.1. Kaplan-Meier Product Limit Method

The Kaplan-Meier (KM) estimator is the standard non parametric estimator of the survival function, S(t), proposed by Kaplan and Meier (1958) which is not based on the actual observed event and censoring times, but rather on the ordered in which events

occur. Kaplan-Meier estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t . The technique becomes a little more complicated but still manageable when censored times are included. Let ordered survival times are given by $0 \leq t_1 \leq t_2 \leq t_j \leq \infty$, then:

$$\hat{S}(t) = \begin{cases} 1, & if\, t < t1 \\ \prod_{j:t_j \leq t}[1 - \frac{d_j}{r_j}], & \text{if } t \geq t_1 \end{cases} \quad (7)$$

Where; d_j is the observed number of events at time t_j and r_j is the number of individuals at risk at time t_j. The Kaplan-Meier estimator, $\hat{S}(t)$ is a step function with jumps at the observed event times. The size of the jump at a certain event time t_j depends on the number of events observed a t , as well as on the pattern of the censored event times before t_j. The variance of the Product-Limit estimator is estimated by Greenwood"s formula (Greenwood, 1926), and is given by;

$$\text{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j \leq t} \frac{d_j}{r_j(r_j - d_j)};\ j = 1, 2, \ldots, r \quad (8)$$

Since the distribution of survival time tends to be positively skewed, the median is preferred for a summary measure. The median survival time is the time beyond which 50% of the individuals under study are expected to survive, i.e., the value of t_{50} at $\hat{S}(t_{50}) = 0.5$.The estimated median survival time is given by $t_{50} = \min\{t_i / \hat{S}(t) < 0.5\}$, where t_i is the observed survival time for the i^{th} individual, i=1,2,…, n. in general, the estimate of the p^{th} percentile is:

$$\hat{t}(p) = \min\{t_i / \hat{S}(t) < 1 - \frac{P}{100}\} \quad (9)$$

A confidence interval for the percentiles can be obtained using delta method (Hosmer&Lemeshaw, 1999).

The variance estimator for the p^{th} percetile is given by:

$$\text{Var}[\hat{S}(t_{(p)})] = (\frac{d\hat{S}(t(p)}{dt(p)})^2 \text{var}(t_{(p)}) = (-f(t_{(p)}))^2 \text{var}(t_{(p)}) \quad (10)$$

The standard error of $t_{(p)}$ is given by:

$$SE(\hat{t}_{(p)}) = \frac{1}{f(\hat{t}(p))} SE[\hat{S}(t_{(p)})]$$

The standard error of $\hat{S}(t_{(p)})$ can be obtained by using Greenwoods formula

$$\hat{f}((\hat{t}(p)) = \frac{\hat{s}(\hat{U}(p)) - \hat{s}(\hat{l}(p))}{\hat{l}(p) - \hat{U}(p)} \quad (11)$$

Where,

$$\hat{U}(p) = \max[S(t_j) \geq 1 - \frac{P}{100} + \delta]$$

$$\hat{l}(p) = \min[S(t_j) \leq 1 - \frac{P}{100} - \delta]$$

Where, $t_j$is the j^{th} ordered event time, j= 1,2,…, r.

Then, the 95% confidence interval for $t_{(p)}$is given by:

$$\hat{t}_{(p)} \pm 1.96 * SE(\hat{t}_{(p)}) \quad (12)$$

3.4.2. Log-Rank Test

The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for these groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally,

it could also be due merely to chance variation. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests. Among the various non-parametric tests one can find in the statistical literature, the Mantel-Haenzel (1959) test, currently called the "log-rank" is the one commonly used non-parametric tests for comparison of two or more survival distributions. The log rank test statistic for comparing two groups is given by:

$$Q = \frac{[\sum_i^m w_i(d_{1i} - \hat{e}_{1i})]^2}{\sum_i^m w_i^2 \hat{V}_{1i}} \sim \chi^2_{k-1} \quad (13)$$

where: $\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i}$ And $\hat{V}_{1i} = \frac{n_{1i} n_{0i} d_i (n_{1i} - d_i)}{n_i^2 (n_i - 1)}$

n_{oi} is the number at risk at observed survival time $t_{(i)}$ in group 0

n_{1i} is the number at risk at observed survival time $t_{(i)}$ in group 1

$n_i$is the total number of individuals or risk before time $t_{(i)}$

d_{1i} is the number of observed event in group 1

d_i is the total number of event at $t_{(i)}$

k is number of groups in each category

3.5. Semi parametric Comparison of Survival Distributions

Semi parametric survival model asks fewer assumptions than typical parametric methods but more assumptions than those nonparametric methods. In particular, and in contrast with parametric models, it makes no assumptions about the shape of the so-called baseline hazard function.

3.5.1. Cox PH Regression Model

The non-parametric method does not control for covariates and it requires categorical predictors. When we have several prognostic variables, we must use multivariate approaches. One very popular model in survival data is the Cox proportional hazards model. The Cox proportional hazards (PH) regression model (introduced in a seminal paper by Cox, 1972), a broadly applicable and the most widely used method of survival analysis. Survival models are used to quantify the effect of one or more explanatory

variables on failure time. This involves specification of a linear -like model for the log hazard. A model based on the exponential distribution may be parameterized as follows:

$$\log h_i(t/x) = \alpha + \beta_1 x_{i1} + \beta_{21} x_{i2} + ... + \beta_k x_{ik}$$

Equivalently;

$$h_i(t/x) = \exp(\alpha + \beta_1 x_{i1} + \beta_{21} x_{i2} + ... + \beta_k x_{ik}) = \exp(\alpha)(\beta' X)$$

In this case the constant α represents the log-baseline hazard since $\log h_i(t) = \alpha$ when all the x's are zero. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time.

$$\log h_i(t/x) = \alpha(t) + \beta_1 x_{i1} + \beta_{21} x_{i2} + ... + \beta_k x_{ik}$$

$$h_i(t/x) = h_0(t)\exp(\alpha + \beta_1 x_{i1} + \beta_{21} x_{i2} + ... + \beta_k x_{ik})$$

$$h_i(t/x) = h_0(t)\exp(X_i^T \beta) \quad (14)$$

Where, $h_0(t)$ is the baseline hazard function; X_i is a vector of covariates and β is a vector of parameters for fixed effects.

The corresponding survival function for Cox-PH model is given by:

$$s(t, X) = [S_0(t)]^{\exp\{\sum_{i=1}^{p} \beta_i X_i\}} \quad (15)$$

Where, $S_0(t)$ is the baseline survival function

In this model, no distributional assumption is made for the survival time; the only assumption is that the hazards ratio does not change over time (i.e., proportional hazards) that is why this model is also known as semi -parametric model. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients β , hazard ratio, and adjusted hazard curves.

If all of the x's are zero the second part of equation (12) equals 1 so, $h_i(t) = h_0(t)$. For this reason the term $h_0(t)$ is called the baseline hazard function. With the Cox proportional hazards model the outcome is described in terms of the hazard ratio.

The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates X and X^* is given by:

$$\hat{HR} = \frac{h_0(t)\exp(\hat{\beta}' X)}{h_0(t)\exp((\hat{\beta}' X^*)} \quad (16)$$

This hazard ratio is time-independent, which is why this is called the proportional hazards model. The parameter of the Cox proportional hazard model refers to the hazard ratio of one group in comparison to the other groups for categorical covariates and change in hazard ratio with a unit change of the covariate for the continuous variables when other covariates are fixed. The change in hazard ratio for the continuous covariate is given by:

$$\frac{h_i(t,x_k+1)}{h_k(t,x_k)} = \exp(\beta_k) \quad (17)$$

Which represent change in the hazard when there is a unit change in the covariates while other covariate keeps constant.

For categorical explanatory variable X with levels, the model contains $(a-1)$ dummy variables defined as $D_i = 1$, if $X = i$, and 0 otherwise for i = 1,2,…, a -1. Let $\beta_1, \beta_2, \ldots, \beta_{a-1}$ denote the coefficient of the levels of dummy variables. The ratio of the hazard of two subjects, one with X at level j and other with k (j, k = 1,2,…, a -1), provided that the values of all other explanatory variables for this subject are the same, the hazard ratio between these two categories is given by:

$$\frac{h(t/D_j)}{h(t/D_k)} = \frac{\exp(\beta_j)}{\exp(\beta_k)} \quad (18)$$

The quantity $\frac{\exp(\beta_j)}{\exp(\beta_k)}$ 100% signifies the ratio of hazard function for subject at level j and k of covariates, given that the effect of other covariate keeps fixed.

Partial Likelihood Estimation for Cox PH Model

Fitting the Cox proportional hazards model, we estimated $h_0(t)$ *and* β. A more popular approach is proposed by Cox (1972) in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters $h_0(t)$ in the Cox PH model. In this part, we construct the partial likelihood function based on the proportional hazards model.

The data in survival analysis based on the sample size n are denoted by the triplet (T_i,δ_i, X_i), i= 1, 2… n where Ti is the time at which the i^{th} individual experience the event (in this research context; death), δ_i = 1 if the event has occurred, $\delta_i = 0$ if censored, $X_{i\ is}$ the vector of covariate or risk factors for the i^{th} individual.

We assume;

a. Given Xi the life time and the censoring times are independent (non-informative censoring).
b. $\tau_1 < \tau_1 < \ldots < \tau_D$ be the D ordered distinct event times
c. There are no tied event times.

Let us define by;

1. I_j is the identity of the individual who died at time τ_j
2. V_j the time of the j^{th} failure at time τ_j and all information about censoring in $[\tau_{j-1}, \tau_j]$

The observable data (T_i,δ_i, X_i) is represented by $\{I_j\}$ and $\{V_j\}$. Hence;

$P(Data) = P(\{I_1,V_1,\ldots,I_D,V_D\})$

$$= P(\{I_1,V_1\}) \times P(\{I_2,V_2\}/\{I_1,V_1\}) \times \ldots \times P(\{I_D,V_D\}/\{I_1,V_1,\ldots,I_{D-1},V_{D-1}\})$$

$$= \prod_{j=1}^{D} P(I_j \mid I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j) \times P(V_j \mid \{I_1, V_1, \ldots, I_{j-1}, V_{j-1}\})$$

Due to the non-informative censoring, the second term does not add much information about the parametersβ.

Hence, we de fine the partial likelihood as;

$$L^{partial}(\beta) = \prod_{j=1}^{D} P(I_j \mid \{I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j\}) = \prod_{j=1}^{D} P(I_j \mid H_j)$$

Where, $H_j$is the "history" of the data, up to j^{th} failure and including the failure time, but not the identity of the failing.

At each failure, we note that the quantity $P(I_j \mid H_j)$ is the conditional probability that a specific individual fails at time τ_j given all the individuals that had not fail beforeτ_j.

We denote by R(t) the set of all the individuals under study just prior to time t.

$P(I_j \mid H_j)$= P(individuals I_j fails ׀ one individual fails in R(τ_j))

$$= \frac{P(\text{individuals Ij fails} \mid \text{at risk at } \tau_j)}{\sum_{l \in R(\tau_j)} P(individuallfails \mid atriskat\tau_j)}$$

$$= \frac{\lambda(\tau_j \mid X_j)d\tau_j}{\sum_{l \in R(\tau_j)} \lambda(\tau_j \mid X_j)d\tau_j} = \frac{\lambda_o(\tau_j)\exp(\beta^T X_j)}{\sum_{l \in R(\tau_j)} \lambda_o(\tau_j)\exp(B^T X_l)} = \frac{\exp(\beta^T X_j)}{\sum_{l \in R(\tau_j)} \exp(B^T X_l)}$$

We get the partial likelihood;

$$L^{partial}(\beta) = \prod_{j=1}^{D} \frac{\exp(\beta^T X_j)}{\sum_{l \in R(\tau_j)} \exp(B^T X_l)} \quad (19)$$

This is the partial likelihood defined by Cox. Note that, it does not depend on the underlying hazard function $h_0(.)$. Cox recommends treating this is as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $h_0(.)$

The likelihood function in equation (19) can be expressed by;

$$L^{partial}(\beta) = \prod_{j=1}^{D} \left[\frac{\exp(\beta^T X_j)}{\sum_{l \in R(\tau_j)} \exp(B^T X_l)} \right]^{\delta_i} \quad (20)$$

The partial likelihood given by equation (20), although it describes only part of the data, could be regarded as a likelihood function allowing the estimation of β with standard procedures.

In general, large sample properties like normality and consistency of maximum likelihood estimators of β based on partial likelihood have been shown to be the same as those of any estimator from complete likelihood (Hosmer&Lemeshow, 1999).

3.5.2. Accelerated Failure Time (AFT) Model

The accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models we measured the direct effect of the explanatory variables on the survival time instead of hazard. 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. (Kalbfleisch and Prentice, 2002).The AFT model states that the survival function of an individual with covariate X at time t is the same as the survival function of an individual with a baseline survival function at a

time t*exp(α'**X**), where α'= (α_1, α_2,…, α_p) is a vector of regression coefficients. In other words, the accelerated failure-time model is defined by the relationship:

$S(t \mid \mathbf{X}) = S_0\{t * \exp(\alpha'\mathbf{X})\}$, for all X (25)

Hereby we can consider on a log-scale of the AFT model with respect to time is given analogous to the classical linear regression approach. In this approach, the natural logarithm of the survival time Y =log (T) is modeled. This is the natural transformation made in linear models to convert positive variables to observations on the entire real line. A linear model is assumed for Y;

$Y = \log(T) = \mu + \alpha' x + \sigma\varepsilon$

where:$\alpha' = (\alpha_1, \alpha_2 \ldots \alpha_p)$ is a vector of regression coefficients

μ = intercept

σ= is scale parameter and

ε = is the error distribution assumed to have a particular parametric distribution. When we denote by S_0 the survival function when X = 0 then we find that

$P(T > t \mid X) = P(Y > \log(t) \mid X)$

$= P\{\mu + \sigma\varepsilon > \log(t) - \alpha' X \mid X\}$

$= P\{\exp(\mu + \sigma\varepsilon) > t * \exp(-\alpha' X) \mid X\}$

$= S_o\{t * \exp(-\alpha' X)\}$

The effect of the covariates on the survival function is that the time scale is changed by a factor exp $(-\alpha' X)$, and we call this an acceleration factor.

We note that when

exp $(-\alpha' X) > 1 \rightarrow$ the survival process accelerates.

exp $(-\alpha' X) < 1 \rightarrow$ the survival process decelerates.

If X is an indicator variable, this is equivalent to

$\alpha > 1 \rightarrow$ Time shrinks

$\alpha < 1 \rightarrow$ Time accelerates

For each distribution of ε there is a corresponding distribution for T. The members of the AFT model considered in this study are theWeibull AFT, Exponential AFT, Log- logistic AFT, and Log-normal AFT models. The AFT models are named for the distribution of T rather than the distribution of log T. This model can be related to the accelerated failure-time model representation (19) as in. The survival function of T_i can be expressed by

$S_i(t) = P(T_i \geq t)$

$= P(\log(T_i) \geq \log(t))$

$= P(Y_i \geq \log(t))$

$= P(\mu + \alpha' x + \sigma\varepsilon \geq \log(t))$

$= P\left(\varepsilon_i \geq \frac{logy-\mu-\alpha' x}{\sigma}\right) = S_{\varepsilon_i}\left(\frac{logt-(\mu+\alpha' x)}{\sigma}\right)$ (26)

Exponential AFT Model

For the time data and skewed to the right and with distribution of the time is exponentially, the time of survival for covariates matrix *X*, which is called, accelerated failure time, expressed as:

$$T = \exp(\beta' X + \varepsilon) \quad (27)$$

This model can be transformed by taking the natural log of each side of the equation as:

$$\ln T = \beta' X + \varepsilon \quad (28)$$

whereε is the error component and $\beta'=(\beta_o,\beta_1,\ldots,\beta_k)$.

The exponential model (t~exp (α)) is the simplest parametric model and assumes a constant risk or hazard over time, which reflects the property of the distribution appropriately called “lack of memory”. The survivorship function may be obtained by expressing in terms of time as:

$$S(t,) = \exp(-te^{-\beta' X}) \quad (29)$$

The hazard function of the Exponential regression model is:

$$h(t, X) = \exp(-(\beta' X)) \tag{30}$$

The exponential regression model for the k covariates and i^{th} individual is expressed as:

$$h_i(t, X_i) = h_0 \exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}) \tag{31}$$

For the Exponential regression survival models the hazard ratio, with one unit increase in covariate X_i while other covariates being held fixed, at a time t is $HR(X_i) = e^{-\beta_i}$.

Weibull AFT Model

The Weibull distribution (including the Exponential distribution as a special case) can be parameterized as an AFT model and they are the only family of distributions to have this property. The results of fitting a Weibull model can therefore be interpreted in either framework (Klein &Moeschberger, 2003). Then the Weibull distribution is very flexible model for time-to-event data. It has a hazard rate which is monotone increasing, decreasing, or constant.

From equation (19), the AFT representation of the survival and hazard function of the Weibull model with scale parameter and shape parameter is given by:

$$f(t,\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left(\left(-\frac{t}{\mu}\right)^{\alpha}\right), \text{ where } \mu > 0 \text{ and } \alpha > 0 \tag{32}$$

And the baseline hazard of this model for the i^{th} subject is

$$h_o(t_i;X) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \tag{33}$$

Independent observation (t_i,δ_i), $i=1,2,\ldots,n$ with survival time t_i, and censoring indicator δ_i which has value of one if i^{th} observation is not censored and zero when the i^{th} observation is censored and let β be the unknown parameter. The likelihood function is

$$L(\beta) = \prod_{i=1}^{n}\left\{f(t_i)^{\delta_i}\left(s(t_i)\right)^{1-\delta_i}\right\} = \prod_{i=1}^{n}\left\{\left(\frac{f(t_i)}{s(t_i)}\right)^{\delta_i} s(t_i)\right\} \tag{34}$$

Reparameterizing the Weibull distribution using $\lambda=\mu-\alpha$ then $h_o= \lambda\alpha t^{\alpha^{-1}}$ would be the baseline hazard function. Now incorporate covariates matrix X in the hazard function the Weibull regression model becomes:

$$h(t;X) = \lambda\alpha t^{\alpha^{-1}} exp(X\beta) \quad (35)$$

The model assumes that individuals' i and j with covariates $X_i and X_j$ have proportional hazard functions of the form:

$$\frac{h(t;X_i)}{h(t;X_j)} = \frac{\exp(X_i\beta)}{\exp(X_j\beta)} = \exp((X_i - X_j)\beta) \quad (36)$$

The quantities $exp(\beta)$ can be interpreted as hazard ratios.

A different parameterization is used with intercept v and scaled parameter σ and covariate effects γ_j having relationship with original parameterization as $\beta=-\gamma_j/\sigma$, $\alpha=\sigma^{-1}$ andμ=exp(v).

Log-logistic AFT Model

The log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump shaped that is it initially increases and then decreases. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions. According to the study of (Gupta *et* al., 1999), the Log-logistic distribution is proved to be suitable in analyzing survival data conducted by Cox (1972), Cox and Oakes (1984), Bennet (1983).The cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring (Bennett, 1983). The Log-logistic distribution is very similar in shape to the Log-normal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameters ¼ and ½, where ¼ is the scale parameter and ½ is the shape parameter (Bennett, 1983).

Its pdf is given by;

$$f(t) = \frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^{\rho})^2} \quad (37)$$

The corresponding survival and hazard functions respectively are given by;

$$S(t) = \frac{1}{1+\lambda t^{\rho}} \quad (38)$$

$$h(t) = \frac{\lambda \rho t^{\rho-1}}{1+\lambda t^{\rho}} \quad (39)$$

Where; $\lambda \in R, \rho > 0$

When $\rho \leq 1$, the hazard rate decreases monotonically and when $\rho > 1$, it increases from zero to its maximum point and then decreases to zero.

Suppose that the survival times have log-logistic distribution with parameter λ andρ, under the AFT model, the hazard function for the i[th] individual is

$$h_i(t/x) = h_o\left(texp(-\alpha' x_i)\right)\exp(-\alpha' x_i) = \frac{\rho \exp((\lambda)\text{texp}(-\alpha' x_i)}{1+\exp(\lambda)\{\text{texp}(-\alpha' x_i\}^{\rho}} \quad (40)$$

The Log-logistic AFT model with a covariate x may be written as;

Y= logT= $\mu+\alpha' x_i + \sigma\varepsilon$, where; $\alpha' = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_p)$; εhas standard logistic distribution. The survival with covariate x is given as follows:

$$S_T(t/x) = \frac{1}{1+\lambda\exp(\beta' x)t^{\rho}} = \frac{1}{1+\exp(log\lambda+\beta' x)} \quad (41)$$

$$h_T(t/x) = \frac{\rho t^{\rho-1}\lambda\exp(\alpha' x)}{1+\lambda\exp(\alpha' x)t^{\rho}} = \frac{\rho t^{\rho-1}\lambda\exp(\alpha' x)}{1+\exp(log\lambda+ \alpha' x)} \quad (42)$$

To interprete the factor exp($\beta' x$) for log-logistic model, one can notice that the odds of survival beyond time t for log-logistic model is given by $\frac{S_T(t)}{1-S_T(t)}$.

We can see that the log-logistic distribution has the proportional odds (PO) property. So this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as:

$$\frac{S_T(t)}{1-S_T(t)} = \exp(-\alpha' x)\frac{S_0(t)}{1-S_0(t)} \quad (43)$$

The factor exp ($-\alpha' x$) is an estimate of how much the baseline odds of survival at any time changes when individual has covariate x. And exp ($\alpha' x$) is the relative odds of experiencing the event for an individual with covariate x relative to an individual with the baseline characteristics. As this representation of Log-logistic regression is as accelerated

failure time model with a log logistic baseline survival function, then the Log logistic model is the only parametric model with both a proportional odds and an accelerated failure-time representation.

If T_i has a Log-logistic distribution, then ε_i has a Logistic distribution. The survival function of Logistic distribution is given by:

$$S_{\varepsilon_i}(\varepsilon) = \frac{1}{1+\exp(\varepsilon)} \tag{44}$$

Then, the AFT representation of Log-logistic survival function is given by

$$S_t(t) = \left[1 + t^{\frac{1}{\sigma}}\exp\left(\frac{-\mu-\boldsymbol{\alpha}'\boldsymbol{x}}{\sigma}\right)\right]^{-1} \tag{45}$$

And the associated hazard function for the i^{th} individual is given by

$$h_t(t) = \frac{1}{\sigma t}\left[1 + t^{\frac{-1}{\sigma}}\exp\left(\frac{-\mu-\boldsymbol{\alpha}'\boldsymbol{x}}{\sigma}\right)\right]^{-1} \tag{46}$$

If the plot of $\log\left[\frac{1-S(t)}{S(t)}\right]$against log(t) is linear , the Log-logistic distribution is appropriate for the given data set.

Log-normal AFT Model

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

$$S_o(t) = 1\text{-}\phi\left(\frac{\log t-\mu}{\sigma}\right),\ h_o(t) = \frac{\phi\left(\frac{\log t}{\sigma}\right)}{\left[1-\emptyset\left(\frac{\log t}{\sigma}\right)\right]\sigma t} \tag{47}$$

Where μ and σ are parameters, $\phi(x)$ is the probability density function and $\phi(x)$ is the cumulative density function of the standard distribution. The survival function for the i^{th} individual is

$$S_i(t) = S_o(t/\eta_i) = S_o(t * \exp(\mu + \alpha' x_i)) = 1\text{-}\ \phi\left(\frac{\log t-\alpha' x_i-\mu}{\sigma}\right) \tag{48}$$

Where $\eta_i = \exp(\alpha_1 x_1 + \alpha_2 x_2 \ldots + \alpha_p x_p)$. Therefore the log survival time for the i^{th} individual has normal $(\mu + \alpha' x_i, \sigma)$. The log normal distribution has the AFT propoerty.

In a two group study: we can easily get

$\phi^{-1}[1 - S(t)] = \frac{1}{\phi}(logt - \alpha' x_i - \mu)$, where x_i is the value of a categorical variable which takes the value 0 in one group and 1 in the other group. This implies that the plot $\phi^{-1}[1 - S(t)]$ against log (t) will be linear if the lo-normal distribution is appropriate for the given data.

Parameter Estimation

Parameters of AFT models can be estimated by maximum likelihood method. The likelihood of n observed survival times, t_1, t_2, $t_{3\ldots}t_n$, the likelihood function for right censored data is given by:

$L(\alpha, \mu, \sigma) = \prod_{j=1}^{n} f_i(t_i)^{\delta_i} * S_i(t_i)^{1-\delta_i}$(49)

Where $f_j(t_j)$ the density function of the i^{th} individual at time t_i, $S_i(t_i)$ is the survival function of the i^{th} individual at time t_i, δ_i is indicator variable. The logarithm of the above equation yields;

$\log L(\alpha, \mu, \sigma) = \sum_{j=1}^{n}\{-\delta_i log(\delta t_i + \delta_i log f_i(x_i) + (1 - \delta_i) log S_i(W_i))\}$ (50)

Where, $W_j = \left\{logt_i - \frac{\mu + \alpha_{1i}x_i + \cdots + \alpha_{pi}x_{pi}}{\delta}\right\}$, Z= $\{z_{ji}\}$ is vector of covariates for the j^{th} subject. The maximum likelihood parameters estimates are found by using Newton-Raphson procedure

3.4.7. Model Development

The methods of selecting a subset of covariates in a PHs regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections. Survival analysis using Cox regression method begins with a thorough univariate analysis of the association between survival time and all important covariates (Hosmer and Lemeshow, 1999)

Recommendable procedure in selecting variables in the study

According to Hosmer and Lemeshow (1998), it is recommended to follow the steps given below.

1. Include all variables that are significant in the univariate analysis at relaxed level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.

2. The variables that appear to be important from step one are then fitted together in a model. In the presence of certain variables others may cease to be important. As a result, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

3. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method. This process may result in terms in the model determined at step 2 ceasing to be significant.

3.5. Model Selection

Akaikie (1974) proposed an information criterion (AIC) statistic to compare different models and/or models with different numbers of parameters. For each model the value is computed as:

$$AIC = -2LogL+2(k+c+1) \quad (51)$$

Where k is the number of covariates and c the number of model specific distributional parameters. The preferred model is the one with the lowest value of the AIC.

3.6. Model Assessment

Every step during model fitting uses the upcoming statistical procedures and later at the end we will check all the assumptions needed for the model. Some of the statistical procedures that are used to assess the final model are:

3.6.1. Testing the Assumption of PH Model

The proportional hazards assumptions are vital to use in a fitted proportional hazards model. Variable adds significant information. If the newly added variable is not significant, it can be taken as the assumptions of the proportional hazard assumptions are satisfied. The method of checking the assumption of the Cox proportional hazards model is scatter plots using the Schoenfeld residual (Schoenfeld, 1982). The residuals constructed for each covariate that are included in the model which are expected to predict the death time of Women with cervical cancer. Under the proportional hazard assumption for the respective covariate, a scatter plot of Schoenfeld residuals against event times is expected to scatter in a nonsystematic way about the zero line, and the polygon (Lowess curve)

connecting the values of the smoothed residuals should have a zero slope and cross the zero line several times (Klein &Moeschberger., 2003). If this plot shows some trend the assumption is violated, where as if the plot demonstrates randomly distributed around the reference line then the assumption is satisfied.

3.6.2. Checking the Adequacy of Parametric Baselines

The graphical methods can be used to check if a parametric distribution fits the observed data. Model with the Weibull baseline has a property that the log (-log(S(t)) is linear with the log of time, where $S(t) = \exp(-\lambda t^{\rho})$. Hence, $\log(-\log(S(t))) = \log(\lambda) + \rho\log(t)$. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting $\log(-\log(\hat{S}(t)))$ versus log(t) where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler and Stucki, 2009).

The log-failure odd versus log time of the Log-logistic model is linear. Where the failure odds of Log-logistic survival model can be computed as:

$$\frac{1-s(t)}{s(t)} = \frac{\frac{\lambda t^{\rho}}{1+\lambda t^{\rho}}}{\frac{1}{1+\lambda t^{\rho}}} = \lambda t^{\rho}. \quad (52)$$

Therefore, the log-failure odds can be written as:

$$\text{Log}\left(\frac{1-S(t)}{S(t)}\right) = \log(\lambda t^{\rho}) = \log(\lambda) + \rho\log(t). \quad (53)$$

Therefore the appropriateness of model with the Log-logistic baseline can graphically be evaluated by plotting $\log\left(\frac{\hat{S}(t)}{1-\hat{S}(t)}\right)$ versus log(time) where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler and Stucki, 2011). If the plot is straight line, log-logistic distribution fitted the given dataset well. If the plot $\emptyset^{-1}[1 - S(t)]$against log (t) is linear, the Lo-normal distribution is appropriate for the given data set.

3.6.3. Quantile Quantile Plot

A quantile-quantile or q-q plot is made to check if the accelerated failure time model provides an adequate fit to the data. The plot is based on the fact that, for the accelerated failure-time model,

$$S_1(t) = S_0(\phi\, t) \quad (54)$$

Where S_0 and S_1 are the survival functions in the two groups and ϕ is the acceleration factor. Let t_{op} and t_{1p} be the p^{th} percentiles of groups 0 and 1, respectively, that is:

$$t_{kp}=s_k^{-1}(1-p),\ k=0,1. \quad (55)$$

Using the relation $S_1(t) = S_0(\phi\, t)$, we must have $s_o(t_{op}) = 1-p = s_1(t_{1p}) = s_o(\phi\, t_{1p})$ for all t. If the accelerated failure time model holds, $t_{op} = \phi\, t_{1p.}$ To check this assumption we compute the Kaplan–Meier estimators of the two groups and estimate the percentiles t_{1p}, t_{0p}, for various values of p. If we plot the estimated percentile in group 0 versus the estimated percentile in group 1 (i.e., plot the points t_{1p}, t_{0p} for various values of p), the graph should be a straight line through the origin, if the accelerated failure time model holds. If the curve is linear, a crude estimate of the acceleration factor ϕ is given by the slope of the line (Klein, 2003).

3.6.4. Using Residual Plots

For the parametric regression problem, analogs of the semi parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein, 2003). The first such residual is the Cox–Snell residual that provides a check of the overall fit of the model.

The Cox–Snell residual, r_j, is defined by:

$$r_j = \hat{H}(T_j \mid X_j) \quad (56)$$

where $\hat{H}$ is the cumulative hazard function of the fitted model. If the model fits the data, then the r_j's should have a standard (λ =1) exponential distribution, so that a hazard plot of r_j versus the Nelson–Aalen estimator of the cumulative hazard of the r_j's should be a straight line with slope 1.

3.7. Ethical Consideration

The Research Ethics Review Board of Jimma University has provided an ethical clearance for the study. The data was brought from Tikur Anbessa Specialized Hospital, and to do so the department of statistics asked to write an official co-operation letter to the Hospital from where data was obtained. The study conducted without individual informed consent because it relied on retrospective data. In this research, the information obtained from log book and patients' card kept secured.

4. RESULTS AND DISCUSSION

4.1. Descriptive Statistics

The study intended to find the determinant risk of the time to death of women with cervical cancer at TASH on those patients who started their treatment from the year 2003 E.C and followed up to 2007. The time interval between screening and death was an interest of this research paper. The minimum observed event time was 6 months and the maximum was 60 months. In this study only those who took the cervical cancer treatment at least one time in the hospital were included. A total of 907 cervical cancer patients fulfilling the inclusion criteria were considered. After the medical cards of women were reviewed among those patients of cervical cancer 349 (38.48%) died and the remaining 558 (61.52%) were censored

As shown in Table 4.1 in appendix of total patients 33.19%, 32.75%, 21.28%, 3.42%, 6.5% and 2.87%lived in Addis Ababa, Oromiya, Amhara, Tigray, SNNP and others regions respectively. The death proportion of patients who lived in Addis Ababa, Oromia, Amhara, Tigray, SNNPR and others region were41.83%, 27.79%, 18.62%, 2.58%, 5.73% and 2.87% respectively. Similarly in considering age groups of the patients, 3.2%,16.87%,28.34%,31.97%,16.43% and 3.2% were in the age groups of 20–30, 31–40,41-50, 51-60, 61-70 and > 70 respectively. The death proportion of the patients for the age classes 20-30, 31–40, 41-50, 51-60, 61-70 and > 70 were 2.58%, 8.60%, 15.19%, 44.41%, 25.79% and 3.44% correspondingly. Of the total patients 13.56%were smokes cigarette. Death proportion of smoker's was 23.21%. In this study 14.33% of patients experienced recurrence of cervical cancer. The death proportions for those patients experienced recurrence of cervical cancer was 19.77%.

Out of the total patients, 5.4% were at stage I, 30.54% were at stage II, 53.14% were at stage III and 11.14% patients were at stage IV. The death proportions of stage I, stage II, stage III and stage IV patients were 5.44%, 25.79% 54.15% and 14.65% in that order. In parity case, there were 1.43%, 43%, 23.93% and 31.64% women were with no child, 1-4 children, 5-8 children, 9-12 children and >12 children respectively. The death proportion among no child,1-4 children, 5-8 children, 9-12 children and above 12 children were 0.86%, 26.07% , 23.21% and 49.86%, respectively. From those patients who came to be treated 0.66% have not took radiotherapy, 65.49% took palliative radiotherapy and 33.85% took radical radiotherapy treatment of this 0.36%, 68.77%, 30.09% died

correspondingly. Among the women were included in the study, 0.55% took surgery, 0.55% took chemotraphy, 76.85% took radiography and 22.05% took Combination of two or three. The death proportion of patients who took surgery, chemotraphy, radiography and Combination of two or three were 1.15%, 0.57%, 74.21% and 24.07% in that order. Furthermore, 29% of the women had experience of using contraceptive. The death proportion for patients had family planning history was 32.66%.

Besides out of the total patients 26.65% had history of abortion. The death proportion of patients who had history of abortion was 26.65%. Moreover, 28% patients were living with HIV AIDS. Death proportion of HIV positive patients were 33.81%.Considering the women's age at marriage60.75%, 35.39% and 3.86% patients were married at age ≤15, 16-20 and 21-25 years respectively. The death proportion of the married patients at age ≤15, 16-20 and 21-25 years were 67.62%, 37.63% and 0 .57% respectively. Based on Table 4.1 in the appendix 39.25%, 44.87%, 12.24% and 1.32% patients gave first birth at age of ≤15, 16-20,21-25and >25 respectively while 2.32% patients not giving birth up to the study ends. The proportion of death were 48.14% for those women give birth at age of ≤15 years, 38.97% for those women give birth at age of 16-20 years,9.74% for those women give birth at age of 21-25, 1.72% for those give birth at age of >25 and 1.43% for those women not giving birth until the study ends.

4.2. Non-parametric Survival Analysis

4.2.1. The Kaplan- Meier Estimate of Time-to-death of CC Patients

Non-parametric survival analysis is very important to visualize the survival of time of cervical cancer patients in TASH under different levels of the covariates. Moreover, it gives information on the shape of the survival and hazard functions of cervical cancer patients' data set. Survival time distributions for time-to-death is estimated for each group using the K-M method and in order to compare the survival curves of two or more groups, log-rank test has been employed.

The estimated mean survival time and 95% confidence interval for survival time of women with cervical cancer with different covariates characteristics are summarized in table 4.2 in the appendix. The mean survival time of cigarette smoker women were 16 months with [95% CI: 13.70, 27.83] which was less than that of non-smoker women35months with its [95% CI: 31.00, 42.68]. The mean survival time for patients had no family history of cervical cancer was greater than had family history 33 and 24 months

respectively. The mean survival times of patients live with HIV were less than HIV negative patients which is24 and 38 months respectively. The overall mean survival time of cervical cancer patients in TASH is 26 month with 95% CI [24.01, 27.34].

Table 4.3: Log Rank Tests of Each Covariate

Covariates	Chi square	Df	P-values
Age	101.83	5	0.00
Smoking	49.88	1	0.00
Region	29.32	5	0.00
Recurrence	8.45	1	0.004
Stage	65.08	3	0.00
cycles of chemo	8.89	4	0.07
Sexual partner	70.29	3	0.00
Aim of radiotherapy	31.80	2	0.00
Family history	26.98	1	0.00
Treatment taken	0.79	3	0.85
Number of children	56.67	4	0.00
Family planning	5.21	1	0.22
Abortion	13.52	1	0.02
HIV status	12.59	1	0.04
Age at marriage	15.04	2	0.05
Age at first birth	13.24	4	0.01

DF=degree of freedom

4.2.2. Survival Time-to-Death for Different Groups of Covariates

The Kaplan-Meier survivor estimators for age category plotted in Figure 4.1 in Appendix shows patients with age group 31-40 and 41-50 have better survival prognosis when compared to other age group. The log-rank test also illustrated that, there were significant differences among age categories with respect to survival probability (p=0.00). According to the log rank test in Table 4.2 regions were statistically significant in estimating the survival time of cervical cancer. Similarly KM curves Figure 4.1 appendix, showed patients from Tigray and Oromiya had better survival prognosis when compared to the remaining Regions.

According to Figure 4.2, survival probability was higher for patients who had no experienced recurrence than those had experienced recurrence. Also the log-rank test in

Table 4.2 demonstrated significant difference between patients who had experienced recurrence of cervical cancer and do not experienced (p=0.0036) at 5% level of significance.

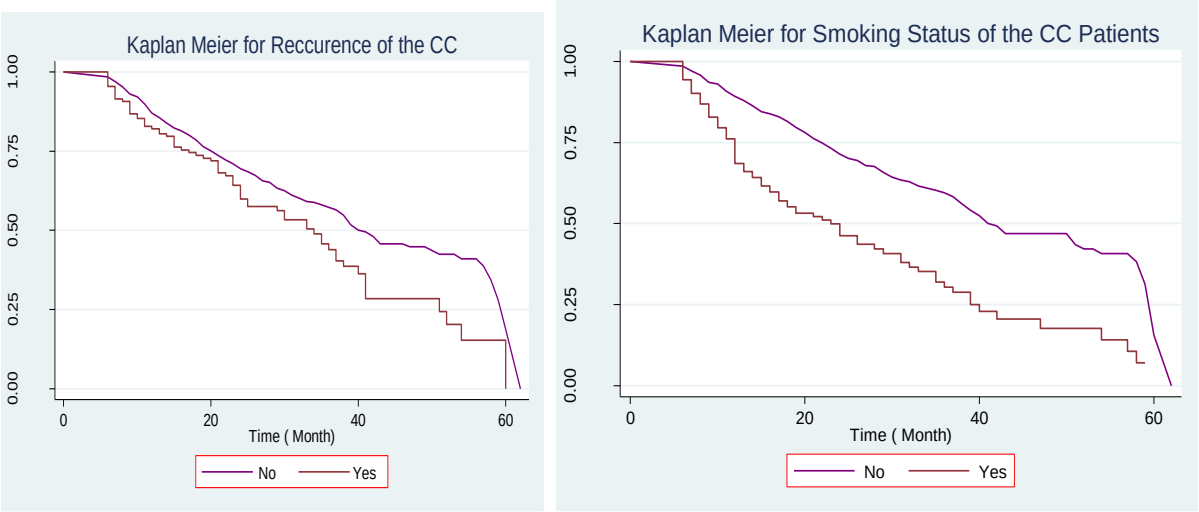

Figure 4.2: K-M survival time plot by Recurrence and Smoking status of CC patients

KM curve in the above Figure indicates nonsmokers had higher survival time than smoker. The log rank test also revealed that smoking cigarettes had significant association with survival time of women (p=0.00) at 5% level of significance Similarly Figure 4.3emphasized that the survival of time of stage III and stage IV women is shorter than Stage I and Stage II women. The result of the log rank test is also revealed the difference is significant (p=0.00) at 5% level of significance.KM curve for number of sexual partner in appendix showed patients had only one sexual partner survived batter than patients had ≥1 sexual partners. Statistical test using log-rank in Table 4.2 also shows that difference was significant (p=0.00) at 5% level of significance.

The other categorical variable included in the study was treatment taken; a Kaplan-Meier survivor estimate for this covariate was plotted in Figure4.4 in appendix. This Figure implied that the risk of death due to cervical cancer for women who took any treatment significantly had a similar survival probability (no survival difference). The log rank test in Table 4.2 also demonstrate treatment taken had no significant association to survival time of women (p=0.853) at 5% level of significance.KM curve for number of children in appendix estimates the patients who had 1-4 and 5-8 children were slightly higher than other. The log rank test in Table 4.2 also demonstrate number of children had significant association to survival time of women (p=0.00) at 5% level of significance.

The survival time plot by aim of radiotraphy is given in Figure 4.5. This plot showed that the risk of death of women with cervical cancer is similar for all group for the first few months. But the a little difference becomes visible at the middle of the curve and becomes similar at the end of the curve. The result of the log rank test Table 4.2also support the significance of this difference (p=0.00) at 5% level of significance. According to the survival time plot by chemotraphy cycles Figure 4.5 (in the appendix) the risk of death for women who had different cycles of chemotraphy were the same. The log rank test in Table 4.2 also revealed that Chemotraphy cycles had no significant association to survival time of patients (p=0.0638) at 5% level of significance. Comparing the survivor function between family histories of cervical cancer, Kaplan-Meier survival estimates for the two groups are plotted in Figure 4.6.This Figure shows that patients who had no family history of cervical cancer had higher survival probability than those who had family history of cervical cancer. The log rank test in Table 4.2 also demonstrated that family history had significant association with survival time of patients (p=0.00) at 5% level of significance.

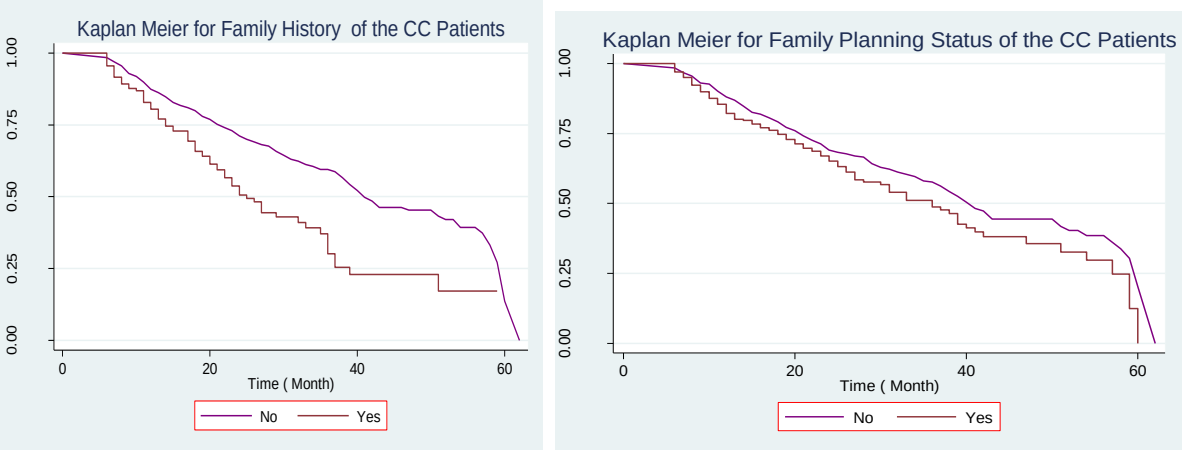

Figure 4.6: K-M plot of survival time by Family History and Family Planning Statusof cervical cancer patients

The survival time plot by family planning status is given in Figure 4.6. This plot showed that the risk of death is slightly different for patients who had history of family planning and no family planning history. But the log rank test in Table 4.2 demonstrated that family planning had no significant association to survival time (p=0.224) at 5% level of significance. The survival time plot of by history of abortion was also given in Figure 4.7 in appendix. This plot showed that the risk of death is different for patients who had history of abortion and had no abortion history. The log rank test also revealed that

abortion history had significant association with survival time of women (p=0.0002) at 5% level of significance.

The survival time plot by HIV status is given in Figure 4.7 in appendix. This plot showed that the risk of death was different for patients who were living with HIV AIDS and free from HIV. The log rank test in Table 4.2 also showed HIV status was significant association with survival time of women (p=0.04) at 5% level of significance. The survival time plot by age at marriage is given in Figure 4.8 (in appendix). This plot showed that the risk of death of women slightly different for all group for the first few months. The result of the log rank test in Table 4.2 also support the significance of this difference (p=0.05) at 5% level of significance. The survival time plot by age at first birth is given in Figure 4.8 in appendix. This plot showed that the risk of death is slightly similar for all age group patients. But the log rank test in Table 4.2 demonstrated that age at first birth had significant association to survival time of women (p=0.01) at 5% level of significance.

4.3. Semi-Parametric Cox PH

4.3.1. Univariate Analysis

In any data analysis it is always a great idea to do some univariate analysis before proceeding to more complicated models. Single covariate Cox proportional hazards model analysis is an appropriate procedure that is used to screen out potentially important variables before directly included in the multivariate model. The relationship between each covariates and survival time of cervical cancer patients are presented in table 4.3in appendix. As shown from this table, survival of the patients is significantly related with age, smoking, region, recurrence, stage, family history, number of children, number of cycles patients took chemotraphy, family planning, abortion history, and HIV status, age at marriage and age at first birth at 10% level of significance.

4.3.2. Multivariate Analysis

Result presented in Table 4.4 indicate the parameter estimates of coefficients for the covariates in the final model along with the associated standard error, Wald statistic, significance level, hazard ratio and 95% confidence interval for the hazard ratio. According to Table 4.3 in appendix the predictor's such as age, smoking, region, recurrence, stage, family history, number of children, number of cycles patients took

chemotraphy, family planning, abortion, HIV status, age at marriage, age at first birth passed the first filtration of variables for multiple covariates analysis and then forward Stepwise variable selection method was used to select the important variables to be included in Cox proportional hazards model. In order to decide whether or not a variable is significant, the p-value associated with each parameter has been estimated and variables that have p-value less than or equal 0.05 cut point or 5% significance level are considered as important variables and hence, are included in the final model.

Table 4.4: Multivariate Analysis of Cox Proportional Hazards with time to event of patients.

Covariates	**Category**	$\hat{\beta}$	**SE**	**Wald**	**Sig.**	**HR**	**95% CI for HR**
	20-30	Ref					
Age	31-40	0.60	0.39	2.38	0.123106	0.55	[0.256, 1.177]
	41-50	-0.50	0.37	1.80	0.180234	0.61	[0.291, 1.261]
	51-60	0.74	0.35	4.36	0.036667*	2.09	[1.047, 4.182]
	61-70	1.02	0.36	7.94	0.004842**	2.77	[1.364, 5.626]
	>70	0.62	0.45	1.84	0.04631*	1.85	[1.761, 4.498]
Smoking Status	nonsmoker	Ref					
	Smoker	0.67	0.14	22.8	1.79e-06 ***	1.96	[1.485, 2.574]
	Addis Ababa	Ref					
	Oromiya	-0.50	0.14	13.88	0.000262***	0.61	[0.468, 0.795]
	Amhara	-0.29	0.15	3.53	0.060420.	0.75	[0.554, 1.013]
Region	Tigray	-0.63	0.36	3.16	0.075604.	0.53	[0.264, 1.068]
	SNNPR	-0.30	0.25	1.47	0.225600	0.74	[0.457, 1.203]
	Others	-0.39	0.31	1.63	0.201729	0.68	[0.370, 1.233]
	I	Ref					
	II						
	III	0.13	0.26	0.26	0.610659	1.14	[0.683, 1.913]
	IV						
Stage		1.06	0.26	16.96	3.75e-05***	2.89	[1.743, 4.774]

		1.78	0.29	38.24	6.26e-10**	5.91	[3.366,10.384]
Family History	No	Ref					
	Yes	0.60	0.14	19.15	1.21e-05***	1.82	[1.390, 2.370]
Abortion	No	Ref					
	Yes	0.67	0.13	27.38	1.67e-07***	1.96	[1.521, 2.514]
HIV status	No	Ref					
	Yes	0.44	0.12	13.12	0.000292***	1.55	[1.222, 1.961]
age at first birth	≤15	Ref					
	16-20	-0.22	0.12	3.33	0.074422·	0.81	[0.636, 1.022]
	21-25	-0.08	0.20	0.12	0.669228	0.92	[0.625, 1.352]
	>25	1.01	0.43	5.63	0.017071*	2.76	[1.198, 6.334]
	Not give birth	-1.38	0.46	8.33	0.002913**	0.25	[0.102, 0.625]

SE=Standard Error, DF=Degree of freedom, HR= Hazard Ratio, CI=Confidence Interval, Ref. Reference

In final model the survival time of women with CC was significantly affected by age, smoking status, region, stage, family history, abortion, HIV status and age at first birth. The values of the Wald statistic for individual $\hat{\beta}_i$ coefficients support that the estimated values $\hat{\beta}_i$ were significantly different from zero at α= 5% level of significance.

4.3.3. Statistical Tests for Proportional Hazards Model Assumptions

Goodness of fit testing approach is appealing because it provides a test statistic and p-value for assessing the PH assumption for a given predictor of interest. rho is a relation between time and residuals. The test of correlation (rho) is insignificant that indicates proportional hazards assumption is fulfilled. The P-values given in the Table 4.5 provide goodness-of-fit tests for each variable in the fitted model adjusted for the other variables in the model are quite high for variables Age, Smoking status, region, family history, HIV status, abortion history and age at first birth, suggesting that all the listed variables satisfy the PH assumption. But variable stage not satisfies the PH assumption. Moreover it is also

possible to see its global test and if it is greater than 0.05 the assumption have satisfied by the covariates in the model. In this study the global test is less than 0.05 the assumptions do not satisfied by the covariate in the model.

Table 4.5: Test of proportional hazards assumption

Covariates	Rho	Chisq	DF	p
Age	-0.094	2.96	1	0.0851
Smoking Status	0.014	0.07	1	0.7871
Region	-0.025	0.26	1	0.6107
Stage	-0.295	29.41	1	0.014
Family History	0.066	1.60	1	0.2063
Abortion	0.066	0.69	1	0.4063
HIV status	-0.07	1.84	1	0.1754
age at first birth	-0.087	2.61	1	0.1065
GLOBAL	-	37.21	1	0.0018

Chisq= chi-squared, DF=degree of freedom

The scatter plots of Scaled Schoenfeld residuals in appendix also used to check PH assumption. If the PH assumption is met, Schoenfeld residuals should look horizontal since the scaled Schoenfeld residuals would be independent of survival time. The plot of stage against survival time was also slightly downward (not horizontal). These also revealed that there is a violation of the proportional hazard assumption for the covariate stage. Thus, we doubt the accuracy of the PH assumption and consider the AFT model for this data set.

4.4. Accelerated Failure Time Model

When PH assumptions were not satisfied, the parametric AFT model should be used instead of the Cox model. Since the p value of the goodness of covariates were significant for the variable stage. In this case AFT model is appropriate for the data as discussed in section 3.5.2.

4.4.1. Univariate AFT Analysis

This study used univariate analysis in order to see the effect of each covariate on survival time before proceeding to the multivariable analysis. The univariate analysis was fitted for every covariate by AFT models using different baseline distributions i.e. exponential, weibull, log-logistic and lognormal. In all univariate analysis of AFT models age, smoking

status, region, stage, family history, abortion history, HIV status and age at first birth of women were significantly associated with survival time of patients at 5% level of significance. The summary of univariate analysis is given in table 4.6 in the appendix. Hence, based on the univariate analysis, all explanatory variables filtrated in Cox PH are candidate predictors for further analysis in AFT models.

4.4.2. **Multivariate AFT Analysis**

For survival time of CC patients data, multivariable AFT models of exponential, weibull, log-logistic and log-normal distribution were fitted by including all the covariates those are significant in the univariate analysis at 5% level of significance. To compare the efficiency of different models, the AIC was used. It is the most common applicable criterion to select model. Based on AIC, a model having the minimum AIC value was preferred.

Accordingly, Log-normal AFT model (AIC=1185.05) found to be the best for the survival time of CC patients data set from the given alternatives when include all the covariate those are significant in the univariate analysis. All covariates significant in the univariate become significant multivariate analysis model. Finally, the effect of interactions terms were also tested and found to be statistically insignificant in multivariable log-normal AFT model at 5% level of significance. The final model kept the main effect of the covariate models age, smoking status, region, stage, family history, abortion history, HIV status and age at first birth of women. All AFT models and the corresponding AIC values are displayed in Table 4.7.

Table 4.7: Comparison of AFT models using AIC criteria for of cervical cancer patients' data

Baseline Distribution	**AIC**
Exponential	1410.30
Weibull	1211.67
Log- logistic	1188.95
Log- normal	1185.05

AIC=Akaike's information criteria

4.4.3. Interpretation and presentation of the final AFT model

Model diagnostics are presented in section 4.5 suggested that the AFT model is in good fit the CC dataset. Thus the acceleration factors in the lognormal model were interpreted as follows. The output of the final log-normal AFT model is presented in Table 4.8. This output showed patients with age groups 51-60, 61-70 &>70, women who smoking cigarettes, patients with stage III &IV, women who had family history of CC, patients with previous abortion history and living with HIV AIDS were statistically significantly shorten survival time of women with cervical cancer while patients not giving birth up to the study ends prolong the survival time of cervical cancer patients in TASH.

Table 4.8: Summary result the final Log-Normal AFT model

Covariates	Category	$\hat{\beta}$	SE	Sig.	Φ	95% CI for ϕ
	20-30	Ref				
Age	31-40	0.20	0.19	2.72e-01	1.23	[0.853,1.759]
	41-50	0.12	0.18	5.08e-01	1.12	[0.794, 1.592]
	51-60	-0.41	0.17	1.78e-02*	0.67	[0.475, 0.932]
	61-70	-0.54	0.18	2.28e-03*	0.58	[0.411, 0.824]
	>70	-0.63	0.23	1.50e-02*	0.53	[1.476, 1.753]
Smoking	nonsmoker	Ref				
Status	Smoker	-0.39	0.08	6.63e-07*	0.68	[0.584, 0.792]
Region	Addis Ababa	Ref				
	Oromiya	0.28	0.07	6.25e-05*	1.33	[1.154, 1.523]
	Amhara	0.22	0.80	7.04e-03*	1.24	[1.06, 1.451]
	Tigray	0.29	0.17	8.73e-02	1.34	[0.957, 1.88]
	SNNPR	0.11	0.13	3.88e-01	1.11	[0.873, 1.416]
	Others	0.05	0.16	7.49e-01	1.05	[0.758, 1.466]
	I	Ref				
	II	-0.08	0.14	6.16e-01	0.94	[0.718, 1.219]
Stage	III	-0.59	0.13	3.73e-06*	0.55	[0.429, 0.711]
	IV	-0.97	0.15	3.93e-11*	0.38	[0.283, 0.507]
Family	No	Ref				
History	Yes	-0.30	0.08	6.94e-05*	0.74	[0.638, 0.857]
Abortion	No	Ref				
	Yes	-0.30	-0.30	7.77e-06*	0.74	[0.644, 0.885]
HIV status	No	Ref				
	Yes	-0.32	-0.32	7.41e-07*	0.73	[0.645, 0.826]
	<=15	Ref				
	16-20	0.14	0.06	2.12e-02*	1.15	[1.020, 1.303]
age at first	21-25	0.12	0.10	2.49e-01	1.12	[0.922,1.361]
birth	>25	0.31	0.242	2.00e-01	1.36	[0.455,1.176]
	Not give birth	0.96	0.231	9.73e-03*	2.609	[1.359,3.359]

ϕ *Indicates Acceleration factor; * significant at 5% level; 95%CI for* ϕ*: 95% confidence interval for acceleration factor; SE: standard error for* estimates*; Ref. Reference*

Under the log-normal AFT model, when the effect of other factor keep fixed, the estimated acceleration factor for patients with age group 51-60 is estimated to be 0.67with [95% CI: 0.475, 0.932].The confidence interval for the acceleration factor did not include one and P-value is small (p=1.78e-02*). This indicates patients with age group51-60 have less survival time than the age group of 20-30 patients. Similarly acceleration factor for patients with age group 61-70 are estimated to be 0.58 with [95% CI: 0.411, 0.824] the ϕ CI did not include one and P-value is small (p=2.28e-03*). This implied that women with age group 20-30have longer survival time than women with age group 61-70.Likewise acceleration factor for patients with age group >70 is estimated to be 0.53 with [95% CI: 0.476, 0.953] the ϕ CI did not include one and P-value is small (p=1.50e-02*). This implied that women with age group 20-30 have longer survival time than women with age group >70. But the age groups 31-40 & 41-50 patients were not significantly different from the age group 20-30 at 5% level of significance.

As shown in Table 4.8 the estimated acceleration factor for patients from Oromia and Amhara were estimated to be 1.33 & 1.24 with [95% CI 1.154, 1.523; 1.06, 1.451] the confidence interval for the acceleration factor did not include one and P-value is very small (p=6.25e-05*, 7.04e-03*).This indicates patients from Oromia and Amhara have prolonged survival time than Addis Ababa patients. But patients from Tigray, SNNPR, Others region were not significantly different from the Addis Ababa at 5% level of significance. The acceleration factors for patients who smoke cigarette were estimated to be 0.68 with [95% CI: 0.584, 0.792]. This implied that non-smoker had longer survival time than smoker or in the other way Patients not smokes cigarettes survived 32% longer than patients who smokes cigarettes. According to Table 4.8 the acceleration factors of patients with stage III and IV were estimated to be 0.55 and 0.38 with [95% CI: 0.429, 0.711 & 0.283, 0.507] respectively. This oblique that the stage I patients have longer survival time than stage III and IV, however the difference is not significant for women with Stage II and stage I (P=6.16e-01).The acceleration factor for women that had family history of cervical cancer was 0.74 with [95% CI: 0.638, 0.857 and P=6.94e-05*] which indicates that the survival of time of patients that had family history of cervical cancer shorten than patients did not have family history.

Similarly the estimated acceleration factor for patients who had history of abortion is estimated to be 0.74 with [95% CI: 0.644, 0.885]. The confidence interval for the acceleration factor did not include one and P-value is very small (p=7.77e-06*) which

indicate patients who had no history of abortion prolong the survival time than patients who had history of abortion. In the same way patients that were no living with HIVAIDS prolong the survival time than patients living with HIV AIDS with acceleration factor 0.73 and [95% CI: 0.645, 0.826]. Finally the estimated acceleration factor for women give birth at age of 16-20 is estimated to be 1.15 with [95% CI: 1.020, 1.303] the ϕ CI did not include one and P-value is small (P=2.12e-02*). This indicates women who got first baby at age of 16-20 have prolonged survival time than women that got the first baby at age of ≤15 years. The estimated acceleration factor for women did not giving birth up to the study time is estimated to be 2.61 with [95% CI: 1.359, 3.359] the confidence interval for the acceleration factor did not include one and P-value is small (P=9.73e-03*). This indicates women did not give birth up to the end of study time have prolonged survival time than women gave birth at age of ≤15 years or patients who did not give birth at the study time survived 161% longer than patients who gave birth at age of ≤15 years. But those women gave birth at age of 21-25 and >25 were not significantly different from the reference age group at 5% level of significance.

4.5. Model Diagnostics

After the model has been fitted, it is desirable to determine whether a fitted parametric model adequately describes the data or not.

4.5.1. Checking Adequacy of Parametric Baselines using Graphical Methods

To check the adequacy of our baseline hazard the exponential is plotted by the -log(S(t) with the time of the study; the Weibull is plotted by log (-log(S(t)) with the logarithm of time of the study; the log-logistic is plotted by $\log(\frac{\hat{S}(t)}{1-\hat{S}(t)})$ with the logarithm of time of the study and the log-normal baseline by $\phi^{-1}[1-S(t)]$ against log (t) of time of the study. If the plot is linear, the given baseline distribution is appropriate for the given dataset. Accordingly, their respective plots are given in figure 4.13 below and the plot for the Log-normal baseline distribution make straight line better than exponential, Weibull and Log-logistic baseline distribution. This evidence also strengthens the decision made by AIC value that log-normal baseline distribution is appropriate for the given dataset.

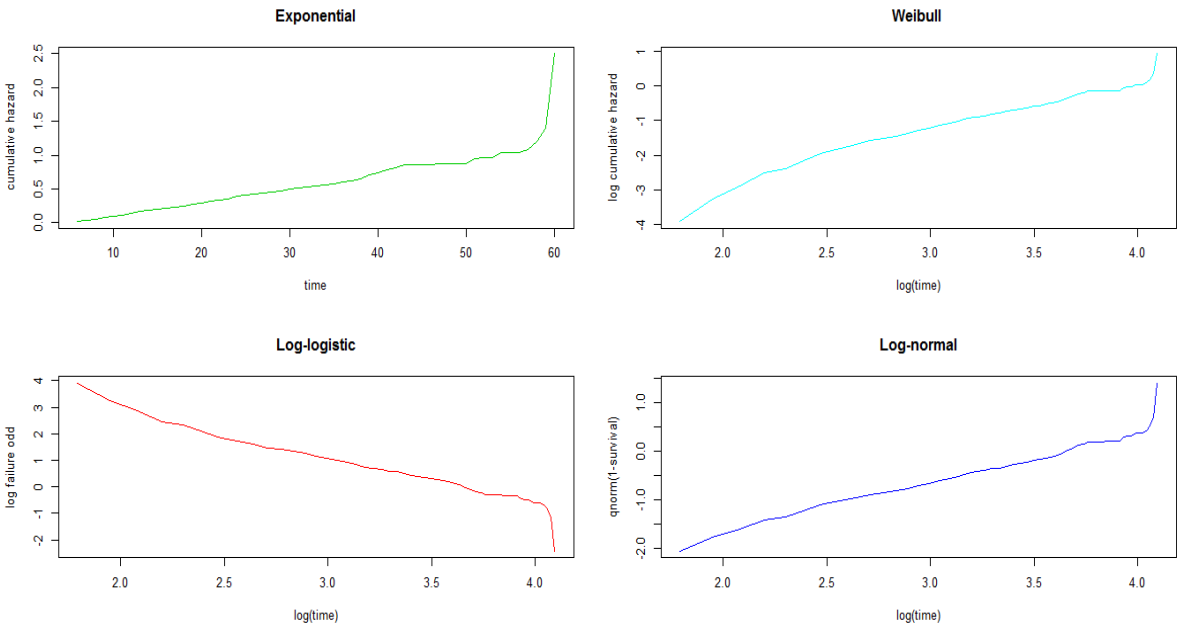

Figure 4.13: Graphs of Exponential, Weibull, Log-logistic, and Log-normal baseline distributions for survivaltime of cervical cancer patients' data set.

Table 4.9: the likelihood ratio and significance of the Lognormal AFT

Loglik(intercept only)	Loglik(model)	Chisq	DF	Sig.
-1724.6	-1560.9	327.39	21	0.00

The likelihood ratio test in Table 4.9 shows that the model is significant and the log likelihood values of the null model and the full model indicates that the model has a significant improvement after the covariates were added in the model.

4.5.2. Cox- Snell residuals plots

The Cox-Snell residuals are one way to investigate how well the model fits the data. The plot for fitted model of residuals for log-normal to our data via maximum likelihood estimation with cumulative hazard functions is given in figure 4.14. If the model fits the data, the plot of cumulative hazard function of residuals against Cox-Snell residuals should be approximately a straight line with slope 1. The plot makes straight lines through the origin for log-normal baseline distribution suggesting that it is appropriate for survival time of cervical cancer patients' data set.

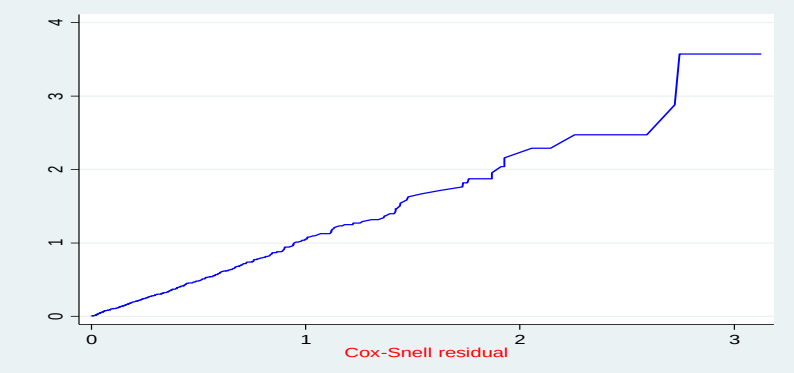

Figure 4.14: Cox- Snell residuals plots of log-normal baseline distribution for survival time of cervical cancer patients' data.

4.5.3. Quantile-Quantile Plot

A quantile-quantile plot is made to check if the AFT provided an adequate fit to the data by using two different groups of population. We shall graphically check the adequacy of the model by comparing the significantly different groups of patients by family history, smoking cigarettes, HIV status and abortion history. The figures appear to be approximately linear for all covariates family history, smoking status, HIV status and abortion history of patients as shown in figure 4.15. Therefore the accelerated failure time appears to be the best to describe survival time of cervical cancer data set.

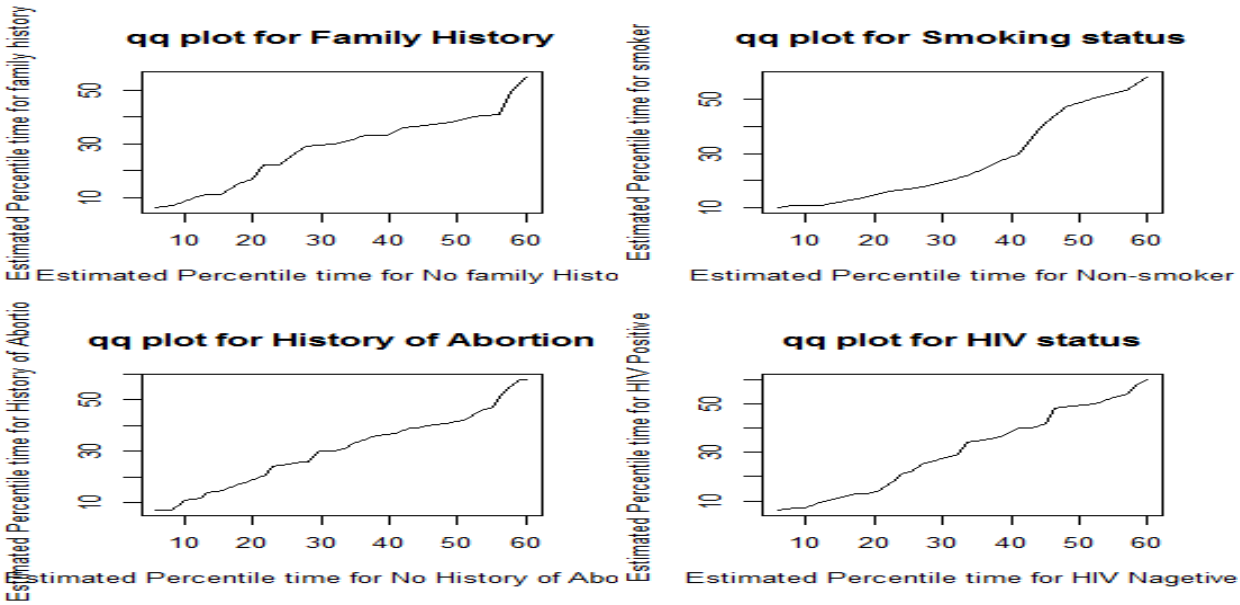

Figure 4.15: Quantile- Quantile plot to check the adequacy of the AFT model

4.6. Discussion

The main purpose this study was modeling the determinant of survival time of cervical cancer dataset which was obtained from TASH. Cox PH was applied for this data. But since PH assumption was violated AFT models with baseline distributions: Exponential, Weibull, Log-logistic and Log-normal distributions were applied to the same data. Covariates which were included in the study were age, smoking status, region, recurrence, stage of the disease, treatment taken, cycles of chemotrapy, sexual partner, number of children, aim of radiography, history of abortion, HIV status, family history, age at marriage and age at first birth and the outcome variable of interest was the survival time of cervical cancer patients measured in months.

The Cox's proportional hazard model fitted using single and multi-case analysis. Multi case analysis found eight variables that jointly serve as predictive factors on the survival of cervical cancer patients. These variables are age, smoking status, region, stage of the disease, family history, history of abortion, HIV status and age at first birth. Out of this the stage of the disease of cervical cancer patients didn't satisfy the proportionality of Cox regression model and then AFT model was fitted. The current study is consistent with other findings Muluye.G (2011), Orbe.J et al (2010), Frankel P; Longmate J. (2010) and Wang SJ *et al.* (2010) with regards the Cox PH and AFT models.

The univariate AFT analysis is revealed that age, smoking status, region, stage, family history, history of abortion, HIV status and age at first birth were significantly associated with time to death of cervical cancer patients' at 5% level of significance. All significant covariates in univariate analysis were included in multivariable analysis of AFT model and comparison was done within the models using AIC criteria where the model having minimum AIC value is selected to be the best Munda 2012), Chapman *et al.* (1992), Komarek (2004). AFT model fitted using eight covariates that jointly serve as predictive factors on the survival of cervical cancer patients.

Log-normal AFT model was found to be the best over Exponential, Weibull and Log-logistic AFT models based on AIC value and graphical evidence .Age, smoking status, region, stage of the disease, family history, history of abortion, HIV status and age at first birth were significantly associated with survival time of cervical cancer patients. This finding is in accordance with the studies Ayele.G (2015), Elisa T. Lee (2007), Mohammad

Amin et al (2008), Chapman *et al.* (1992), Royston (2001) with regards the Lognormal AFT model.

The findings of this study revealed that detected at early age and early stage of the disease and not giving birth until the end of study time prolong the timing death of cervical cancer patients while detected at old age and advanced stage of the disease, smoking cigarettes, family history of cervical cancer, history of abortion, living with HIV AIDS, giving birth at early age shorten timing death of cervical cancer patients. Kumari *et al*. and Bingham et al. (2013), Kim, et al. (2009) and World health organization (2015) support this finding. The overall estimated mean survival time was 26 months. The result of this finding is almost similar with the Zalimr. *et al.* (2011) reported that 27 months.

The numbers of cervical cancer cases were not related to population size of the regions in Ethiopia at the study period. The number of cases in other region of the country depends on their distance from Addis Ababa; the farther the region, the fewer number of cases. Because patients may not got the opportunity to go and be diagnosed. Those patients who are relatively near to the hospital such as patients in Addis Ababa, Oromiya and Amhara are able to go the hospital and diagnosed. But, those patients who are far from the hospital are not able to go the hospital easily due to many reasons like transportation cost, accommodation, job and family related cases. Similar finding was obtained from Abate.S (2015), Sule ST, Shehu MS (2007).

Patients treated at old age were significantly associated to shorten to time death of cervical cancer. This implied survival time for cervical cancer patients was highest in the youngest women and decreases with increasing age. The current study is consistent with other findings of Kantelhardt *et al*. (2014),Yancik. R **(**2013), Wright JD. (2014), Castillo et al., (2009)&Kumari *et al.*, (2013) reported that increases in age have been related to poorer survival time of cervical cancer.

The results of this study suggested that smoking was significant predictive factor for survival time of the patients. Non-smokers had longer survival time than smokers. The current study is consistent with other findings of Tigist Y. (2014), S. E. Waggoner et al. (2010), Kumari *et al*. (2013), Garland et al., Yetimalar H, Kasap B andCukurova. K *et al.* (2011).

Accordingly the stage of the disease also had significant association to survival timing of patients. Stage I and Stage II patients had long survival time than stage III and IV patients. This is due to early to an early stage cervical cancer can simply be cured by available cancer treatments. This result is consistent with Tigist Y. (2014), Kantelhardt *et al.* (2014) and Kumari *et al*. (2013).

Family history of cervical cancer is also one factor that significantly predicts the survival of the patients. In this study, family history is significantly associated with survival time cervical cancer patients. The current study is consistent with other findings Kantelhardt *et al.* (2014), IARC (2014) and Bellinger.JD (2013reported that women who have a mother or sister diagnosed with cervical cancer have a greater risk of developing this cancer than women without any cervical cancer family history

The history of abortion also had significant association to survival time of cervical cancer patients'. Those women who had no history of abortion had long survival time than those who had abortion history. This result is consistent with Ahmed Ib. *et al*. (2011) reported that previous abortion plays a role in cervical cancer. And also a study by R. Somers *(2012)* in Danish found a significantly positive trend between those women had abortion and the death of cervical cancer.

The results of this study indicated that status of HIV AIDS was significant predictive factor for survival time of cervical cancer patients in TASH. Women who lived with HIV AIDS had smaller survival time than women who are free from HIV. This happens since cervical cancer makes CD4 counts lower and they are more likely to die. Abel G. *et al* (2013), Massad LS (2008) and Abel G (2013) found a significantly positive trend between those women having had HIV and the death of cervical cancer.

The results of the study revealed that having first baby at early age ($\leq$15) years has a significant association with survival time of cervical cancer patients. Women who has had first child at early age had smaller survival time. Similarly Similar finding was obtained from Ahmed Ib. et al (2011) reported having your first baby before the age of 16 gives a higher risk, compared to women who had their first baby after the age of 25.This result is also in accordance with the studies Brinton. L.,*et al* (2002) and Chelimo C, *et al* (2013).

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study used survival time of cervical cancer patients' dataset of those patients who started their cancer treatment from 2003-2007 years with the aim of modeling the determinant of time-to-death of women with cervical cancer in TASH. Out of the total 907 women who started cancer medicine (treatments), about 38.48% died at the end of the study. The estimated mean survival time of women was 26 months.

To model the determinants of survival time of cervical cancer patients, Cox Ph model was used. Then AFT model was fitted because the assumption of Cox proportional model was violated. Different AFT models by using different baseline distributions were applied. Among this using AIC, Log-normal model is better fitted survival time of cervical cancer patients' dataset than other AFT base line distributions.

The result of Log-normal AFT model showed that age, smoking status, stage, region, family history, abortion history, HIV status and age at first birth were found significant predictors for survival time of patients in TASH. Of which not giving birth up to study the end of time prolong the timing death of cervical cancer patients. Similarly age classes51-60, 61-70 &>70, smoking cigarettes, stage III & IV, family history of cervical cancer, abortion history, HIV status and having first baby at early age ($\leq$15) were statistically significantly shorten timing of death of women with cervical cancer.

Goodness of the fit of baseline distribution by means of graphical method in figure 4.13 and Cox-Snell residuals plots in figure 4.14 revealed that Log-normal distribution is better when compared to Exponential, Weibull and Log-logistic baseline distributions to explain survival time of women with cervical cancer dataset in Tikur Anbessa Specialize Hospital.

5.2 Recommendations

1) Based on the findings of the study the following recommendations are made for ministry of health, policy makers, the community at large, Tikur Anbessa Specialized Hospital and researcher.

2) The ministry of health and policy makers should work on awareness by letting to know the risk factors for the cervical cancer and to complete the prescribed treatment without considering cervical cancer as incurable disease and to follow up their cancer status to minimize the risk of death and recognizes cervical cancer as an important health problem and establishing screening test and early detection policies for most risky groups (for old age and married women).
3) In addition it will be important to open cancer diagnosing and treatment center in each region of the country.
4) Awareness has to be given for the society on causes of cervical cancer. The mass media can play an effective role in this regard and special attention should be given to old age women, because they are the most risky groups for cervical cancer.
5) Tikur Anbessa Specialize Hospital need to improve public and professional awareness, early detection and prompt treatment using feasible, effective regimens and include detailed patient characteristics in the cancer registry data. This hospital based cancer registry is older which need integration with WHO international coding system.
6) Further studies should be conducted in each regional states of Ethiopia and identify other factors that are not identified in this study. Based on that study, regional governments should take actions of screening to reduce death of cervical cancer.

5.3 Limitation of the Study

As the data is gathered from the card of patients in the study there were a lot of patients with insufficient information; lack of published literature on the country related to the survival time of cervical cancer. Even if the number of patients is high in number, on follow-up it is often a physical and financial burden for them to return hospital for follow-up in the day of appointment; most patients didn't come again to the Hospital to be treated.

REFERENCES

Aalen O. (2008).Survival and Event History Analysis. Springer-Verlag, New York.

Abate. S (2015).Trends of Cervical Cancer in Ethiopia, Oncology Pharmacy Case Team, TikurAnbessa Specialized Hospital, College of Health Sciences, Addis Ababa University, Ethiopia

Abel. G, Ayalew. A, Gizachew.A (2013). The Prevalence of Cervical Cancer Lesion among HIV-Infected Women inSouthern Ethiopia: A Cross-Sectional Study

AbrehamBerhane (2011). Statistical Modeling for the Recurrence of Cervical Cancer: A Case Study at TikurAnbessa Referral Hospital, Addis Ababa, Ethiopia. Master Thesis Hawassa University, Department of Statistics

Agency for Health Care Policy and Research (2009).Evidence report/technology assessment no. 11, evaluation of cervical cytology. AHCRP Publication No. 109-E010. Rockville, MD: US Department of Health and Human Services. 16.

Ahmed Ib. Mohhamde Y.(2011).Predictors of cervical cancer being at an advanced stage at diagnosis in Sudan.Int J women health 3:385-389

Akaike's H. (1974). A new look at the statistical model identification. IEEE Trans Automatic Control.

American Cancer Society (2012).Guidelines for the early detection of cancer.American Cancer.Society.2012aRetrievedfromhttp://www.cancer.org/Healthy/FindCancerEarly/CancerScreeningGuidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer.

American Cancer Society (2015), "Prostate Cancer: Detailed Guide 4th edition,"

American Congress of Obstetricians and Gynecologists(2013).Cervical cancer screening.Retrievedfromhttps://www.acog.org/~/media/For%20Patients/faq085.pdfdm

Andersen, P. K., Borgan, O., Gaill, R. D., and Keiding, N(1993). Statistical Methods Based on Counting Processes.Springer, New York.

Ayele.G(2015). Survival Analysis of Time-to-First Birth after Marriage among Women in Ethiopia: Application of Parametric Shared Frailty Model.

Ball C, Madden JE (2009). Update on cervical cancer screening. Current diagnostic and evidence-based management protocols. Pos grad Med, 113, 59-64

BellingerJD (2013). The role of family history of cancer on cervical cancer screening behavior in a population-based survey of women in the Southeastern United States

Bennett S. (1983). Analysis of survival data by the proportional odds model.Statistics in Medicine, 2, 273.

BerksonJ., and Gage R. P. (1950) .Calculation of survival rates for cancer. Proceedings of the staff meetings of the Mayo clinic 270.286

Bingham A, Bishop A, Coffe P, Winkler J, Bradley J, Dzuba I, et al(2013). Factors affecting utilization of cervical cancer prevention services in low-resource settings. SaludPublica Mex. 45 Suppl 3:S408-16.

BogalechFufa(2015). Assessment of Knowledge, Attitudes and Practices on Cervical Cancer Screening among Health Service Providers at Marie Stopes International Ethiopia Centers

Bosch FX, de Sanjose S, Muñoz N, Herrero R, Castellsagué (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Eng J Med 348:518–527.

Brinton LA, Hamman RF, Huggins GR, et al(2002). Sexual and reproductive risk factors for invasive squamous cell cervical cancer. J NATL CANCER INST. 79(1):23–30.

Central Statistical Agency (2011).Ethiopian Demographic and Health Survey. Addis Ababa, Ethiopia

Castillo, O.O., R. Fernandez, R. Arauz, E. Zarak, J. Diaz, C. Alcedo, R. I. Lopez and F. Cebamanos(2009).Prognostic factors for death in elderly Women with cervical carcinoma. Journal of Clinical Oncology, 26: 16564.

Chapman, J. W., Trudeau, M. E., Pritchard, K. I., Sawka, C. A., Mobbs, B. G., Hanna, W.M., Kahn, H., McCready, D. R., Lickley, L. A (1992). "A comparison of all-subset Cox and accelerated failure time models with Cox step-wise regression for node-positive breast cancer", Breast Cancer Research and Treatment, 22(3): 263 – 272

Chelimo C, Wouldes TA, Cameron LD, Elwood JM(2013). Risk factors for and prevention of uman papillomaviruses (HPV), genital warts and cervical cancer. J INFECT;66(3):207–217.

Chidiebere.M.I (2009). Factors influencing cervical cancer survival uptake among women attending Mahalapye District Hospital in Botswana

Collett, D. (2003). Modeling Survival Data in Medical Research Chapman & Hall, London

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

Cox D. R. and Oakes D. (1984). Analysis of Survival Data Chapman and Hall, London.

Datwyler, C., and Stucki,, T., (2011). Parametric Survival Model, Handout.

David G. Kleinbaum Mitchel Klein (2005). Survival Analysis, A Self-Learning Text, Second Edition.

Dress, K.A. (1986). The Effect of Gender Identify on Conversation. Social Psychology Quarterly, Vol 49: P 294-301

Elisa T. Lee (2007). Statistical methods for Survival Data Analysis, 3rd Edition

mortality and prevalence worldwide," *Cancer Journal for Clinicians,* vol. 55, no.5, pp.74-108

Eralp, Y., P. Saip, B. Sakar, S. Kucucuk, A. Aydiner, M. Dincer, I. Aslay and E Topuz (2006). Prognostic factors and survival in patients with metastatic or recurrent carcinoma of the uterine cervix. *International Journal of Gynecological Cancer* 13 (4):497-504. 33. Franco, E. D

Ethiopian Cancer Association (2010).Assessment of Major Risk factor of cancer in Ethiopia

Frankel P, Longmate J. (2002). Parametric models for accelerated and long-term survival: A comment on proportional hazards.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, ParkinDM(2009). GLOBOCAN, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 10.

FIGO Global Guidance (2009).Cervical Cancer Prevention and Control.

Greenwood, M (1926). The natural duration of cancer. Reports on Public Health and Medical Subjects 33, 1-26.

Gupta, R. and Kundu, D. (1999).Generalized exponential and log-logistic distributions.

Hosmer .D and Lemeshow S. (1998).Applied Survival Analysis Regression Modeling of Time to Event Data. John Wiley and Sons, New York

Hosmer DW, Lemeshow(1999). Regression modeling of time-to-event data. Wiley, NewYork.

Imam SZ, Rehman F, Zeeshan MM, et al (2009). Perceptions and practices of a Pakistani Population regarding cervical cancer screening.Asian Pacific J Cancer Prev, 9, 42-4.

International Agency for Research on Cancer; World Health Organization (2009). World Cancer Report, Lyon: IARC

International Agency for Research on Cancer (2014). List of Classifications by cancer sites with sufficient or limited evidence in humans(link is external). Volumes 1 to 105

Kalbfleisch, J. D., and Prentice, R. L (2002). The statistical Analysis of Failure Time Data, 2nd ed.Wiley, New York.

Kantelhardt ,Moelle. U,Begoihn .M, Addissie.A, Yonas.B,Hezkiel.P,Stagn.A,Thomssen.C, Vordermark,Gemechu.T (2015),Cervical cancer in Ethiopia: survival of 1,059 patients who received oncologic therapy

Kaplan EL and Meier P. (1958).Nonparametric estimation from incomplete observations.

Journals of American Statistical Association 53:457-81.

Karadag G, Gungormus Z, SurucuR (2014). Awareness and practices regarding breast AndSexual morbidity in very long term survivors of vaginal and cervical cancer

Kim, et al. (2009). Image-guided stereotactic body radiation therapy in patients with Isolated Para-aortic lymph node metastases from uterine cervical and corpus cancer. INT. J. RADIAT. ONCOL. BIOL. PHYS. 74 147–153

Klein, J. P., and Moeschberger, M. L (1997). Survival Analysis: Techniques for Cen-sored andTruncated Data. Springer, New York.

Klein, J. P., &Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

Klembaum D. G (1996). Survival Analysis: A Self learning text. Springer, New York.

Komarek, A., Lesaffre, E., and Hilton, J.F (2004). Accelerated failure time model for arbitrarily censored data with smoothed error distribution, Technical Report 0451, IAP statistics network.

Kruiroongroj S, Chaikledkaew U, Thavorncharoensap M (2014).Knowledge, acceptance, and willingness to pay for Human Papilloma Virus (HPV) vaccination among female parents in Thailand. Asian Pac J Cancer Prev, 15, 5469-74.

Kumari K, Gupta S, Parida SK, Chattopadhyay D, Prasad M (2013). Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [Setariaitalica (L.)]. DNA Res 20: 207

Louie K, Sanjose S &Mayaud P (2009). Epidemiology and prevention of human papilloma virus and cervical cancer in sub- Saharan Africa: a comprehensive review. Tropical Medicine and International Health Vol. 14 NO. 10, pp 1287 1302

Manjunath G, Tighiouart H, Ibrahim H, MacLeod B, Salem DN, Griffith JL, et al(2010). Level of kidney function as a risk factor for atherosclerotic cardiovascular outcomes in the community. J Am CollCardiol, 41(1):47-55.

Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22, 719-748.

MassadLS(2008. Squamous cervical lesions in women with human immunodeficiency virus: Long-term follow-up. *Obstetrics and Gynecology* 111(6):1388–1393.

Mohamad Amin et al (2008) Comparing Cox Regression and Parametric Models for

Survival of Patients with Gastric Carcinoma

MonsonegoJ, Autillo-Touati A, Bergeron C (2004). Liquid-based cytology for primary Cervical cancer screening: a multi-centre study. Cancer 84:360–6.

Moosa NY, Khattak N, Alam MI, et al (2014). Comparison of cervical cell morphology usingtwo different cytology techniques for early detection of pre-cancerous lesions. AsianPac J Cancer Prev, 15, 975-81.

Muluye.G (2011) Modeling Time-to-Menarche in Ethiopia: Cox Proportional Hazard versus Accelerated Failure Time Models

Munda M., Rotolo F. and Legrand C. (2012). Parametric Frailty Models in R. Journal of American Statistical Association: 55:1-21

Nagpal J, Osmond C &Sachdev HP (2012). Effect of administration of intestinal anthelmintic drugs on haemoglobin: systematic review of randomised controlled trials. BMJ **334**, 1095.

National Cancer Institute (2014). Defining Cancer in standard way, Estimates of

Healthcare professionalNational Health Service (2015). National Department of Health, NHS Performance Assessment Framework, http://www.doh gov.uk/indicat/indicat.htm.

NC IPress Office (2012). NCI Issues Clinical announcement on Cervical Cancer: Chemotherapy Plus radiation Improves.

Orbe J., Ferreira E., Núñez-Antón V. (2010). Comparing proportional hazards and AFT models for survival analysis.*Stat in Med,* 21(22),3493- 3510

Parkin DM, Whelan SL, Ferlay J, TeppoL(2014).Cancer incidence in five continents 5th edition Lyon: IARC Press; 2014.

Parkin DM (2000).Global cancer statistics The Lancet Oncology ,2 (9):533-543.

Pourhoseingholi MA, Pourhoseingholi A, Vahedi M (2011). Alternative for the Cox regression model: using parametric models to analyze the Survival of Cancer Patients.Iran J Cancer Prev; 4:1- 9.

Program for Appropriate Technology in Health(2000). Planning appropriate cervical cancer prevention programs. 2nd Edition.http://www.path.org/publications/files/cxca-planning-appr o-prog-guide.pdf

Rebecca S. and Jemal.A (2015). Global Cancer Facts & figures 2nd Edition.

Royston, P (2001). The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors, StatisticaNeerlandica, 55:89-104

R. Somers (2001). Risk of Admission to Psychiatric Institutions among Danish Women who Experienced Induced Abortion: An Analysis on National Record Linkage," Dissertation Abstracts International, Public Health 2621-B, Order No. 7926066

Schoenfeld (1982) .Schoenfeld D. Residuals for the proportional hazards regresssion model. Biometrika, 69(1):239-241.

Sogukpinar N, KaracaSaydam B, Ozturk Can H, et al (2013).Assessment of cervical cancer risk in women between 15 and 49 years of age: case of Izmir. Asian Pac J Cancer Prev, 14, 2119-25

S. E. Waggoner et al (2010). Smoking behavior in women with locally advanced cervical Cancer: a gynecologic oncology group study, American Journal of Obstetrics and Gynecology, vol. 202, no. 3, pp. 283.e1–283.e7

Sule ST,ShehuMS(2007). Cervical cancer management in Zaria, Nigeria. African Journal of Health Sciences 14:149-153

T. Ibrahim (2009). Research Methodology and Dissertation Writing forHealth and Allied Health Professionals, Cress Global Link, Abuja, Nigeria

TC Saglık, Bakanligi (2014). Knowledge, Attitudes and Behaviors of Women Over 20 Years Old on Cervix Cancer in Istanbul, Turkey

Tigist Y.(2014). Survival and associated factors among cervical cancer patients in Black LionHospital,

Wang SJ, Kalpathy J., Kim JS, Fuller D. and Thomas CR. (2010). Parametric Survival Models for Predicting the Benefit of Adjuvant Chemo radiotherapy in Gallbladder Cancer.AMIAAnnuSymp Proc. 847–51

WHO /ICO (2009).Human papiloma virus and related cancers in Ethiopia. In Summary report.

Wittet .S, Tsu.V (2009). Cervical cancer prevention and the Millennium Development Goals. Bull World Health Organ 86: 488-490.

World Bank (2014). Causes of and population dynamics for Ethiopia including life expectancy by age and history, age adjusted rates and rankings for the top 50 causes and total cause by death.

World Health Organization (2015). Comprehensive cervical cancer control: A guide to essential Practice

Wright, JD. (2014). *Cervical intraepithelial neoplasia: Terminology, Incidence, pathogenesis, and prevention*. Accessed Jul.3,2014, from http://www.uptodate.com/contents/cervical-intraepithelial-neoplasia-terminology-incidence-pathogenesis-and-prevention

Yancik (2012). Associations between Gene Expression Variations and Cervical Cancer Risk Alleles Identified from Genome Wide Association Studies.

Yetimalar H, Kasap B, CukurovaK (2011). Cofactors in human papillomavirus infection and cervical carcinogenesis.*ArchGynocolObstet*, 285**,** 805-10

Zali MR. et al (2011) alternative for cox regression model using AFT model to analysis survival of cervical cancer in kenya

ZurHausen H (2004). Papillomaviruses in human cancers.MolCarcinog, 1, 147-50

APPENDIXIES

Table 4.1 Descriptive Summary of the Data at TASH (2003-2007)

		Status of Patient		
Variable	**Category**	**Number of Event (%)**	**Number of Censored (%)**	**Total (%)**
Age	20-30	9(2.58)	20(3.58)	29(3.20)
	31-40	30(8.60)	123(22.04)	153(16.87)
	41-50	53(15.19)	204(36.56)	257(28.34)
	51-60	155(44.41)	135(24.19)	290(31.97)
	61-70	90(25.79)	59(10.59)	149(16.43)
	>70	12(3.44)	17(3.44)	29(3.20)
Smoking habit	non-smoker	268(76.79)	516(92.47)	784(86.44)
	Smoker	81(23.21)	42(7.53)	123(13.56)
Region	Addis Ababa	146(41.83)	155(27.78)	301(33.19)
	Oromiya	97(27.79)	200(35.84)	297(32.75)
	Amhara	65(18.62)	128(22.94)	193(21.28)
	Tigray	9(2.58)	22(3.94)	31(3.42)
	SNNPR	20(5.73)	39(6.99)	59(6.5)
	Others	12(2.87)	14(2.51)	26(2.87)
Recurrence	No	280(80.23)	497(89.07)	777(85.67)
	Yes	69(19.77)	61(19.77)	130(14.33)
Stage	I	19(5.44)	28(5.02)	47(5.18)
	II	90(25.79)	187(33.51)	277(30.54)
	III	189(54.15)	293(52.51)	482(53.14)
	IV	51(14.61)	50(14.61)	101(11.14)
Sexual partner	None	3(0.86)	10(1.79)	13(1.43)
	One	91(26.07)	91(26.07)	390(43.00)
	Few	81(23.21)	81(23.21)	217(23.93)
	Multiple	174(49.86)	113(20.25)	287(31.64)
Aim of radiography	no RT	4(0.36)	2(0.36)	6(0.66)
	Palliative	240(68.77)	354(63.44)	594(65.49)
	Radical	105(30.09)	202(36.20)	307(33.85)
Family History	No	275(78.80)	499(89.43)	774(85.34)
	Yes	74(21.20)	59(10.57)	133(14.66)

	Surgery	4(1.15)	1(0.18)	5(0.55)
Treatment taken	Chemotherapy	2(0.57)	3(0.54)	5(0.55)
	Radiotherapy	259(74.21)	438(78.49)	697(76.85)
	Combination of two or three	84(24.07)	116(20.79)	200(22.05)
	No chemo	228(65.33)	406(72.33)	634(69.90)
Cycles of chemotrapy	First cycle	71(20.34)	78(13.98)	149(16.43)
	Second cycle	18(5.16)	32(5.73)	50(5.16)
	Third cycle	15(4.30)	23(4.12)	38(4.19)
	Fourth cycle	17(4.87)	19(3.41)	36(3.97)
	No child	10(2.87)	23(4.12)	33(3.64)
Number of children	1-4	24(6.88)	91(16.31)	115(12.68)
	5-8	72(20.63)	228(40.86)	300(33.80)
	9-12	134(38.40)	130(23.30)	264(29.11)
	Above12	109(31.25)	86(15.41)	195(21.50)
Family planning	No	235(67.34)	409(73.30)	644(71.00)
	Yes	114(32.66)	149(26.70)	263(29.00)
	No	256(73.35)	466(83.51)	722(79.60)
Abortion	Yes	93(26.65)	92(16.49)	185(20.40)
HIV status	No	231(66.19)	422(75.63)	653(72.00)
	Yes	118(33.81)	136(24.37)	254(28.00)
Age at marriage	>=15	236(67.62)	315(56.45)	551(60.75)
	16-20	111(37.63)	210(37.63)	321(35.39)
	21-25	2(.57)	33(5.91)	35(3.86)
Age at first birth	<=15	168(48.14)	188(33.69)	356(39.25)
	16-20	136(38.97)	271(48.57)	407(44.87)
	21-25	34(9.74)	77(13.80)	111(12.24)
	>25	6(1.72)	6(1.08)	12(1.32)
	Not give birth	5(1.43)	16(2.87)	21(2.32)

Kaplan Meier survival time plot of Cercal Cancer Patients by Different Covariates

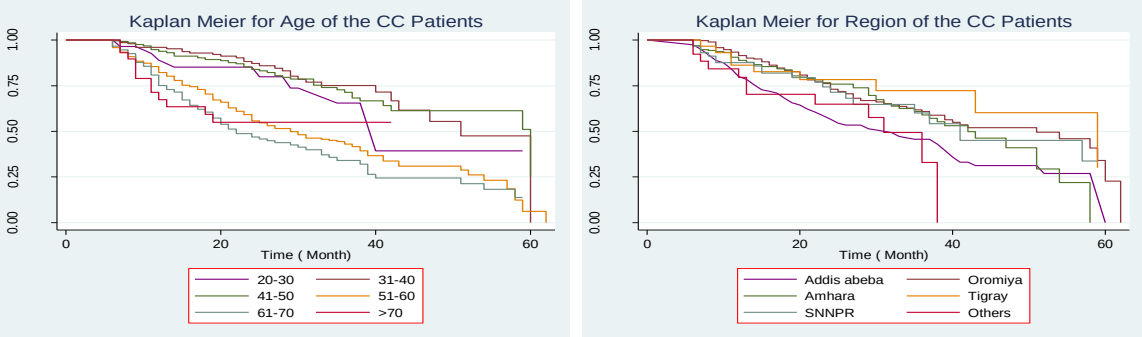

Figure 4.1: K-M survival time plot by Age and Region of patients

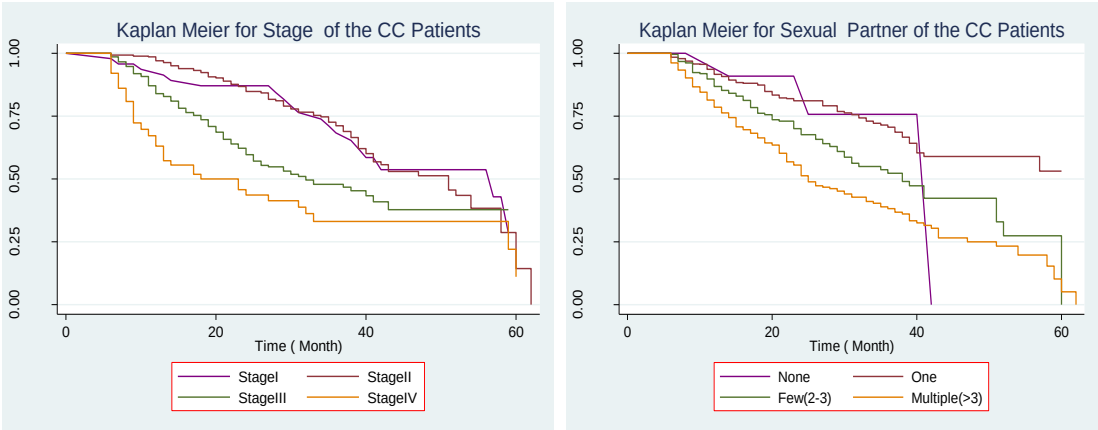

Figure 4.3: K-M survival time plot by Stage and Number of sexual partner of patients

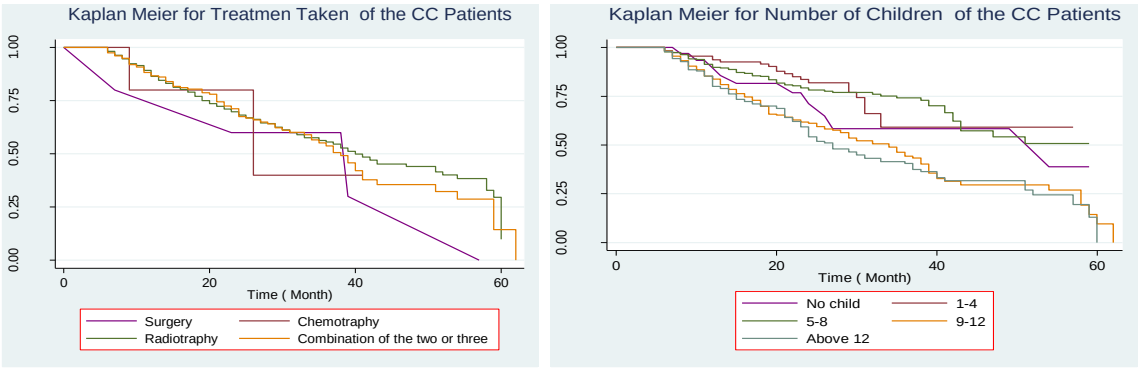

Figure 4.5: K-M survival time plot by treatment taken and number of children of patients

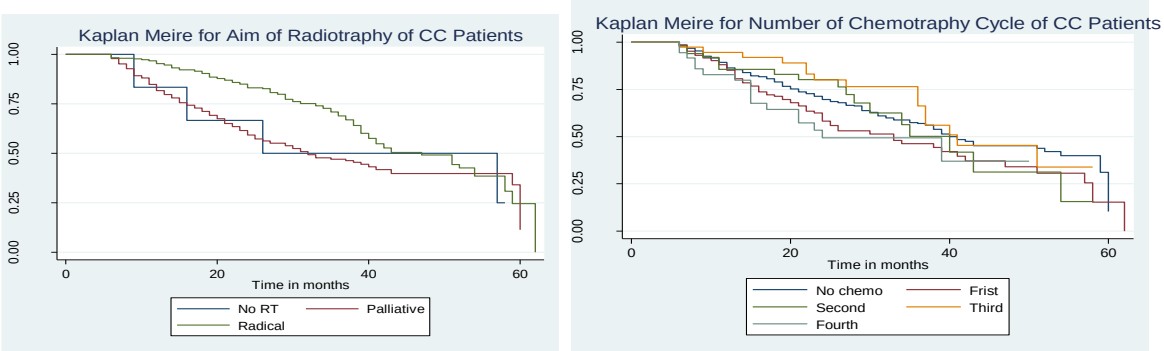

Figure 4.5: K-M survival time plot by Aim Radiotrphy and chemotrphy cycles of patients

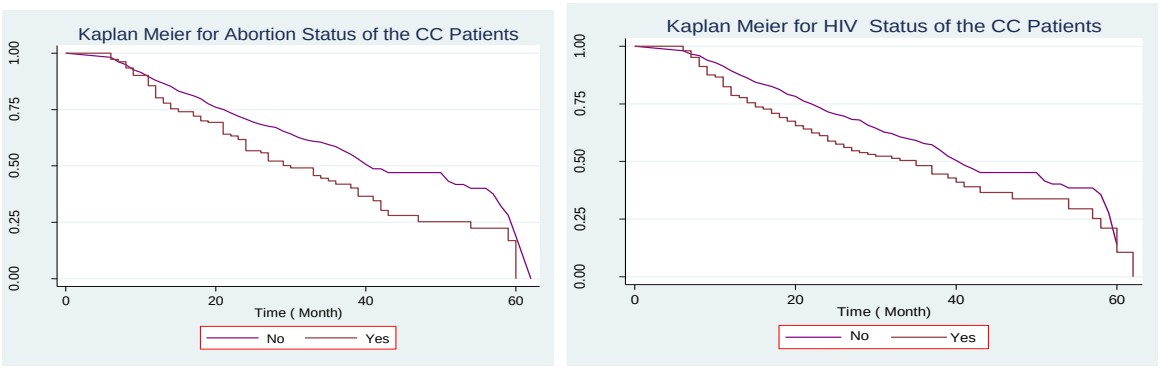

Figure 4.7:K-M survival time plot by Abortion and HIV status of patients

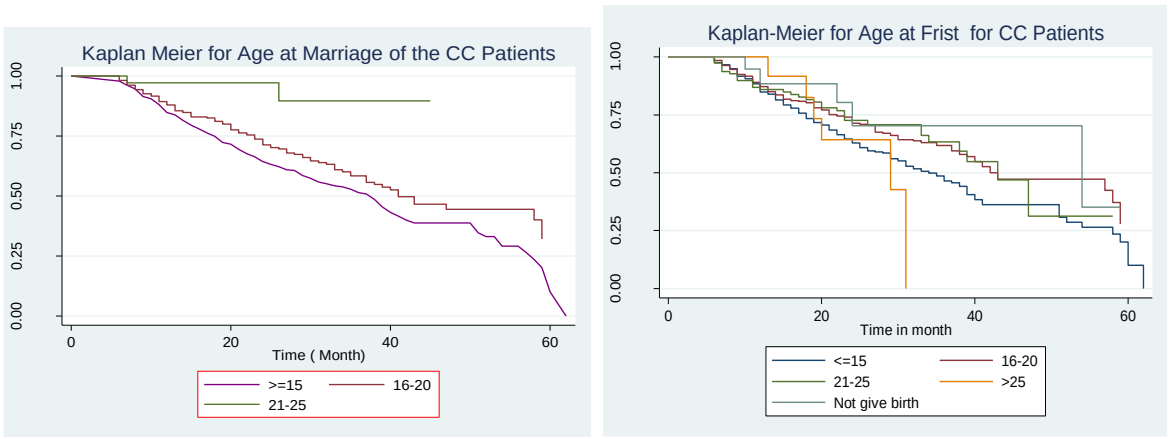

Figure 4.8:K-M survival time plot by age at marriage and Age at first birth of patients

Table 4.3: Mean survival time and confidence interval by different levels of covariates.

Covariates	Categories	mean survival in time (months)	95% Conf. Interval
Age	20-30	41	[33.00, 47.57]
	31-40	36	[32.59, 39.73]
	41-50	29	[23.97, 32.20]
	51-60	22	[21.36, 26.55]
	61-70	19	[18.70, 27.57]
	>70	11	[10.99, 26.86]
Smoking	nonsmoker	35	[31.00, 42.68]
	Smoker	16	[13.70, 27.83]
Region	A.A	21	[20.90, 26.46]
	Oromiya	33	[29.67, 35.44]
	Amhara	30	[25.65, 34.19]
	Tigray	36	[35.25, 39.12]
	SNNPR	26	[23.32, 28.20]
	Others	23	[22.02, 27.20]
Recurrence	No	31	[28.08, 31.81]
	Yes	23	[19.39, 26.78]
Stage	I	37	[35.50, 40.04]
	II	31	[29.34, 37.27]
	III	25	[23.289, 28.48]
	IV	18	[14.34, 35.30]
cycles of chemotraphy	No chemo	39	[37.67, 41.63]
	First cycle	33	[31.03, 38.95]
	Second cycle	35	[30.73, 38.11]
	Third cycle	28	[25.75, 31.45]
	Fourth cycle	30	[28.39,33.04]
Sexual partner	None	39	[30.59, 43.17]
	One	35	[33.24, 48.17]
	Few	27	[24.13, 30.93]
	Multiple	21	[18.59,28.64]
Aim of radiotherapy	no RT	33	[30.57, 43.93]
	Palliative	27	[24.01,33.52]
	Radical	37	[23.42, 41.53]

Family history	No	33	[28.47, 37.12]
	Yes	24	[21.87,33.27]
Treatment taken	Surgery	33	[21.17,41.43]
	Chemotherapy	27	[20.64, 35.56]
	Radiotherapy	35	[31.10, 40.86]
	Combination of two or three	24	[23.03,30.82]
Number of children	No child	41	[33.02, 45.82]
	1-4	37	[36.25, 41.17]
	5-8	40	[39.68, 42.32]
	9-12	31	[30.16, 36.85]
	Above12	28	[25.55, 32.73]
Family planning	No	35	[33.89, 40.95]
	Yes	29	[24.90, 30.79]
Abortion	No	38	[35.41, 42.25]
	Yes	31	[29.67, 36.36]
HIV status	No	38	[36.13, 41.91]
	Yes	24	[21.61, 29.94]
Age at marriage	<=15	27	[24.92, 29.13]
	16-20	31	[30.70,34.04]
	21-25	36	[29.71, 40.16]
Age at first birth	<=15	28	[24.17, 33.14]
	16-20	35	[31.08, 39.88]
	21-25	31	[30.96, 36.97]
	>25	29	[22.03, 30.51]
	Not give birth	40	[25.36,45.05]

Table 4.3: Univariate Analysis of Cox Proportional Hazards with time to event of CC patients.

Covariates	Category	$\hat{\beta}$	SE	Wald	Sig.	HR	95% CI for HR
	20-30	Ref					
Age	31-40	-0.42	0.38	1.21	0.27237	0.66	[0.313 , 1.388]
	41-50	-0.35	0.36	0.94	0.33145	0.70	[0.347 , 1.429]
	51-60	0.71	0.34	4.24	0.03918*	2.03	[1.036, 3.98]
	61-70	0.95	0.35	7.45	0.00631**	2.60	[1.310, 5.163]
	>70	0.68	0.44	2.37	0.02460*	1.97	[1.829, 4.681]
Smoking	nonsmoker	Ref					
Status	Smoker	0.87	0.13	47.20	6.62e-12***	2.39	[1.866 , 3.071]
Region	A.A	Ref					
	Oromiya	-0.60	0.13	20.41	6.25e-06 ***	0.55	[0.425, 0.713]
	Amhara	-0.48	0.15	10.21	0.00139 **	0.62	[0.462, 0.832]
	Tigray	-0.83	0.35	5.85	0.01558 *	0.43	[0.221, 0.854]
	SNNPR	-0.48	0.24	4.08	0.04339	0.62	[0.385, 0.986]
	Others	0.12	0.30	0.15	0.70119	1.12	[0.622, 2.024]
	No	Ref					
Recurrence	Yes	0.39	0.13	8.41	0.00372**	1.48	[1.135, 1.923]
	I	Ref					
	II	-0.33	0.26	0.02	0.8978	0.97	[0.589, 1.590]
Stage	III	0.70	0.24	8.11	0.0044**	2.01	[1.242, 3.237]
	IV	1.22	0.27	20.28	6.7e-06***	3.93	[1.993, 5.773]
Sexual	None	Ref					
partner	One	-0.05	0.59	0.01	0.9343	0.95	[0.302, 3.010]
	Few	0.55	0.59	0.71	0.3471	1.74	[0.549, 5.504]
	Multiple	0.99	0.58	2.88	0.0898·	2.68	[0.858, 8.414]
Aim of	no RT	Ref					
radiography	Palliative	0.21	0.51	0.18	0.674	1.24	[0.458, 3.349]
	Radical	-0.45	0.51	0.77	0.380	0.64	[0.234, 1.739]

Family	No	Ref					
History	Yes	0.67	0.13	26.07	3.29e-07***	1.958	[1.513, 2.535]
	Surgery	Ref					
Treatment	Chemotherapy	0.88	0.88	0.10	0.758	0.77	[0.140, 4.185]
taken	Radiotherapy	0.51	0.51	0.58	0.446	0.68	[0.253, 1.830]
	Combination of two or three	0.52	0.51	0.40	0.525	0.72	[0.264, 1.972]
	No chemo	Ref					
cycles of	First cycle	0.32	0.14	5.38	0.0203*	1.37	[1.051, 1.797]
chemotrapy	Second cycle	0.02	0.25	0.01	0.9203	1.03	[0.634, 1.656]
	Third cycle	-0.20	0.27	0.57	0.4492	0.82	[0.484, 1.379]
	Fourth cycle	0.43	0.25	2.89	0.0890·	1.53	[0.937, 2.513]
	No child	Ref					
Number of	1-4	-0.28	0.38	0.50	0.4789	0.77	[0.365, 1.604]
children	5-8	-0.25	0.34	0.56	0.4530	0.78	[0.400, 1.505]
	9-12	-0.57	0.33	2.98	0.0842·	1.76	[0.926,3.359]
	Above12	-0.62	0.33	4.00	0.0455*	1.94	[1.013, 3.709]
Family	No	Ref					
planning	Yes	0.26	0.11	5.20	0.0226 *	1.298	[1.037, 1.624]
	No	Ref					
Abortion	Yes	0.45	0.12	13.45	0.000245***	1.56	[1.23, 1.979]
	No	Ref					
HIV status	Yes	0.40	0.11	12.38	0.000434***	1.49	[1.194, 1.864]
Age at	<=15	Ref					
marriage	16-20	-0.27	0.12	5.31	0.0212*	0.76	[0.608, 0.961]
	21-25	-0.57	0.34	2.78	0.0956·	0.57	[0.292 , 1.105]
Age at first	<=15	Ref					
birth	16-20	-0.35	0.12	9.08	0.00258**	0.71	[0.562, 0.885]
	21-25	-0.37	0.19	3.91	0.04813 *	0.69	[0.476, 0.997]
	>25	0.23	0.42	0.30	0.58240	1.26	[0.556, 2.845]
	Not give birth	-0.74	0.45	2.62	0.10525	0.48	[0.197, 1.167]

SE=Standard Error,DF=Degree of freedom,HR= Hazard Ratio, CI=Confidence Interval, Ref. Reference

Test of proportional hazards assumption by Scaled Schoenfeld residuals

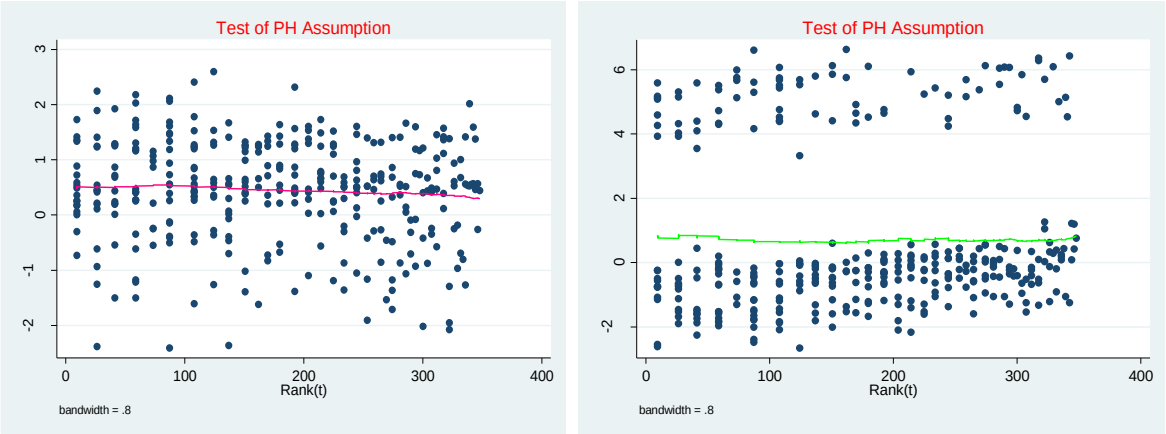

Figure 4.9 The plot of Scaled Schoenfeld residuals for Ageand Smoking to to check PH assumption

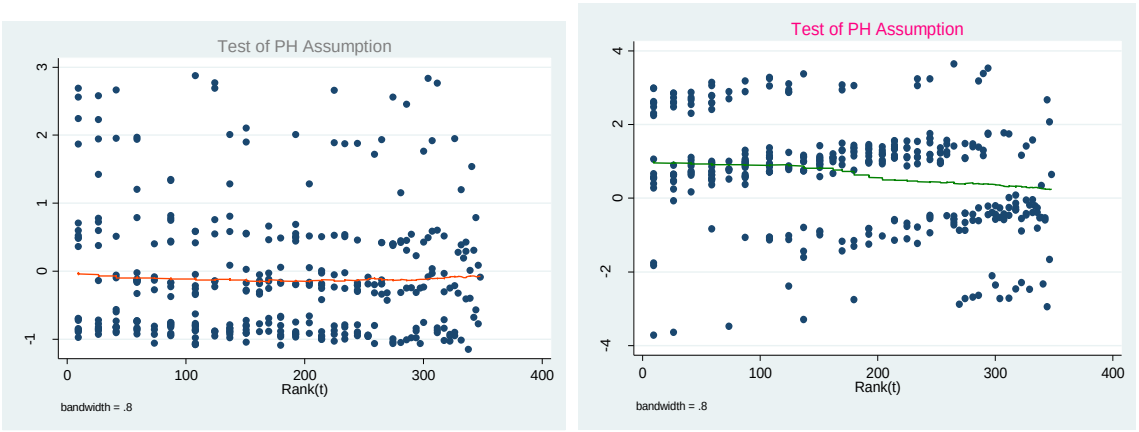

Figure 4.10: The plot of Scaled Schoenfeld residuals for Region and Stage to check PH assumption

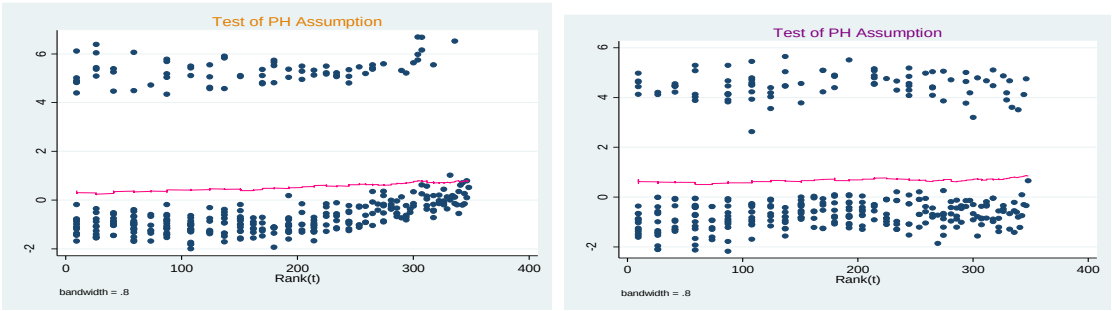

Figure 4.11: The plot of Scaled Schoenfeld residuals for Abortion and Family history to check PH assumption

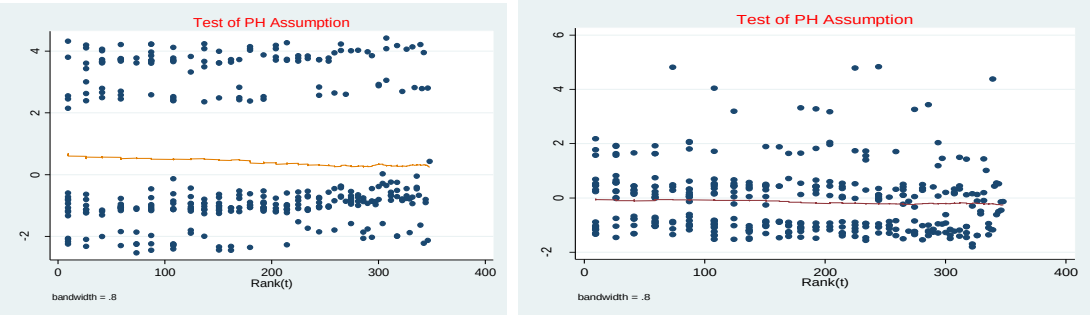

Figure 4.12: The plot of Scaled Schoenfeld residualsfor HIV status and Age at first birth to check PH assumption

Table 4.6: Univariate AFT analysis for survival time of patients modeling using different baseline hazard functions

		Baseline Distributions			
		Exponential	**Weibull**	**Log- logistic**	**Log-normal**
Covariates	**Categories**	$\hat{\beta}$ **(95% CI for** ϕ**)**	$\hat{\beta}$ **(95% CI for** ϕ**)**	$\hat{\beta}$ **(95% CI for** ϕ**)**	$\hat{\beta}$ **(95% CI for** ϕ**)**
Age	20-30	Ref	Ref	Ref	Ref
	31-40	0.42[0.724,3.212]	0.23[0.798,1.983]	0.23 [0.812,1.967]	0.24[0.826,1.971]
	41-50	0.35[0.701,2.879]	0.18[0.780,1.850]	0.17 [0.774,1.803]	0.17[0.783,1.799]
	51-60	-0.67*[0.261,0.999]	-0.45*[0.424,0.964]	-0.49*[0.406,0.925]	-0.48*[0.415,0.932]
	61-70	-0.88*[0.210,0.827]	-0.60*[0.362,0.837]	-0.68*[0.336,0.774]	-0.63*[0.350,0.806]
	>70	-0.561[0.240,1.354]	-0.386[0.397,1.163]	-0.529 [0.339,1.022]	-0.506[0.355,1.024]
Smoking Status	nonsmoker	Ref	Ref	Ref	Ref
	Smoker	-0.81*[0.348,0.517]	-0.53*[0.505,0.621]	-0.62[0.451,0.647]	-0.60*[0.459,0.661]
Region	Addis	Ref	Ref	Ref	Ref
	Ababa	0.52*[1.299,2.172]	0.35*[1.215,1.665]	0.40*[1.264,1.771]	0.40*[1.266,1.772]
	Oromiya	0.45*[1.173,2.106]	0.29*[1.120,1.602]	0.36*[1.189,1.741]	0.33*[1.154,1.686]
	Amhara	0.72*[1.050,4.037]	0.54*[1.133,2.570]	0.56*[1.133,2.691]	0.48*[1.077,2.437]
	Tigray	0.43[0.960,2.445]	0.30[1.016,1.794]	0.32[1.009,1.863]	0.27 [0.972,1.761]
	SNNPR	-0.05[0.526,1.708]	-0.07[0.654,1.336]	-0.03 [0.647,1.467]	-0.07[0.622,1.399]
	Others				
Recur	No	Ref	Ref	Ref	Ref
	Yes	-0.38*[0.524,0.887]	-0.23[0.672,0.932]	-0.23*[0.657,0.955]	-0.25*[0.647,0.941]
	I	Ref	Ref	Ref	Ref

Stage	II	0.08[0.662,1.781]	-0.01 [0.745,1.316]	0.98 [0.756,1.334]	0.06[0.793,1.414]
	III	-0.48*[0.388,0.996]	-0.47*[0.475,0.816]	-0.53*[0.447,0.770]	-0.48*[0.467,0.813]
	IV	-0.97*[0.223,0.641]	-0.75* [0.348,0.642]	-1.02*[0.264,0.498]	-0.93*[0.287,0.542]
Family History	No	Ref	Ref	Ref	Ref
	Yes	-0.60*[0.426,0.712]	-0.42*[0.561,0.771]	-0.43*[0.542,0.775]	-0.42*[0.549,0.793]
Abortion	No	Ref	Ref	Ref	Ref
	Yes	-0.41*[0.521,0.839]	-0.27[0.660,0.877]	-0.29[0.636,0.884]	-0.27*[0.646,0.901]
HIV status	No	Ref	Ref	Ref	Ref
	Yes	-0.38*[0.546,0.851]	-0.26*[0.671,0.882]	-0.31*[0.632,0.854]	-0.30*[0.637,0.859]
age at first birth	<=15	Ref	Ref	Ref	Ref
	16-20	0.36*[1.137,1.788]	0.22*[1.085,1.436]	0.21*[1.061,1.437]	0.19*[1.041,1.408]
	21-25	0.40*[1.031,2.155]	0.23 [0.999,1.577]	0.23[0.995,1.601]	0.20[0.961,1.534]
	>25	-0.16[0.378,1.926]	-0.20 [0.977,1.354]	-0.06 [0.557,1.602]	-0.004[0.554,1.791]
	Not give birth	0.71 [0.833,4.936]	0.47[0.925,2.675]	0.48[0.918,2.837]	0.46[0.919,2.720]

95% CI ϕ: 95% confidence interval for acceleration factor, **Indicates significant at 10%level of significance, Ref= reference*