DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

POSTGRADUATE STUDIES

DEPARTMENT OF STATISTICS

TIME TO DEATH AND ASSOCIATED FACTORS OF COLORECTAL CANCER PATIENTS IN TIKUR ANBESSA SPECIALIZED HOSPITAL, ADDIS ABABA, ETHIOPIA

By:

YESEWZER ENAWGAW

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF DEBRE BERHAN UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS (BIOSTATISTICS

JUNE, 2021

DEBRE BERHAN

ETHIOPIA

DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

POSTGRADUATE STUDIES

DEPARTMENT OF STATISTICS

TIME TO DEATH AND ASSOCIATED FACTORS OF COLORECTAL CANCER PATIENTS IN TIKUR ANBESSA SPECIALIZED HOSPITAL, ADDIS ABABA, ETHIOPIA

By:

YESEWZER ENAWGAW
ADVISOR: BEZAREDE M. (ASSI.PROF)
CO-ADVISOR: FEKADE G. (MSC)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF DEBRE BERHAN UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS (BIOSTATISTICS)

JUNE, 2021

DEBRE BERHAN,

ETHIOPIA

DECLARATION

I hereby declare that work which is being presented in this thesis entitled "**Time To Death And Associated Factors Of Colorectal Cancer Patients In Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia**" is performed under the supervision of my research advisor Mr. Bezarede Mekonnen (Assi. Prof) and Mr. Fekade Getabil. It is my original work, has not been presented for a degree or diploma in any other university, and that all sources of the material used for the thesis have been duly acknowledged.

Yesewzer Enawgaw

Signature ………………………………….

This thesis has been submitted for examination with approval as university advisors

Bezarede M. (Assi. Prof)   …………………   ………………

Advisor   Signature   Date

Fekade G.(MSc)   ………………….   ………………

Co-advisor   Signature   Date

APPROVAL SHEET

We, the undersigned member of the broad examiners of the final open defense by Yesewzer Enawgaw have read and evaluated his thesis entitled **"Time To Death And Associated Factors Of Colorectal Cancer Patients In Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia**" and examined the candidate. This is therefore to certify the thesis has been accepted in partial fulfillment of the requirement for the degree of master of sciences in statistics with specialization of biostatistics.

_____   _____   _____

Name of Advisor   Signature   Date

_____   _____   _____

Name of Co-Advisor   Signature   Date

_____   _____   _____

Name of External examiner   Signature   Date

_____   _____   _____

Name of Internal examiner   Signature   Date

_____   _____   _____

Associate Dean, College Res/CS& Postgraduate   Signature   Date

_____   _____   _____

Dean College of Graduate Studies   Signature   Date

Final approval and acceptance of the thesis is contingent upon the submission of the thesis of the final copy of the thesis to the college of graduate studies (CGS) through the department of the graduate committee (DGC) of the candidates.

Stamp of CGS   Date _____

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Contents   **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF GLOSSARY TERMS

Event: Death of patients due to colorectal cancer.

Follow-up period: The time from the beginning of the study period to an event, end of the study, or loss of contact or withdrawal from the study.

Time to death: Time from the first confirmed diagnosis date of colorectal cancer to death.

Comorbidity: According to the International Classification of Disease-10, Disease from Charles comorbidity index was used during data collection. The co-occurrence of any of these diseases with colorectal cancer at the time of diagnosis is labeled as a "yes" response (Lin et al., 2019).

Incomplete data: When one of the independent variables was not registered (stage, primary site, age, residence, comorbidity, etc).

Body Mass Index (BMI) according to disease prevention and control.

Underweight: BMI less than 18.5 Kg/m^2

Healthy weight: BMI 18.5 to 24.9 Kg/m^2

Overweight: BMI from 25 to 29.9 Kg/m^2

Obese: BMI 30Kg/m^2 or higher (Center of Disease Control and Prevention, 2018)

Stage at diagnosis: According to the American Joint Committee of Cancer (AJCC)

Stage I: Tumor invades muscularis propria, submucosa, no lymph node, and no metastasis

Stage II: Tumor invades muscularis propria, penetrates to the surface of the visceral peritoneum, adherent to other organs or structure, no lymph node and no metastasis

Stage III: Tumor metastasis in seven or more regional lymph nodesherent to other organs or structure, no lymph node, and no metastasis

Stage IV: Tumor metastasis into different organs (Amin et al., 2017).

LIST OF ABBREVIATIONS

ACA: American Cancer Association

ACS: American Cancer of Society

ACG: American College of Gastroenterology

ACIM: Australian Cancer Incidence and Mortality

AFT: Accelerated Failure Time

AIC: Akaike's Information Criterion

AHR: Adjusted Hazard Ratio

BIC: Bayesian Information Criterion

BMI: Body Mass Index

CRC: Colorectal Cancer

FMH: Federal Ministry of Health

GLOBOCAN: Global Organization Board of Cancer Association Network

HR: Hazard Ratio

LMIC: Low and Middle Income Country

NCD: Non-Communicable Disease

NCI: National Cancer Institute

NCCPE: National Cancer Control Plan of Ethiopia

TASH: Tikur Anbessa Specialized Hospital

WHO: World Health Organization

US: United States

ABSTRACT

Colorectal cancer (CRC) is a cancer of the large intestine. Anatomically, it also is known as colon cancer or rectal cancer but when both present with similar features they are termed colorectal cancer. It is a common health problem, representing the third most commonly diagnosed cancer worldwide and causing a significant burden in terms of morbidity and mortality. In Ethiopia CRC is the third most common cancer next to breast and uterine cancer for females, but the first most common cancer among the male population. The main objective of the study was to investigate the associated factors that affect the time to death of CRC patients in TASH, Addis Ababa, Ethiopia by using survival models. The retrospective cohort study was conducted on 325 CRC patients who enrolled between January 1, 2017, and December 30, 2020, in TASH, Addis Ababa, Ethiopia. From 325 patients, 111 (34.15%) died. The overall median survival time of colorectal cancer patients is 23 months. Kaplan-Meier survival curves and Log-Rank test were used to compare the survival experience of different categories of patients. Five AFT fitted models (Exponential, Weibull, Log-normal, generalized gamma, and Log-logistic) were compared by using AIC and BIC. The log-logistic AFT model was found to be the best model to fit the data. Based on the log-logistic model, marital status, stage, family history, alcohol consumption, physical exercise, tumor grade, and treatment modality were found to be the most prognostic factors of time to death of CRC patients at 5% levels of significance. The result showed that non-alcohol user and married patients were prolonged the survival time whereas, CRC patients who were diagnosed at stage IV (metastatic), patients who did not do physical activity, patients who had no family history, patients diagnosed with poorly differentiated tumor grade, patients diagnosed as chemotherapy alone and chemo plus surgery were shortened the survival time. From our findings, it is better to implement colorectal cancer early screening and detection programs to improve survival outcomes.

Keywords: colorectal cancer (CRC), Cox-PH, time to death, log-logistic

1. INTRODUCTION

1.1. Background of the Study

Cancer is a group of diseases that cause cells in the body to change and spread out of control. It is an important global health problem in developed and developing countries that is the most important cause of morbidity and mortality (Niksic et al., 2016; Bray et al., 2018). Colorectal cancer (CRC) is a cancer of the large intestine. Anatomically, it also is known as colon cancer or rectal cancer but when both present with similar features they are termed colorectal cancer (CRC) (Olsen, 2015).

The American Cancer of Society (ACS) and American College of Gastroenterology (ACG) documented that CRC starts in the colon or the rectum which originates from pre-cancerous growths or polyps that grow in the colon or rectum, but their progression to CRC could be halted if it is detected early and polyps are removed (Karlitz et al., 2017; Wolf et al., 2018). Clinical presentation of CRC depends on its size, presence, or absence of metastatic and tumor location. Most colorectal cancers occur sporadically and are characterized by a sequenced carcinogenesis process that involves the progressive accumulation of mutations in a period that lasts on average 10–15 years (Arvelo et al., 2015; Ang et al., 2017). Early CRC often has no symptoms, as a tumor grows, it may bleed or block the intestine. The most common symptoms are bleeding from the rectum, blood in the stool, a change in bowel movements, weight loss, and fatigue (NCI, 2014).

Globally, CRC is the third most commonly occurring cancer and the second most common cause of cancer death next to lung cancer in men and breast cancer in women (WHO, 2018). According to the Global organization board of cancer association network (GLOBOCAN) estimated that about 1.8 million new CRC cases (9.2%) whereas the death rate was 9.0% (Bray et al., 2018). In the US, the incidence rate was higher in black 56.1 per 100,000 than 45.6 per 100,000 for whites, of those males had an incidence rate 53.4 per 100,000 than females 39.9 per 100,000 with an overall incidence of 45.9 per 100,000, while the mortality rate was 14.5 per 100,000 men and women annually (Ansa et al., 2018). In Iran, it is the third most common cancer. The increasing incidence of CRC in the past decades in Iran has made it a major public health problem

(Dolatkhah et al., 2015). According to the Iranian Annual National Cancer Registration Report, CRC is the third most common cancer in Iranian women and the fifth common cancer in men (Pourhoseingholi and Zali, 2012).

CRC is the 4th most common cancer in the World Health Organization-Africa region. The crude incidence of CRC in Sub-Saharan Africa (SSA) for both men and women was found to be 4.04 per 100 000 population (4.38 for men and 3.69 for women), about 24,711 new cases were estimated annually (Graham et al., 2012). Recent studies have shown that the incidence and mortality of CRC are increasing in low and middle-income country's (LMIC), especially in Sub-Saharan Africa (Katsidzira et al., 2017). This rising burden is mirrored in Nigeria, where more than half of the patients die within one year of diagnosis. In the African population, patients with CRC tend to present at a younger age with advanced disease (Katalambula et al., 2016; Ibrahim et al., 2011). Currently, in Sub-Saharan Africa, CRC is the commonest cancer that confirmed cases for 5.6% in males and 3.7% in females (McCormack and Newton, 2019).

In Ethiopia, CRC is the third most common cancer next to breast and uterine cancer for females, but the first most common cancer among the male population (Woldu et al., 2017). The organization of healthy in Ethiopia mostly focused on the switch of communicable disease and, there is insufficient screening center, treatment facilities, and, unwell organized referral (FMH, 2020). Therefore, non-communicable disease (NCD) like colorectal cancer is increasing rapidly (FMH, 2020). Having these, this study was to investigate the associated factors that affect the time to death of CRC patients in TASH using survival models. Survival analysis is a statistical method for data analysis where the response variable is the time to the occurrence of an event, death of CRC patients in Tikur Anbessa Specialized Hospital (TASH) in this study.

1.2. Statement of the Problem

The global burden of CRC raised from 1.36 million to 1.80 million within 2012-2018, of which about 881,000 death cases were documented (WHO, 2018; Ferlay et al., 2015). The GLOBOCAN in 2012 estimated that the incidence of males varied from less than 5 per 100,000 in African countries to higher than 40% in Europe, Northern American, and Oceania, with the three highest rates seen in Slovakia (61.6%), Hungary (58.9%) and the Republic of Korea (58.7%). The lowest rate was seen in sub-Saharan Africa, specifically in Gambia and

Mozambique both 1.5 per 100 000 (Arnold et al., 2017; Gandomani et al., 2017).CRC has a great impact on the quality of life involving physical, psychological, and socioeconomic dimensions (Quach et al., 2015).

According to the US National Cancer Institute (NCI), the expenditure rate on CRC care increased from 14.1 US dollars to 16.3 US dollars (Banegas et al., 2018). This indicates that patients couldn`t afford to pay for treatment and transportation during a long hospital stay. Despite cost-effective increment in treatment coverage and realization of early screening strategies in Sub-Saharan Africa and Southeast Asia, the death rate has been greater; whereas the US implemented early screening for CRC every 3 years (Ginsberg et al., 2012).

In Ethiopia, the 2011-2014 Addis Ababa cancer registries reported that the incidence rate of CRC was 19% from all cancer have been reported among the male population (Timotewos et al., 2018). Some studies have been conducted related to CRC in Ethiopia. For instance, Atinafu et al., (2020) conducted research about survival status and predictors of mortality among colorectal cancer patients by using cox proportional hazard regression model. Etissa et al., (2021) also carried out a study about the Prognosis of colorectal cancer by using cox proportional hazard regression model.

And (Teka et al., 2021) also conducted histological characteristics, survival patterns, and prognostic determinants among colorectal cancer patients by using the Cox PH model. However, under certain circumstances, the Cox PH model can give less precise estimates to analyze survival data than parametric models (Efron, 1977). So, this study was attempted to fill the research gap by using non-parametric, semi-parametric, and parametric survival models and identify the important socio-demographic and other associated factors that affect the time to death of CRC patients.

Research Questions

❖ How can estimate the survival probability and compare the survival curves of time to death of CRC patients among different levels of covariates?

❖ Which factors have major effects on the time to death of CRC patients in TASH?

1.3. Objective of the Study

1.3.1. General Objective

❖ The general objective of this study was to investigate the associated factors that affect the time to death of CRC patients in TASH enrolled from January1^{st}, 2017 to December30^{th}, 2020.

1.3.2. Specific Objectives

❖ To estimate the survival probability of CRC patients in TASH.
❖ To compare the survival probability of CRC patients among different levels of the covariate.
❖ To assess the association between the covariates and hazard ratio.
❖ To identify the factors significantly associated with time to death of CRC patients in TASH.

1.4. Significance of the Study

Studying time to death of colorectal cancer patients has important practical implications for patients and society at large to know how their prognosis is changing over time and what their life expectancy is based on disease status, to provide an essential indicator for early detection and improvement in CRC treatment modalities and improve quality of care.

The result of this study was important for nurses to provide effective high quality based care to CRC patients. Knowing factors that affect the time to death of CRC patients helps nurses to reduce treatment-related errors, to select the best care to be given for the patients, and to provide a precise decision to clinicians and patients. This study could be to educate and motivate patients, how to increases the survival time.

Moreover, this study would be an important input to policymakers, oncology program managers, and health professionals to implement early detection, prioritize intervention, estimate the survival rate of patients, and make an evidence-based decision on CRC, to guide the national

cancer control program, to support the planning systems for better cancer control and prevention program.

1.5. Limitation of the study

One of the limitations of this study is that important variables such as educational status, socio-economic status of patients, and multidisciplinary care were not available.

1.6 Organization of the Study

This section shows the hierarchical developments of the study. The introduction of the study consists background of the study, a statement of the problem, the objective of the study, the significance of the study, and the limitation of the study. The literature review was organized under Chapter two; in this chapter, all related literature was discussed. In chapter three, data and methodology were organized and the following section was discussed; description of study area, source of population, study population, study design, study variable, data extraction and procedure, statistical models, estimation and inference, variable and model selection. Proportional hazard and checking and model diagnostics among these. Statistical data analysis was under Chapter 4, all statistical analysis, variable selection, and model diagnostics were reported. The last chapter of this study is the discussion, conclusion, and recommendations of the study.

2. LITERATURE REVIEW

Cancer is a disease characterized by the unchecked division of abnormal cells. When this type of growth occurs in the colon or rectum, it is called CRC. It is the third common cancer and the second cause of mortality among males and females tends to be the major cause of the health-related problem (Peterse et al., 2018).

2.1. Prevalence of Colorectal Cancer

CRC is the third prevalent cancer leading to death in western countries (Merrill et al., 2013). In 2017, about 1,348,087 people living with colorectal cancer in the United States. There are an estimated 50,630 deaths from colorectal cancer (CRC) in 2018 in the US, second only to lung cancer (Siegel, 2018). According to the national cancer institute the prevalence of colorectal cancer disease ware 9 percent of females and 10 percent of males in the US. It is the second most commonly diagnosed invasive cancer in Australia, with 13,076 people being diagnosed in 2005 (representing 13.0% of all new cancer diagnoses) and 4,168 dying from the disease (10.7% of all cancer deaths and 3.2% of all deaths) (ACIM, 2009). CRC accounted for over 600000 of those deaths, with 70% occurring in (LMIC)(WHO 2012, Meetoo, 2008).In Ethiopia, CRC is the third most prevalent cancer among the entire adult population, and patients often present with advanced stages of cancer (NCCPE, 2015). It becomes prevalent to account (12.2%) in males and (4.4%) in females of all ages (Solomon and Mulugeta, 2019).

2.2. Factors Associated with Colorectal Cancer

Family history: Up to 30% of CRC patients have a family history of the disease, making this one of the most important and actionable risk factors (Lowery et al., 2016; Jones et al., 2020). People with a first-degree relative (parent, sibling, or child) who has been diagnosed with CRC have 2 to 4 times the risk of developing the disease compared to people without this family history, with a higher risk for diagnosis before age 50 and/or multiple affected relatives(Lowery et al., 2016).

BMI: Risk is believed to increase with increased BMI since there's an increase in circulating estrogens and a decrease in insulin sensitivity; therefore, this is assumed to influence CRC risk (Rasool et al., 2013).

Physical inactivity: Physical activity is strongly associated with a reduced risk of CRC (Rasool et al., 2013). Studies suggest a dose-response effect, that reflects the regularity and intensity of physical activity which is inversely related to CRC, and studies find as little as 7 hours a week could lower the risk of CRC (Giovannucci and Willett, 1994).

Smoking: In November 2009, the International Agency for Research on Cancer reported that there is sufficient evidence to conclude that tobacco smoking is a cause of CRC (Secretan et al., 2009).

Alcohol: Alcohol consumption is a known risk factor determinant for the onset of CRC at a young age (Haggar and Boushey, 2009). Increased alcohol consumption leads to increase CRC risk. Alcohol intake is related to a higher risk of colorectal adenoma, which consequently provides an increased risk (Giovannucci and Willett, 1994).

Age: The CRC is most common in people older than 50, and the chance of getting CRC goes up with each decade past age 40 (Fiorot et al., 2018).

Gender: CRC is more common in men. Men and women are equally at risk for colon cancer, but men are more likely to get rectal cancer (Nasaif and Al Qallaf, 2018).

Stage and Tumor grade: These are important prognostic factors for CRC (Lyall et al., 2006). Lower-stage tumors and well-differentiated tumors have better outcomes of CRC (Yamauchi et al., 2012). Tumors without lymphovascular invasion or distant metastases have a more favorable prognosis than those with invasion or metastases (Kawazoe et al., 2015).

Surgery: the gold standard treatment in non-metastatic disease is surgery, with 5- year survival rates ranging from 44% to 93% depending on the stage(O'Connell et al., 2004). It remains the primary course of treatment in cases of early diagnosis but is no longer effective in advanced cases where cancer has metastasized, as is the case in about 25% of diagnoses (Kekelidze et al., 2013).In such patients, the efficacy of neoadjuvant, cytotoxic therapies has been stifled by the rapid evolution of drug resistance and cancer recurrence (Colussi et al., 2013).

Sex, comorbidity, and stage of the disease are the most important factors for CRC patient mortality (Saidi et al., 2011).

2.3. Empirical Literature about Associated Factors That Affects Mortality of Colorectal Cancer Patients

2.3.1. Socio-Demographic Factors

Atinafu et al., (2020) conducted research about survival status and predictors of mortality among CRC patients by using the Cox proportional hazard model. They found that socio-demographic factors such as age, marital status, smoking, alcohol consumption, and comorbidity were significant predictors of CRC mortality. Patients aged 70 and over were 1.7 times at higher risk to die than those aged below 40 years old. A study conducted in Malaysia revealed that older age CRC patients are less survival time than patients at a younger age (Hassan et al., 2016).

A study revealed patients who were age <50 years old had a significantly longer survival time than patients who wear age>50 years old (Fiorot et al., 2018). In addition to this, younger age group CRC patients had better survival compared to older age patients (>60 years) (Hassan et al., 2016). The age-stratified analyses revealed that age was a statistically significant predictor of mortality risk; there was a 3% higher mortality risk with an annual increase in age (Tannenbaum et al., 2014).

The research conducted in Netherland indicated that the 5- year survival rate of patients age <63.26, 63.21- 71.61, 71.61-79.49, >79.49 was 58.2%, 58.8%, 51.5%, 40.8% respectively, with a median survival of colon cancer patients was 5.13 years, to some extent lower median survival in rectal cancer patients which was 4.67years, so patents age (age) was significant determinant factors of CRC mortality (van Eeghen et al., 2015).

Patients who married 2.4 times (95% CI: 1.5-3.8), widowed 2.4 times (95% CI: 1.2-4.6), and divorced 2 times (95% CI: 1.1-3.7) were at higher risk of mortality than single marital status (Atinafu et al., 2020). In contrast, a study in Taiwan showed that married status had better survival than single, divorced, and widowed status similarly another study in Florida designated that patients who divorced 1.22 times, single 1.29 times and windowed 1.19 times were at higher risk of mortality than married marital status (Tannenbaum et al., 2014). The same study publicized that sex also determines the survival outcomes of colorectal cancer patients as shown by AHR of 1.00 in males and 0.85 in females 95% CI(0.82-0.88) at p<0.001. Another follow-up

study in Jordan indicated that the 5-year survival rate of CRC of males and females was 54.8% and 58.1% respectively (Sharkas et al., 2017). The risk of death was 1.1 times (95% CI: 1.08–1.15) greater in men than in women (Yung-Heng Lee, et al.,2019).

The survival analysis revealed that married patients with CRC have superior survival compared to unmarried (single) patients and that unmarried patients were 30 % higher risk of death due to CRC compared to the married patients (AHR 1.30; CI 1.17, 1.44) (Alyabsi et al., 2021)

CRC Patients having comorbid conditions were 1.8 times at high hazard to die than patients with non-comorbid conditions CI: 1.3-2.5 (Atinafu et al., 2020). Moreover, a study conducted in Japan showed that patients in CRC with comorbid conditions were 1.2 times (95% CI, 1.08–1.34) at greater risk to die than patients with non-comorbid conditions (Morishima et al., 2018). Similarly, a study conducted in Spain found that CRC patients who had comorbid conditions experienced lower survival (56.0%) than patients with non- comorbid conditions (Parés-Badell et al., 2017). BMI (HR, 0.96; 95% CI, 0.96-0.975) has a significant effect on patient's CRC-related mortality (Moamer et al., 2017).

CRC patients who smoke cigarettes' and alcohol user were 1.6 and 1.5 times at high hazard to die than non-smokers (CI: 1.1-2.3) and non-alcohol users (CI: 1.07 -2.2) respectively (Atinafu et al., 2020). Similarly, a prospective study in Oslo University revealed that non-smokers had 0.79 times (95%CI 0.64-0.99) lower risk to die as compared to smokers (Japuntich et al., 2019). Heavy drinking is also associated with poorer survival after a CRC diagnosis than light drinking. Mainly, lifetime heavy drinkers exhibited poorer overall (AHR: 1.37; 95% CI: 1.06, 1.78) and disease-free (AHR: 1.38; 95% CI: 1.09, 1.74) survival (Walter et al., 2017).

Research done in Iran indicated that urban residence CRC patients have better survival outcomes than rural residence (Semnani et al., 2016). Colorectal cancer patients with having family history had an 11% reduction in the risk of death compared to patients with no family history (HR=0.89,95% CI: 0.81-0.98, P=0.02) (Morris et al., 2013). Conversely, an analysis of 2090 incidence CRC cases within the prostate, lung, colorectal, and ovarian cancer screening trial (PLCO) showed an increased risk of CRC-pacific mortality(HR, 3.31; 95% CI,1.02-1.69) individuals with a family history of CRC, compared to those without a family history o CRC(Schoen et al., 2015).

In men, the lowest income class showed a higher risk of rectal cancer (OR 1.37 [1.18- 1.59]) than the highest income class (Kim et al., 2012).

The adjusted risk for cancer in the proximal colon was statistically significantly decreased for participants with vocational secondary education (HR0.76, 95% CI 0.58–0.98) and a primary education or less (HR0.73, 95% CI 0.57–0.94) (Leufkens et al., 2012).

2.3.2. Clinical and Pathological (clinicopathological) Related Factors

Primary tumor site, distant metastasis, lymph node involvement, and treatment modalities were significantly associated with the survival rate of CRC patients (Etissa et al., 2021). According to (Etissa et al., 2021) Patients diagnosed with rectal cancer had a 76% (HR 1.761, 95% CI: 1.173–2.644) increased risk to die compared to colon cancer. A study conducted in Iran told that tumor grade is a significant predictor of CRC patient's death (Ahmadi et al., 2015). Tumor grade and disease stage at diagnosis are significant predictors of CRC mortality (Atinafu et al., (2020). Based on (Atinafu et al., 2020) CRC patients diagnosed as undifferentiated tumor grade were 1.7 times at high hazard to death than those who were a well-differentiated type of tumor (CI:1.17-2.4).

A retrospective cohort study showed that using cox proportional hazard modeling, the cancer stage is significantly associated with time to death. The study conducted in Shahid Beheshti University of medical sciences used the Weibull model stage III (HR, 1.69; 95% CI, 1.246-2.315) and Stage IV (HR, 4.51; 95% CI,2.91-6.99) have a significant effect on patient's CRC related mortality (Moamer et al., 2017). According to (Moamer et al., 2017) Patient who was diagnosed at stage III were 1.69 times (95% CI, 1.246-2.315) and stage IV were 4.51 times (95% CI, 2.91-6.99) at high hazard to die than those who were diagnosed as stage I. Patients who were diagnosed at disease stage IV were 17.6 times at high hazard to die than those who were diagnosed as disease stage I (CI: 6.1-50.0) (Atinafu et al., (2020). The death rate was 2.7 times higher for those diagnosed at stage IV compared to stage I and II (AHR = 2.66, 95%CI: 1.44–4.91). Whereas, the rate of death was nearly 5 times higher for mucinous or signet-ring cell carcinoma than adenocarcinoma NOS (AHR=4.92, 95%CI: 1.75–13.8) (Teka et al., 2021).

Node-positive patients were 3.146 times (95% CI: 1.629 6.078) higher to die compared to node-negative patients (Etissa et al., 2021). A retrospective study in the republic of Korea indicated that locally advanced primary tumor (high patient stage, positive regional lymph node, and local residual primary tumor) was associated with lower outcome survival in incurable stage IV colorectal cancer patients (Kim et al., 2018). The risk of mortality for metastatic cancer was 4.221 (95% CI: 2.788–6.392) times higher than non-metastatic patients (Etissa et al., 2021). Mucinous carcinoma type of CRC had a significant indicator of an outcome as shown the survival rate of 81.4% than non-mucinous (87.4%) (Park et al., 2015).

2.3.3. Treatment-Related Factors

A patient who underwent surgical removal of a primary tumor and received chemotherapy had a median overall survival of 18.3 months compared with 8.4 months (95% CI) if they were treated with chemotherapy alone (P<.0001) (Bekaii-Saab et al., 2019). Chemotherapy is a factor significantly associated with CRC patient's death (Saidi et al., 2011).

Observation of population cohort study in Scotland showed that the five-year Cause-specific survival rate of patients cared for in a multidisciplinary team was 63.1% which is greater than patients who was not cared for by the multidisciplinary team (48.2%) with 0.73 times lower risk of die than the non-multidisciplinary team (95% CI:0.53-1.00,p=0.047). Therefore being cared for by the multidisciplinary team can influence the survival outcomes of the patient in different CRC stages (Munro et al., 2015).

Retrospective study analysis Japan revealed that 3 years overall survival rate patients treated with adjuvant chemotherapy and surgery alone were 93.5% and 81.7%; p<0.001 respectively, so being treated with adjuvant therapy improves the survival outcome because of addressing the left tumor from the primary therapy and reduce relapse rate (Tashiro et al., 2014). Cox proportional analysis revealed that use of chemotherapy (HR=50.47, 95% CI: 50.41-0.54), SRPT (HR=50.49, 95% CI: 50.41-0.58), second-line Chemotherapy (HR=50.47, 95% CI: 50.45-0.64), and metastasectomy (HR=50.54, 95% CI: 50.45-0.64) were associated with superior survival (Bekaii-Saab et al., 2019). The risk of mortality was 36.1% lesser (HR: 0.639 (95% CI: 0.418–0.977)) and 47.9% lesser (HR: 0.521 (95% CI: 0.279–0.973)) in those patients who received

adjuvant chemotherapy and who received adjuvant chemotherapy plus adjuvant radiotherapy, respectively compared to patients who had only surgical resection (Etissa et al., 2021).

2.4. Overview of Survival Model

Survival analysis is the analysis of survival data in which the outcome variable of interest is time until some event occurs. Yet relatively little has been written about their formal statistical theory (Kaplan and Meier, 1958) gave a comprehensive review of earlier work and many new results. (Cox, 1972) was largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and more generally to the incorporation of regression-like arguments into life table analysis. Survival models have the capability of handling censored data. (Cox, 1972) and Cox and Oakes (1984) used survival analysis in modeling human lifetimes. Fergusson et al. (1984) used hazard functions to study the time to marital breakdown after the birth of the child. Hazard functions had been also used in studies of time to shift in attention in the classroom (Felmlee*et. al*., 1983), in the study of relapse of mental illness (Lavori et al., 1984), marital dissolutions (Morgan et al., 1988), and human lifetimes (Gross *et al*., 1975).

 (Cox, 1972), introduced a semi-parametric survival model. This model is based on the assumption that the survival times of distinct individuals are independent of each other. This assumption holds in many experimental settings and widely applicable. However; there are instances in which this assumption may be violated.

The parametric AFT model provides an alternative to the PH model for statistical modeling of survival data (Wei, 1992). AFT model is used in industrial fields and is seldom used in the case of survival data. If the appropriate parametric form of the AFT model is used then it offers a potential statistical approach in the case of survival data which is based upon the survival curve rather than the hazard function. It is known as the Accelerated failure time model because the term "failure" indicates the death, disease, etc. and the term "Accelerated" indicates the responsible factor for which the rate of failure is increased. That factor is called the "Accelerated factor" .the AFT model is known as the log-location scale model given by Lawless (1982). According to the literature found, Pike (1966) proposed the AFT model in the case of carcinogenesis data. He develops the basic statistical methodology and discussed likelihood estimation for the Weibull distribution.

3. Data and Methodology

3.1. Study Area

Data were used from Addis Ababa Population-Based Cancer Registry (AAPBCR), which was established in 2011 under the TASH radiotherapy center. TASH is a tertiary-level hospital equippedwith cancer diagnostic and treatment facilities and is one of the cancer treatment centers in Ethiopia. TASH is a teaching, central tertiary generalized referral hospital with approximately 800 inpatient beds**.** It is the largest and well-known public hospital which was built in the early 1960s. The hospital hosts a Cancer Treatment Center. The registry uses hospitals, higher diagnos tic clinics, and pathology services as the main source of cases. The Hospital is geographically located between $9^0 0^' 0''$ to $9^0 10^' 0''$ north latitude and $38^0 40^' 0''$ to $38^0 50^' 0''$ east longitude with an altitude of 2379 meters above sea level.

3.2. Source of Data and Study Population

All CRC patients in the TASH Oncology unit were used as a source of data for this study which was obtained from the TASH oncology unit, Addis Ababa of Ethiopia and the population of the study was all medical records of CRC patients in the TASH Oncology unit who were diagnosed January 1, 2017, up to December 30, 2020, who fulfill eligibility criteria.

3.3. Study Design

A hospital-based Retrospective cohort study was applied to obtain data on CRC patients that was recorded in the oncology department of TASH, Addis Ababa, Ethiopia.

3.4. Study Variables

In statistical models, there were two types of variables called response variables and explanatory variables. The response variable of this study was the time to death of CRC patients, which was measured in months. Several predictors were considered in this study to investigate the associated factors of time to death of CRC patients. These covariates are described together with their coding scheme in Table3.1.

13

Table3.1. Description and Coding of the predictor Variables

No	Variable	Description	Categories and codes
Socio-demographic factors			
1	Sex	Sex of Patients	1= Male , 2=Female
2	Age	Age of patients	1=< 40 , 2=40-49, 3=50-59 , 4=>=60
3	Family history	Family history of CRC	1=Yes , 2= No
4	Alcohol consumption	Alcohol consumption	1=Yes, 2=No
5	Residence	Residence of patients	1=Urban, 2=Rural
6	Marital status	Marital status of patients	1= Single ,2=Married
7	Smoking status	Smoking status of patents	1= Smoker , 2= Non- Smoker
8	BMI	BMI of patients	1=Under weight, 2=Healthy weight, 3=Over weight
9	Comorbidity	Comorbidity illness of patients	1=Yes 2=No
10	physical exercise	physical exercise	1=Yes, 2=No
11	Religion	Religion of patients	1=Orthodox, 2=Muslim, 3=Protestant,
Clinical and pathological related factors			
12	Site of tumor	The primary site of a tumor	1=Rolon , 2= Rectal
13	Stage	Stage of the disease	1= Stage I and II ,2=Stage III, 3=Stage IV
14	Tumor grade	Tumor grade	1=Well differentiated, 2=Moderately differentiated,

			3=Poorly differentiated
15	Histology type	Histology type	1=Adenocarcinoma,
			2=Mucinous /signet ring -cell carcinoma
Treatment-related factors			
16	Treatment	Treatment modality	1=Radiotherapy alone
			2=Surgical treatment alone ,
			3=Chemotherapy alone,
			4=Surgery plus chemo,
			5=Surgery +chemo+ radiation

3.4. Data Extraction and Procedures

All medical records of confirmed colorectal cancer patients in TASH during the defined period (2017–2020) were included while incomplete patient charts, `charts that were missed at the time of data collection, Patients who had confirmed diagnosis at other hospital and were referred to TASH for advanced management were excluded from the study.

The information available in the eligible patients' medical records was observed and then recorded using a data extraction tool prepared by adapting from different studies (Magaji BA et al., 2017; Walter V et al., 2016,) which consists of socio-demographic factors, clinical and pathological related factors, and treatment-related factors. Then, all charts of colorectal cancer patients, diagnosed between 1st January 2017 to 30th December 2020 in TASH were retrieved and then reviewed (both baseline and follow-up records), death certificate supplemented was identified from TASH cancer registries by their medical record number. Then, the records of all the study participants were selected according to the eligibility criteria.

3.5. Methods of Data Analysis

3.5.1. Survival Data Analysis

Survival analysis is a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. It was used in analyzing the time-to-even data arising in several applied fields like medicine, biology, public health, epidemiology, demography (Aalen et al., 2008). Censoring is an important issue in survival analysis, it present when we have some information about a subject's event time, but we don't know the exact event time. There are three categories of censoring such as right censoring, left censoring, and interval censoring (Klein and Moeschberger, 2003). The presence of the patient in the data set who have not yet experienced a failure by the end of the study period.

3.5.2. Descriptive Statistics

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group.in summarizing survival data, the two common functions applied are the survival function and the hazard function (Hosmer and Lemeshow, 1999).

3.5.2.1. Survival function

The survival 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 survived beyond a specified time. Moreover, the distribution of survival time is characterized by three functions: (a) the survivorship function, (b) the probability density function, and(c) 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 survival function, $S(t)$ is

$$S(t) = p(T > t) = 1 - F(t), t \geq 0$$

Where, $F(t)$ is cumulative distribution function, which represents the probability that a subject selected at random would have a survival time less than or equal to some sated value t, is

$$F(t) = P(T \le t) = \int_0^t f(u)du, \, t \ge 0$$

The probability density function, $f(t)$, is given by:

$$f(t)\frac{d}{dt}F(t) = -\frac{d}{dt}S(t) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots (3.1)$$

3.5.2.2. Hazard Function

The hazard function is a measure of the risk of the event happening at any point of time. It is the instantaneous probability of having an event at time t (per unit time) given that one has survival (i.e not had an event) up to time t(Kleinbaum and Klein, 2011). It is given by:

$$\lambda(t) = \frac{f(t)}{S(t)}, \, F(t) = -\frac{d}{dt}\ln S(t)$$

The cumulative hazard function is given by:

$$\Lambda(t) = \int_0^t \lambda(u)\,du = \ln S(t)$$

Thus;

$$S(t) = e^{-\Lambda(t)} \dots\dots\dots\dots\dots\dots\dots\dots\dots..(3.2)$$

3.6. Non –Parametric Estimation of Survivorship Function

In practice, when using actual data, we usually obtain the estimated survivor function and obtain curves that are step functions, rather than smooth curves.

3.6.1. Kaplan-Meier Estimate of the Survival Function

The number of observed events at t (j), j = 1… r. Then the K-M estimator of S (t) is defined as the Kaplan-Meier estimator is the standard non-parametric estimator of the survival function used for estimating the survival probabilities from observed survival times both censored and uncensored (Kaplan and Meier, 1958).

Suppose that r individuals have failures in a group of individuals, let $0 \le t_{(1)} \le \dots < t_{(r)} < \infty$ be the observed ordered death times. Let $r_{(j)}$ be the size of the risk set $t_{(j)}$ where risk set denotes the collection of indivi0duals alive and uncensored just before $t_{(j)}$. Let $d_{(j)}$ be the number of observed events at $t_{(j)}$ $j=1\dots r$. Then the K-M estimator of (t) is defined by:

$$\hat{S}(t) = \prod_{j:t(j)<} \left[1 - \frac{d_{(j)}}{r(j)} \right]$$

The cumulative hazard function of the KM estimator can be estimated as:

$$\hat{H}_{KM}(t) = -\ln(\hat{S}_{KM}(t) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3.3)$$

Where $\hat{S}(t)$ is KM estimator

3.7. Non-Parametric Comparison of Survival Functions

The Kaplan –Meier pots were used to see whether there was a difference in survival time or not between groups of covariates under the investigation. But, the KM plot cannot be used to decide whether the survival time of patients living with CRC in each covariate were real differences or significant differences or not. Instead, used log-rank test.

3.7.1. Log-rank test

The log-rank test, developed by Mantel and Haenszel, was a non-parametric test for comparing two or more independent survival curves. Since it was a nonparametric test no assumption about the distributional form of the data was required. This test was most power full in detecting a higher cured proportion in one group than other groups (Mantel and Haenszel, 1959). The log-rank test statistic for comparing two groups was given by:

$$\chi^2{}_{logrank} = \frac{[\sum_{j=1}^{r}(d_{1j}-n_{1j}\frac{d_j}{n_j})]^2}{\sum_{j=1}^{r}\frac{n_{1j}n_{2j}d_j(n_j-d_j)}{n_j^2(n_j-1)}} \sim \chi^2{}_{(1)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3.4)$$

Where:

❖ r is the total number of rank-ordered event (death) times.

❖ d_{1j} is the numbers of failure in j^{th} time of 1st group

- ❖ d_{2j} is the numbers of failure in j^{th} tine of 2nd group

- ❖ d_j is the numbers of failure in j^{th} time $d_{1j} + d_{2j}$

- ❖ n_{1j} is the number at risk at j^{th} time of 1st group

- ❖ n_{2j} is the number at risk at j^{th} time of 2nd group

- ❖ n_j is the number at risk at j^{th} time $(n_{1j} + n_{2j})$

The log-rank test can also be extended for comparing three or more groups of survival experience.

3.8. Survival Models

3.8.1. Semi-Parametric Proportional Hazard Model

The Cox proportional hazard model was widely used for survival analysis because it is simple; it can easily accommodate right censoring. It was used to relate several risk factors or exposure considered simultaneously, to survival time. In a Cox –proportional hazard regression model, the measure of effect was hazard rate, which is the risk of failure (i.e the risk or the probability of suffering the event of interest), given that the participant has survived up to a specific time (Cox, 1972). Thus, the relationship between the predictors and the time to event in survival analysis was given thought hazard function as follows:

$$\lambda(t|Z) = \lambda_0(t)e^{\beta'Z} = \lambda_0(t)e^{\beta_1 Z_1 + \beta_2 Z_2 \cdots\cdots\cdots\cdots + \beta_p Z_p}$$

Were:

- ❖ $\lambda(t|z)$ was the hazard at a time t for patients with a set of predictors $Z_1, Z_2 \ldots\ldots\ldots\ldots.Z_P$
- ❖ $\lambda_0(t)$ was the baseline hazard function
- ❖ $\beta_1, \beta_2 \ldots\ldots\ldots\ldots.\beta_P$ Were the parameters describing the effect of the predictors on the overall hazard rate.

The corresponding survival function for the Cox PH model was given by:

$$S(t|Z) = [S_0(t)]^{e^{\beta_1 Z_1 + \beta_2 Z_2 \cdots\cdots\cdots + \beta_P Z_P}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (3.5)$$

Where:

$S_0(t)$ Was the baseline survival function and $S(t|Z)$ was survival function at time t for patients with a set of predictors $Z_1 + Z_2 \ldots \ldots \ldots \ldots \ldots + Z_P$.

3.8.1.1. Assumption of Cox proportional Hazard Model

1) The baseline hazard depends on t, but not on the covariates $x_1, x_2, \ldots, x_p$.

2) The hazard ratio, i.e.,($\beta'X$) depends on the covariates $X = (x_1, x_2, \ldots, x_p)$ but not on time t.

3) The covariates x_i are time-independent.

3.8.1.2. Partial Maximum Likelihood Estimation for Cox PH Model

The general methodology used for proportional hazards which cancels out the baseline function is also used in determining the partial likelihood. To illustrate, the partial likelihood of an event occurring at time t for an individual can be written as:

P (individual i has experienced an event at time t(i) one event at a time t_i

$$L = \frac{h(t, x_{i)}}{\sum_{j \in R_t}(i) \ h(t, x_{i)}} = \frac{h_0(t) \exp^{(\beta' x_i)}}{\sum_{j \in R_t}(t) \exp^{(\beta' x_j)}}$$

It assumes that there are no tied values among the observed survival times. Suppose we have m distinct failure times and let x_i is the vector of covariates at ordered failure time t(i). We define the Partial Likelihood as:

$$L_p(\beta) = \prod_{i=1}^{m} \left[\frac{\exp^{(\beta' x_i)}}{\sum_{j \in R_t}(i) \exp^{(\beta' x_j)}} \right] d_i \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (3.6)$$

Where d_i is the number of deaths, $d_i = 1$ we assume there are no tied so excluded for di=0. And,$Rt(i)$ is the set of subjects at risk at a time just before ti (ti-0). And the summation in the denominator is over all subjects in the risk set at time ti denoted by$Rt(i)$.

3.8.1.3. Checking the Assumption of Proportional Hazard

The main assumption of the Cox PH model is proportional hazards. Proportional hazard means that the hazard function of the individuals is proportional to the hazard function of the other

individuals; i.e the hazard ratio is constant over time. There are several methods for varying that a model satisfies the assumptions of proportionality.

3.8.1.3.1. Graphical Method

We can obtain the Cox PH survival function by the relationship between hazard function and survival function.

$$s(t, X) = s0(t)^{e^{\sum_{i=1}^{p} \beta i x i}}$$

Where x1,x2…….xp are Explanatory variables. When taking the logarithm twice we can easily gate

$$\ln(-lns(t, x)) = \sum_{i=1}^{p} \beta i x i + ln - lns0(t) \dots \dots \dots \dots \dots (3.7)$$

By plotting estimated $\ln(-lns(t, x))$ versus survival time for two groups, we will the parallel curves if the hazards are proportional. This method does not work well for continuous predictors or categorical predictors that have many levels. Looking at KM curves and $\ln(-lns(t, x))$ is not enough to be certain of proportionality since they are univariate analyses and do not shows whether hazards will still be proportional.

3.8.1.3.2. Tests Based on Schoenfeld Residuals

This overcomes the problem that other residuals depend heavily on observed survival time and cumulative hazard function. They are computed for each individual and covariate(Schoenfeld, 1982). It follows that, the Schoenfeld residual for the ith individual and k^{th} covariate is defined as:

$$\hat{S}_{ik} = \delta_i \left[X_{ik} - \frac{\sum_{j \in R(t_i)} X_{ik} exp^{(\beta' x_j)}}{\sum_{j \in R(t_i)} exp^{(\beta' x_j)}} \right] \dots \dots \dots \dots \dots \dots (3.8)$$

Where, X_j is a vector of p fixed covariates for the jth individual, X_{jk} is the value of kth covariate on the jth individual. Because of that, Schoenfeld residuals are defined only for the uncensored observations in which

$$\hat{S}_{ik} = X_{ik} - \frac{\sum_{j \in R(t_i)} X_{ik} exp^{(\beta'x_j)}}{\sum_{j \in R(t_i)} exp^{(\beta'x_j)}}$$ and for each covariate, it must sum to zero. In addition, they are uncorrelated and with an expected value of zero (Schoenfeld, 1982).

If the plot of the scaled Schoenfeld residuals with each continuous covariate versus time or log time is parallel, then the proportional hazard assumptions are fulfilled.

3.9. Parametric Accelerated Failure Time (AFT) Model

The AFT models were an alternative to the PH model for the analysis of survival time data when the Proportional hazard assumption doesn't hold. The key difference between the Cox-PH model and AFT models were the baseline hazard function and ways of estimating coefficient (Kleinbaum and Klein, 2011). The AFT was obtained by regressing the logarithm of the survival time over the covariates and the effect of the explanatory variables on the survival time is directly measured. Some of the standard parametric AFT models were exponential, Weibull, log-normal, and log-logistic (Datwyler and Stucki, 2011).

The survival function of an individual which covariate X at time t, in the accelerated failure time models is the same as the baseline survival function at time $t *$ $exp(\beta_1 X_{1i} + \beta_2 X_{2i} \dots \dots, +\beta_p X_{pi})$, where $\beta_1, \beta_2 \dots \dots \beta_p$ were coefficient of the regression models. Thus, the survival function of time t, $S(t|X) = S_0[t * exp(\beta_1 X_{1i} + \beta_2 X_{2i} \dots \dots, +\beta_p X_{pi})]$ for all $t \geq 0$. The effect of the covariates on the survival function is that the time scale was changed by a factor $exp(\beta'X)$, called an accelerated factor. The AFT model treats the logarithm of survival time as the response variable and includes an error term that is assumed to follow a particular distribution. The AFT model can be written as follow.

$$Log(T_i) = \mu + a_1 x_{1i} + a_2 x_{2i} + \cdots \dots \dots a_1 x_{\rho i} + \sigma \varepsilon_{i,} \dots\dots\dots\dots\dots(3.9)$$

This model shows the log-linear representation of the AFT model for the ith individual, where: μ is intercept, $log T_i$ is the log-transformed survival time, $X_1, X_2 \dots \dots \dots X_p$ were explanatory variables with coefficients $\beta_1, \beta_2 \dots\dots\dots \beta_p$, ε_i represent residual or unexplained variation in the log-transformed survival times, μ and σ were the intercept and scale parameters, respectively.

3.9.1. Exponential Distribution

The exponential distribution is the only distribution with a constant hazard i.e $\lambda(t,\lambda) = \lambda, \lambda > 0$.this implies that the conditional 'probability' of an event is constant over time. In other words, the risk of an event occurring is flat for time. The survivor function is $S(t,\lambda)=\exp\{-\lambda(t)\}$ and the density is $f(t,\lambda) = \lambda\exp(-\lambda t)$. It can be shown that $E(T) = \frac{1}{\lambda}$ and $Var(t) = \frac{1}{\lambda^2}$

3.9.2. Weibull Distribution

T is Weibull with parameter $\lambda > 0$ and $\rho > 0$, denoted $T\sim W(\lambda,\rho)$. The cumulative hazard is $\Lambda(t,\lambda,\rho,) = \lambda^\rho$, the survival function is $S(t,\lambda,\rho) = exp\{-\lambda t^\rho\}$, and hazard is $\lambda(t,\lambda,\rho)=\rho\lambda t^{\rho-1}$. The Weibull model is more general and flexible than the exponential model and allows for hazard rates that are non-constant but monotonic.

3.9.3. Log-logistic Distribution

A random variable T has the log-logistic distribution with the following hazard, density, and survivorship function $\lambda(t,\lambda,\rho)=\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^p}$, $S(t,\lambda,\rho) = \frac{1}{(1+\lambda t^\rho)}$, $and\ f(t,\lambda,\rho) = \frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^p)^2}$ where scale parameter $\lambda>0$, shape parameter $\rho > 0$.

3.9.4. Log normal distribution

The lognormal distribution is also defined for random variables that take positive values and so used as a model for survival data (Collett, 2015). If the survival times are assumed to have a lognormal distribution, the baseline survival function and hazard function respectively are given by.

$S_0(t)$=1- $\Phi\left(\frac{logt-\mu}{\sigma}\right)$ $and\ h_0(t)\frac{\phi\left(\frac{logt}{\sigma}\right)}{\left[1-\Phi\left(\frac{logt}{\sigma}\right)\right]\sigma t}$, Where $\mu\ and\ \sigma$ are parameters, (t) is the probability density function, and (t) is the cumulative density function. The survival function for the i^{th}individual is

$$S_i(t) = 1 - \Phi\left(\frac{logt-a'X_i-\mu}{\sigma}\right)\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.10)$$

3.9.5. Generalized gamma distribution

The probability density function of the generalized gamma distribution with three parameters, λ, α, and γ is defined by

$$f(t) = \frac{\alpha \lambda^{\alpha \gamma}}{\Gamma(\gamma)} t^{\alpha \gamma - 1} exp[-(\lambda t)^{\alpha}] \quad t > 0, \gamma > 0, \lambda > 0, \alpha > 0, \dots\dots\dots\dots\dots\dots(3.11)$$

Where γ is the shape parameter of the distribution. the survival function and the hazard function do not have a closed-form for the generalized gamma distribution. The exponential, Weibull, and log-normal models are all special cases of the generalized gamma model. it is easy to see that this generalized gamma distribution becomes the exponential distribution if $\alpha = \gamma = 1$. The Weibull distribution if $\gamma=1$, and the log-normal distribution if $\gamma \to \infty$.

3.10. Parameter Estimation for Parametric Accelerated Failure Time (AFT) Model

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

$$L(\partial, \mu, \delta) = \prod_{j=1}^{n} f_i(t_i)^{\delta_{i^*}} s_i(t_i)^{1-\delta_i}$$

Where $f_i(t_i)$ the density function of the tth individual at time t_i, S_i is the survival function of the tth individual at time t_i , δ_i is indicator variable. The logarithm of the above equation yields;

$$\log L(\partial, \mu, \delta) = \sum_{j=1}^{n}\{-\delta_i log \ (\delta t_i + \delta_i log f_i(x_i) + (1 - \delta_i) log S_i(W_i))\}\dots\dots\dots\dots\dots(3.10)$$

Where $W_j = \left\{log t_i - \frac{\mu + \alpha_{1i}\dots\dots + \alpha_{pi}x_{pi}}{\delta}\right\}, Z = \{z_{ji}\}$ is a vector of the covariate for the jth subject. The Maximum likelihood parameters estimates are found by using the Newton-Raphson procedure which can be done by software.

3.11. Method of Variable and Model Selection

3.11.1. Methods of Variable Selection

The methods of selecting a subset of covariates in Cox-PH and AFT models are essentially similar to those used in any other regression models. Hosmer and Lemeshow (1998) recommended the following steps in selecting the variables:

1. The first step is to fit a model that contains each of the variables one at a time.
2. We begin by fitting a multivariable model containing all variables significant in univariable analysis at the 20-25 percent level, as well as any other variables not selected with this criterion but judged to be of clinical importance.
3. Use backward selection to eliminate non-significant variables and examine the effect of the remaining variables.
4. Starting with step (3) model, consider each of the non-significant variables from step (2) using forward selection and do the analysis
5. Fit the final model by omitting variables that are non-significant and adding significant variables.

3.11.2. Method of Model Selection

There are some model selection criteria, such as Akaike Information Criterion (AIC)(Akaike,1974) and Bayesian information criterion(BIC)(Schwarz,1978). But most of the time for comparing models that are not nested, the Akaike Information Criterion(AIC) is used which is defined as:

$$AIC= -2LogL+2p \ldots\ldots\ldots\ldots\ldots\ldots (3.11)$$
$$BIC=-2LogL+Pln(n)\ldots\ldots\ldots\ldots\ldots.(3.12)$$

Where Log L is denoted the fitted log-likelihood, P is the number of parameters and n is the sample size. a model with lower AIC and BIC is preferred over one with higher AIC & BIC.

3.12. Model Diagnostic

After a model has been fitted, the adequacy of the fitted model needs to be assessed. The model checking procedure could be done using residuals. In linear; regression methods, residuals are

defined as the difference between the observed and predicted values of the dependent variables. However, when censored observations are present and partial likelihood function is used in the Cox PH model, the usual concept of residual is not applicable. Several residual has been proposed for the use of model diagnostics for the Cox PH and AFT model. in this study, we used Cox-Snell residual for model diagnostics(Cox .,1968, Therneau et al.,1990, Schoenfeld .,1982).

4. STATISTICAL DATA ANALYSIS

In this section we discussed, summary statistics of the covariate, estimate the survival time, and also comparing the survival curve in different groups of variables, fitting the model, and finally, the result was interpreted.

4.1. Socio-Demographic Characteristics of the Study Participants

Out of the 325 study participants, 214(65.85%) were censored and 111(34.15%) died. About 203(62.46%) of study participants were males and 60.31% were from urban areas. About 122 (37.54%) of patients were female whereas 203 (62.46%) of patients were male. Out of those 116(35.69%), participants had a family history and the remaining 209 (64.31%) were not having a family history.

As we have been considered the age of patients, 107(32.92%), 64(19.9%), 80 (24.62%), 74(22.77%) of patients were less than 40, 40 to 49, 50 to 59, and the age of 60 and above years old respectively. When we have been considered the smoking status of participants, we observed that 96(29.54%) of the participants were smokers, of which 33.33% have died and 59(27.57 %) were censored. Slightly more than half of (52.31%) the participants were in the normal condition (healthy weight) of BMI, out of which 123(57.48) patients were censored and 48(43.24) died. More than one quarter (29.5%) had comorbid conditions, of which 36.94% have died. About 129 (39.69%) patients were alcohol consumers, out of those patients 72(64.86%) were died.

Table 4.1:- socio-demographic characteristics of colorectal cancer patients in TASH

Variable	Category	Survival status		Total No.%
		Death No. (%)	Censored No. (%)	
Sex	Male	68(61.26)	135(63.08)	203 (62.46)
	Female	43(38.76)	79(36.92)	122(37.54)

	< 40	36(32.43)	71(33.18)	107(32.92)
	40-49	23(20.72)	41(19.16)	64(19.69)
Age of patients	50-59	28(25.23)	52(24.30)	80(24.62)
	>=60	24(21.62)	50 (23.36)	74(22.77)
Alcohol consumption	Yes	72(64.86)	57 (26.64)	129 (39.69)
	No	39(35.14)	157(73.36)	196 (60.31)
Family history	Yes	30(27.03)	86(40.19)	116(35.69)
	No	128 (59.81)	128 (59.81)	209 (64.31)
Smoking status	Smoking	37(33.33)	59(27.57)	96(29.54)
	Non Smoking	74(66.67)	155(72.43)	229(70.46)
Residence	Urban	69(62.16)	127(59.35)	196(60.31)
	Rural	42(37.84)	87(40.65)	129(39.69)
Marital status	Single	20(18.02)	28(13.08)	48(14.77)
	Married	91(81.98)	186(86.92)	277(85.23)
BMI	Under weight	47(42.34)	62(28.97)	110(33.85)
	Healthy weight	48(43.24)	123(57.48)	170(52.31)
	Over weight	16(14.41)	29 (13.55)	45(13.85)
Comorbidity	Yes	41(36.94)	56(26.17)	97(29.85)
	No	70(63.06)	158(73.83)	228(70.15)
physical exercise	Yes	11(9.91)	67(31.31)	78(24.00)
	No	100(90.09)	147 (68.69)	247(76.00)
Religious	Orthodox	72(64.86)	125(58.41)	197(60.62)
	Muslim	22(19.82)	45(21.03)	67(20.62)
	Protestant	17(15.32)	44(20.56)	61(18.77)

TASH: Tikur Anbesa Specialized Hospital.

4.2. Clinical and pathological and treatment-related characteristics

More than half (52%) of the primary site of the tumor was found to be rectal. Of those patients, 48.65% died. A large percentage (83.08%) of the patients were diagnosed at late stages (28.00% at stage III, and 55.08 % at stage IV). Slightly more than three-fifth (61.26%) of the patients that had been diagnosed at stage IV died. 61.85% of the tumor grade was well-differentiated; about 229 (70.46%) were adenocarcinoma type. Concerning the type of treatment given, 115(35.38%) of the cases were given chemotherapy alone, of which 41(36.94%) have died and 27.38% of the cases were served surgery plus chemotherapy out of those patients 42(37.84%) were died.

Table 4.2:- Clinical and pathological and treatment-related factors of colorectal cancer patients in TASH

Variable	Category	Survival status		Total No. (%)
		Death No. (%)	Censored No. (%)	
primary site of tumor	Colon	57(51.35)	99(46.26)	156(48.00)
	Rectal	54(48.65)	115(53.74)	169(52.00)
Stage	Stage I and II	10(9.01)	45 (21.03)	55(16.92)
	Stage III	33(29.73)	58(27.10)	91(28.00)
	Stage IV	68 (61.26)	111 (51.87)	179(55.08)
Tumor grade	Well-differentiated	62(55.86)	139(64.95)	201(61.85)
	Moderately differentiated	17(15.32)	46(21.50)	63(19.38)
	Poorly differentiated	32(28.83)	29(13.55)	61(18.77)
Histology type	Adenocarcinoma	74(66.67)	155(72.43)	229(70.46)
	Mucinous/signet ring cell carcinoma	59(27.57)	37(33.33)	96(29.54)

Treatment modality	Radiotherapy alone	11(9.91)	37 (17.29)	48 (14.77)
	Surgical treatment alone	2(1.80)	34 (15.89)	36 (11.08)
	chemotherapy alone	41(36.94)	74(34.58)	115(35.38)
	surgery plus chemo	42(37.84)	47(21.96)	89(27.38)
	radiation + surgery+ chemo	15(13.51)	22(10.28)	37(11.38)

TASH: Tikur Anbesa Specialized Hospital.

4.3. Non-parametric Survival Analysis

4.3.1. The Kaplan- Meier Survival Estimate for Time to Death of Colorectal Cancer Patients in TASH
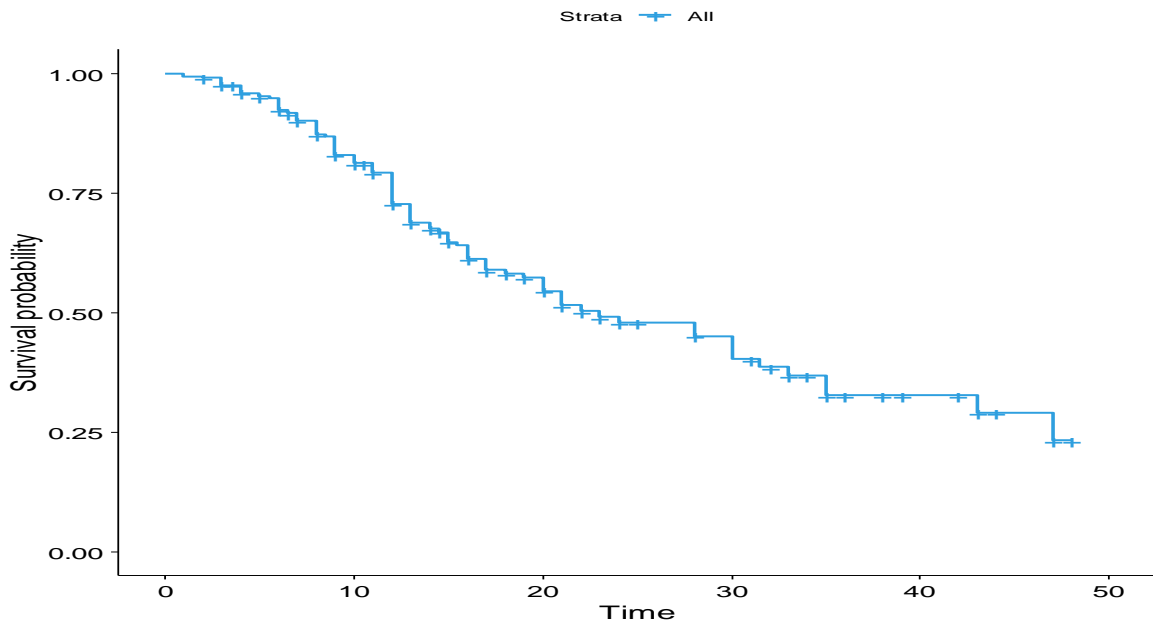
Figure 4.1: Overall Kaplan-Meier estimation of survival functions of colorectal cancer patients

From Figure 4.1, we have observed that the probability of survival was highest at the first month of diagnosis of colorectal cancer, but it relatively declined later as follow-up time increased.

4.3.2. The Overall Median Survival Time of Colorectal Cancer Patients

Table 4.3 showed that the overall median survival time of colorectal cancer patients was 23 months with [95%CI: 20–33].

Table 4.3:- Estimation of overall median survival time of colorectal cancer patients

No of observation	No of events	Median survival time in a month	95% CI	
			Lower	Upper
325	111	23.0	20.0	33
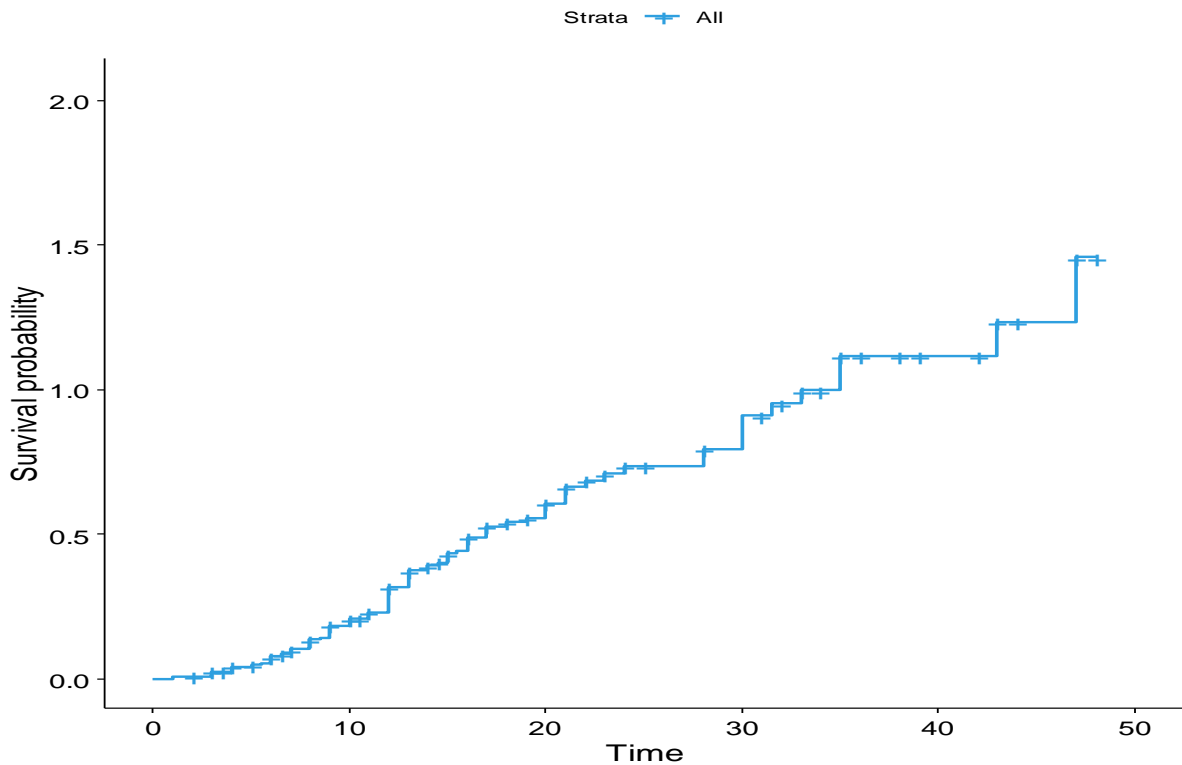
Figure 4.2: Overall Kaplan-Meier estimation of hazard functions of colorectal cancer patients

From the above Figure 4. 2 we observed that the probability of hazard is the lowest at the first month of diagnosis of colorectal cancer, but it relatively highest as follow up time increases.

4.3.3. Survival Function of Different Categorical variables
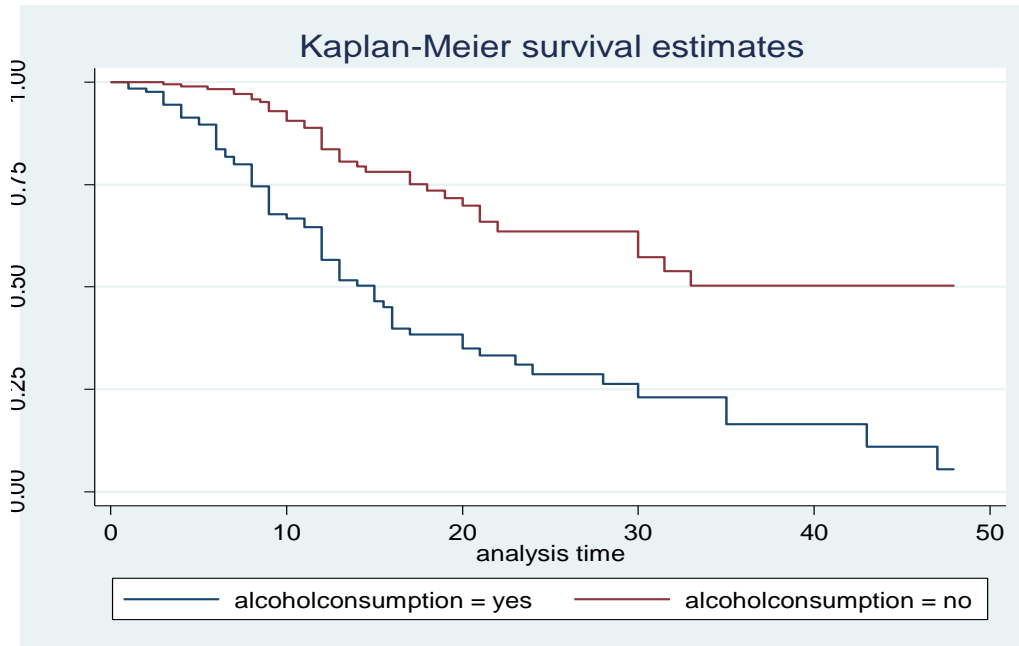
Figure 4.3: Plot of Kaplan-Meier Estimates for alcohol consumption

The survival curves in Figure 4.3 showed that non-alcohol user patients lying above as compared to alcohol user patients. It indicated that non-alcohol user patients had a higher probability of survival than alcohol user patients.
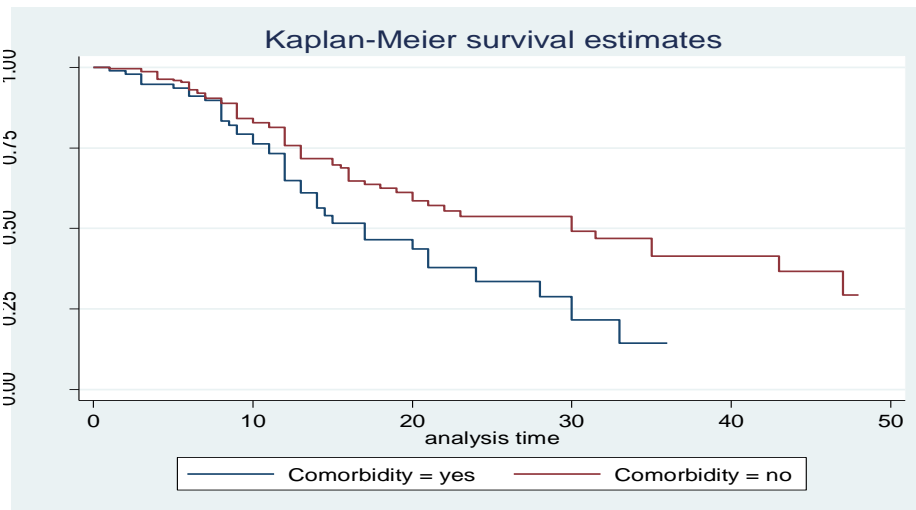
Figure4.4: Plot of Kaplan-Meier Estimates for comorbidity

Figure 4.4 revealed that the survival of the patients having non-comorbid conditions had better survived as compared to the survival of the patients that had a comorbid condition. It showed that comorbid condition patients had a lower probability of survival than non- comorbidity illness patients.
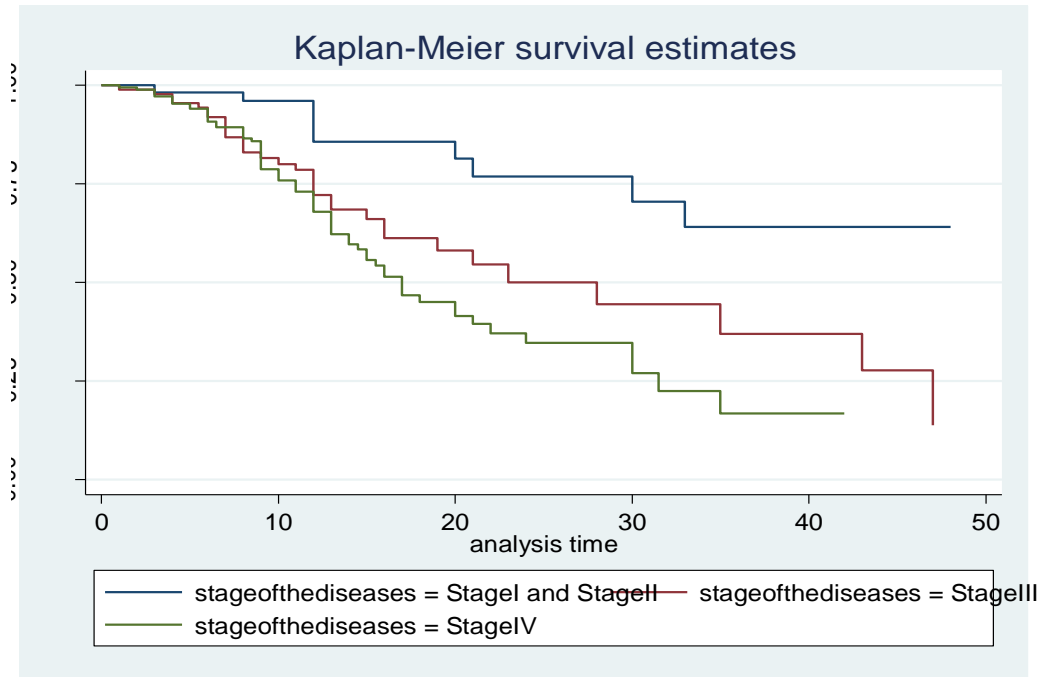
Figure 4.5: Plot of Kaplan-Meier Estimates for the stage

The survival function against survival time for time to death of colorectal cancer patients by stage of the diseases was shown in Figure 4.5. This plot indicated that patients who were diagnosed at stages I and II have a higher probability of survival than stage III and stage IV and patients who were diagnosed at stage III had a higher probability of survival than stage IV.

4.3.4. Comparison of Survival Experiences between Groups

By using log-rank tests there were significant differences in survival experience among groups of family history, alcohol consumption, comorbidity, physical exercise, stage, tumor grade, and treatment modality shown in Table 4.4. These showed that those categorical variables had statistically significant differences in survival probabilities.

Table 4.4:- Results of log-rank test for each categorical variable

Covariates /factors	Log-rank test		
	DF	Test statistics	p- value
Age	3	1.39	0.7072
Sex	1	0.1	0.8
Family history	1	7	0.008*
Alcohol consumption	1	42	<0.0001***
Residence	1	0.6	0.4
Marital status	1	2.13	0.1442
Smoking status	1	1.7	0.2
BMI	2	2.95	0.2290
Comorbidity	1	7.5	0.006*
Physical exercise	1	13.7	< 0.0001***
Religion	2	1.3	0.2
Primary site of tumor	1	0.001	0.9
Stage	2	18.19	0.00011**
Tumor grade	2	11.3	0.004*
Histology type	1	0.75	0.3866
Treatment modality	4	17.51	0.0015*

4.4. Cox proportional hazards model

The fitted Cox-PH model, as shown from (Appendix A Table 4.8) from that the time to death of CRC patients significantly affected by family history, alcohol consumption, marital status, physical exercise, tumor grade, stage of the disease, and treatment modality.

Test of proportional hazard assumption by graphical method
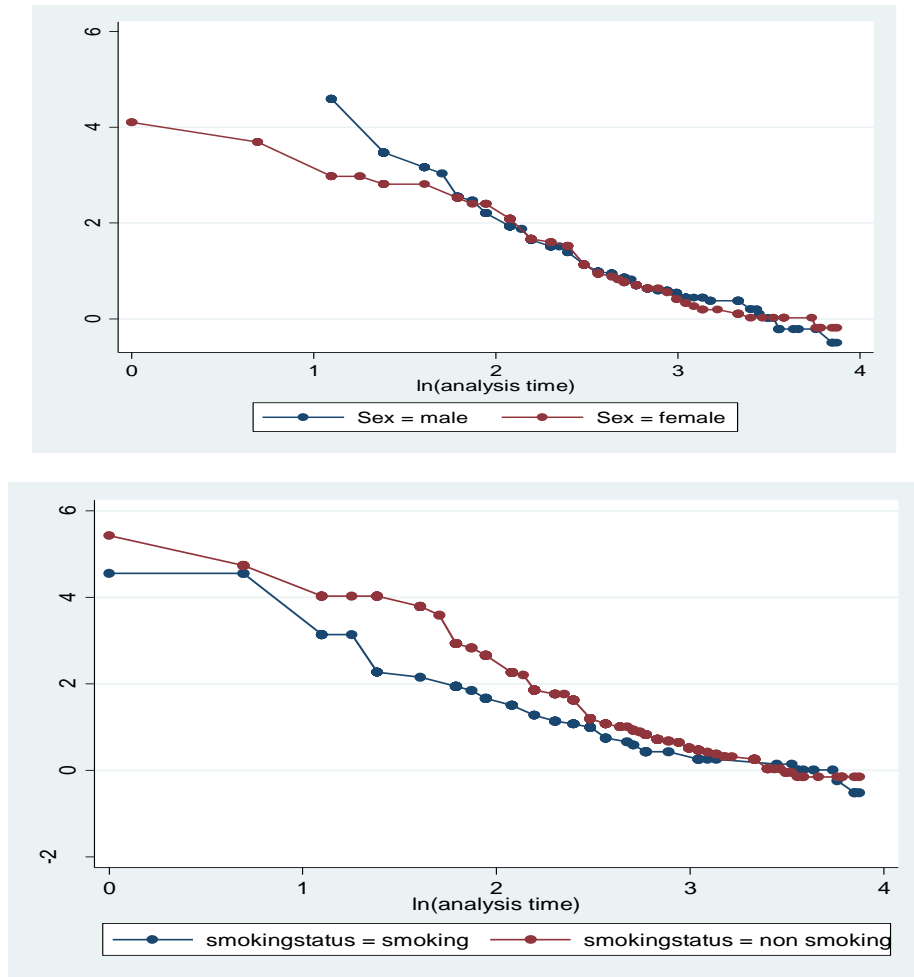
Figure 4.6: check proportional hazard assumption by graphical method

In the above plot the estimated $\ln(-lns(t,x))$ versus survival time for two groups would not parallel curves so proportional hazard assumption has been violated.

Test of proportional hazard assumption by Schoenfeld residual

In this study the p-value was checked for testing the assumption is fulfilled or not. The global test result shows (Appendix A Table 4.9) that the p-value is significant (p-value=0.0259<0.05). This revealed that there is evidence to contradict the proportionality assumption. So the proportionality assumption is not fulfilled.

4.5. Accelerated Failure Time Model

We used univariable analysis before proceeding to the multivariable analysis. The univariable analysis was fitted for each covariate with a p-value less than 0.25 by using different AFT models such as Exponential, Weibull, log-logistic, generalized gamma, and log-normal distribution. Family history, alcohol consumption, marital status, BMI, comorbidity, smoking status, physical exercise, stage, tumor grade, and treatment modality were found to be significant with time to death of CRC patients at 25% levels of significance in all AFT models.

Multivariable analysis of Exponential, Weibull, log-normal, generalized gamma, and log-logistic models was done by using all significant covariates in univariable analysis at 5% levels of significance. We used the backward elimination method to select the final significant covariates /factors.

The covariates such as alcohol consumption, marital status, physical exercise, stage of the diseases, tumor grade, and treatment modality were significant at a 5% level of significance in all AFT models. Whereas the covariate such as Family history was significant in Weibull, log-logistic, generalized gamma, and lognormal AFT models. Model comparison was done using those significant covariates for each AFT model.

4.5.1. Model Selection

The value of AIC and BIC for all AFT models are displayed in Table 4.5. The AIC and BIC value for log-logistic was the smallest compared to others. This revealed that the log-logistic AFT model better fits the colorectal cancer patient's data set.

Table 4.5:- comparisons of AFT models using AIC and BIC

Distributions	AIC	BIC
Exponential	455.06	504.25
Generalized gamma	410.96	463.94
Weibull	408.83	461.80
Lognormal	407.99	460.96
Log logistic	**406.43**	**459.41**

AIC= (Akaike's Information Criteria), BIC= (Biasian Information Criteria)

4.5.2. Log-logistic Accelerated Failure Time Model

The estimated value of regression coefficient for log-logistic AFT model using multivariable analysis with backward elimination is shown in Table 4.6. family history, alcohol consumption, physical exercise, stage, tumor grade, marital status, and treatment modality were significant covariates/factors.

The acceleration factor and its 95%CI for those CRC patients who had non-alcohol users compared with alcohol user patients were ϕ =1.907and [1.529, 2.435] respectively. This revealed that the non-alcohol user CRC patients survival time was longer by a factor$\hat{}$= 1.907 than alcohol user patients.

The acceleration factor for patients who had married when compared to single was 1.477[95% CI: 1.087, 2.007]. This implies that the survival time of married CRC patients increases by a factor$\hat{}$= 1.477 compared with patients who had single marital status at a 5% level of significance.

The estimated acceleration factor and 95% CI of acceleration factor for those who did not do physical activity compared to patients who did do physical activity were 0.475 and [0.333,0.676] respectively. Thus, the estimated survival times of CRC patients decreased by 52.5% for patients who did not do the physical activity than patients who did do physical activity.

With regarding the stage of the disease, the estimated acceleration factor and 95% CI of acceleration factor for the stage of the death of CRC patients who were in stage IV was 0.607and

[0.413, 0.892]. This indicates that CRC patients with stage IV their survival time (death time) was decreased by 39.3% compared with patients who were in stages I and II.

The acceleration factor and its 95% CI for CRC patients diagnosed as poorly differentiated tumor grade were ϕ =0.748 and [0.567, 0.986] respectively. This implies that patients diagnosed with poorly differentiated tumor grades shortened the survival time by 25.2% compared with well-differentiated tumor grades.

CRC patients who had no family history of colorectal cancer decrease survival time by a factor$\hat{}$= 0.743 than patients who had a family history of CRC.

Concerning the type of treatment given, CRC patients diagnosed as chemotherapy alone and surgery+chemo reduced survival time by 34.4% and 36.6% as compared with CRC patients given Radiotherapy alone ϕ=0.634 [95% CI: 0.431,0.931], ϕ =0.656 [95% CI: 0.451, 0.957] respectively.

The value of the shape parameter $\hat{\rho}$=1.538, hence the value is greater than unity the hazard function is unimodal.

Table 4.6:- maximum likelihood parameter estimate of the log-logistic AFT model

Covariates	Categories	$\hat{\beta}$	SE	ϕ	p-value	95% CI (ϕ)
Family history	Yes (Ref)	-	-	-	-	-
	No	-0.297	0.133	0.743	0.026*	0.572- 0.965
Alcohol consumption	Yes (Ref)	-	-	-	-	-
	No	0.646	0.124	1.907	0.00***	1.529- 2.435
Marital status	Single (Ref)	-	-	-	-	-
	Married	0.390	0.156	1.477	0.013 *	1.087-2.007
Physical exercise	Yes (Ref)	-	-	-	-	-
	No	-0.745	0.180	0.475	0.00***	0.333- 0.676
Stage of the diseases	Stage I and II(Ref)					
	Stage III	-	-	-	-	-
	Stage IV	- 0.290	0.212	0.748	0.172	0.493-1.134
		-0.499	0.196	0.607	0.011*	0.413- 0.892

		$\hat\beta$	SE	ϕ	P-value	95% CI (ϕ)
Tumor grade	Well-differentiated(Ref)	-	-	-	-	-
	Moderately differentiated	-0.071	0.159	0.931	0.658	0.681-1.273
	Poorly differentiated	-0.291	0.141	0.748	0.039*	0.567-0.986
	Radiotherapy alone (Ref)	-	-	-	-	-
Treatment Modality	surgical treatment alone	0.571	0.404	1.770	0.158	0.801-3.912
	chemotherapy alone	-0.421	0.192	0.656	0.028*	0.451- 0.957
	surgery + chemo	-0.456	0.196	0.634	0.020*	0.431-0.931
	radiation +surgery+ chemo	-0.361	0.232	0.697	0.120	0.442-2.572
	Intercept	3.968	0.339	52.87	0.0000	27.221- 102.822
shape parameter		$\hat\rho$=1.538				

NB: * Significant (P-value < 0.05), **significant (p-value<0.01), *** significant (p<0.001),

The reference category marked as (Ref)

$\hat\beta$: Estimate coefficient, SE= standard error, ϕ = acceleration factor, 95% CI (ϕ): Confidence Interval for acceleration factor, ρ = shape parameter

4.6. Model Diagnostics for AFT model

4.6.1. The cox snell residual plot
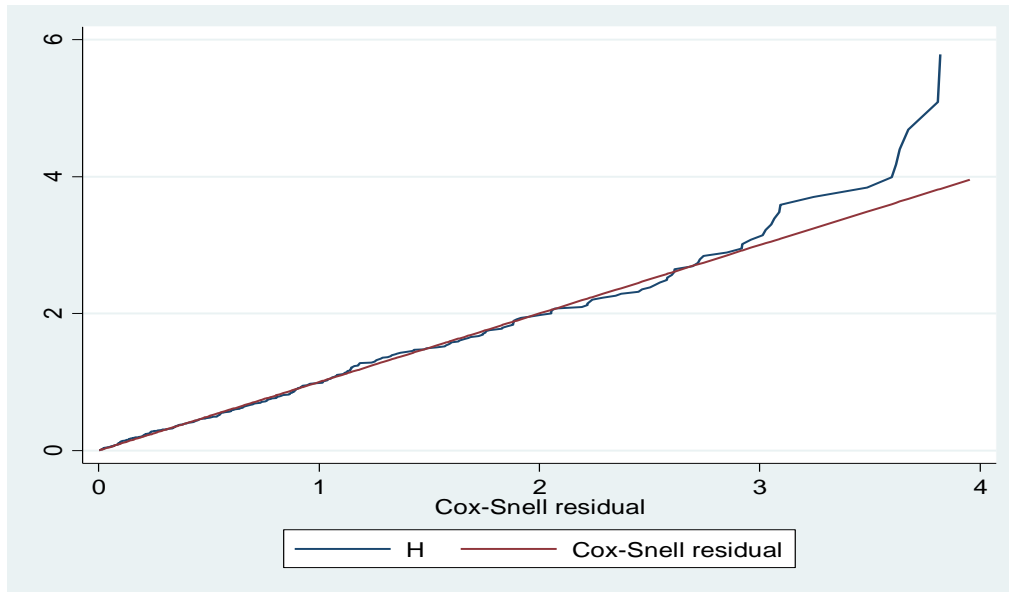

Figure 4.7: cox-snell residual Vs Kaplan-Meier cumulative hazard for a log-logistic model

The cumulative hazard plot of the cox-snell residual of the log-logistic AFT model is presented in Fig 4.7. the point lies on a line that has a unit slope, again the AIC and BIC of the log-logistic model were better. It indicates that the log-logistic AFT model fits the data very well.

5. DISCUSSIONS, CONCLUSIONS, AND RECOMMENDATIONS

5.1. Discussions

This study aimed to identify the factors significantly associated with the time to death of CRC patients in TASH enrolled from January 1, 2017, to December 30, 2020. Covariates that were included in the study were age, sex, Marital status, Family history, alcohol consumption, smoking status, BMI, comorbidity, physical exercise, residence, Religion, Site of a tumor, Stage of disease, Tumor grade, Histology type, and treatment modality.

In this study, out of the total 325 patients, 111(34.15%) died and 214(65.85%) were censored until the end of the study. Large percentages (83.08%) of the patients were diagnosed at late stages and more than three-fifth (61.26%) of the patients that had been diagnosed at stage IV died. This result almost similar to the result reported by (Atinafu et al., 2020), which implies that a large percentage (65.7%) of the patients were diagnosed at late stages (39.3% at stage III, and 26.4% at stage IV). Three-fifth (60.4%) of the patients that had diagnosed at stage IV died.

The overall Kaplan- Meier plot also showed that the follow-uptime of CRC patients increased, the survival probability of patients would be decreased. This was consistent with (Atinafu et al., 2020) and (Teka et al., 2021), which indicated that the probability of survival was the highest on the first day of diagnosis of colorectal cancer, but it relatively fell later as follow up time increases.

The Cox PH model is widely used for analyzing survival data in clinical research. If the assumption of the Cox PH model does not hold, there are various solutions to consider. AFT model is an alternative method for the analysis of survival data when hazards are not proportional.

The Cox PH model expresses the multiplicative effect of covariates on hazard the AFT model provides an estimate of the median survival time ratios. The results from AFT models are easier to interpret, more relevant to clinicians, and provide a more appropriate description of survival data in many situations. In this study, we analyzed the CRC data set using Cox PH and AFT models. For our data, the Cox PH model assumption was violated and AFT models provided a

better description of the data. So the AFT models with baseline distribution: exponential, Weibull, log-logistic, generalized gamma, and log-normal were considered. To compare different AFT models, AIC and BIC were used and the log-logistic AFT model was found to be the best fit for the time to death of CRC patients than others.

The findings of this study revealed that family history, marital status, physical exercise, alcohol consumption, stage, tumor grade, and treatment modality were significant factors for the time to death of CRC patients.

The result of our finding stage was found to be significantly associated with the time to death of CRC patients. This revealed that CRC patients who were diagnosed at stage IV had shortened the survival time by a factor $\phi = 0.607$ as compared to stage I and II. This finding was similar to the study conducted by (Teka et al., 2021) ;(Moamer et al., 2017 and (Etissa et al., 2021), which revealed that the risk of mortality for metastatic cancer was higher than non-metastatic patients.

In our thesis result, alcohol consumption was a significant effect on the time to death of CRC patients. This showed that, non-alcohol user CRC patients high chance of survival than alcohol user patients. This finding in line with (Atinafu et al., 2020), which indicated that CRC patients who alcohol users were 1.5 times at high hazard to die than non-alcohol users (CI: 1.07 -2.2). And our finding is also supported by (Walter et al., 2017), which implies that heavy drinking is also associated with poorer survival after a CRC diagnosis than light drinking. Mainly, lifetime heavy drinkers exhibited poorer overall (a HR: 1.37; 95% CI: 1.06, 1.78) than light drinking.

Family history (who had no family history) of CRC patients were reduced the survival time than patients who have a family history. This showed that those patients who had no family history of CRC reduced survival by 25.7% than patients who had a family history of CRC. This finding supported by a retrospective study conducted by (Morris et al. 2013), which indicated that patients with having family history had an 11% reduction in the risk of death compared to patients with no family history (HR=0.89,95% CI: 0.81-0.98, P=0.02).

From our study, we had seen that marital status was a significant predictor of the time to death of CRC patients. This revealed that, patients who had married marital status increase their survival by a factor $\phi =1.477$ as compared to single marital status. This result supported by (Alyabsi et al.,

2021), which showed that married patients with CRC have superior survival compared to unmarried (single) patients and that unmarried patients were 30% higher risk of death due to CRC compared to married patients (a HR 1.30; CI 1.17, 1.44).

A study conducted by (Atinafu et al., 2020) showed that patients diagnosed as undifferentiated tumor grade were 1.7 times at high hazard to die than those who were a well-differentiated type of tumor (AHR:1.7, CI:1.17-2.4). Our finding also showed that CRC patients diagnosed with poorly differentiated tumor grades shortened the survival time by 25.2% than those who were well-differentiated tumors. Moreover, another other study conducted by (Ahmadi et al., (2015) also showed that tumor grade is a significant predictor of CRC patient's death.

Physical activity is strongly associated with a reduced risk of CRC (Rasool et al., 2013). The result of this study also showed that physical inactivity was a significant effect on the time to death of CRC. i.e CRC patients who did not do physical activity shortened the survival by 52.5% compared to patients who did do physical activity.

The result of this study indicated that treatment modality was a significant effect on time to death. This implies that CRC patients diagnosed as chemotherapy alone and surgery plus chemo reduced survival time by34.4% and 36.6 % respectively as compared with CRC patients given Radiotherapy alone. This result was consistent with (Etissa et al., 2021).

5.2. Conclusions

This study was based on colorectal cancer (CRC) patients data set which was obtained from the TASH oncology unit in Addis Ababa, Ethiopia. The major objective of the study was to investigate the associated factors that affect the time to death of CRC patients in TASH enrolled from January 1, 2017, to December 30, 2020. Based on the descriptive result out of the total 325 patients 111(34.15%) died and 214(65.85%) were censored.

The Kaplan-Meier curve plot of CRC patients shows as the follow-up time increases the survival probability of patients decreases. The estimated median survival time of colorectal cancer patients was 23 months.

A log-logistic model was the best-fitted model for CRC patient's dataset. The result of the log-logistic model revealed that alcohol consumption, physical exercises, family history, marital status, stage, tumor grade, and treatment modality were significant prognostic factors.

Among these non-alcohol users and married patients were longer survival While patients did not do physical activity, patients have diagnosed only chemotherapy and surgery plus chemo, CRC patients who were diagnosed in stage IV, patients diagnosed as poorly differentiated tumor grade and individuals have no family history were shortened the survival time.

5.3. Recommendations

The gov't of Ethiopia has given high priority to CRC patient's survival interventions. This decision has been taken in contexts which strongly support such actions. Moreover, identify the important socio-economic, demographic, and other factors that affect the time to death of CRC patients and acting on them is mandatory. Based on our findings, we make the following recommendations

- ❖ Better if implement colorectal cancer early screening and detection program to improve treatment results and survival outcomes.
- ❖ The health care providers could enhance the awareness of treatment adherence.
- ❖ Could give special attention to the patient with poorly differentiated tumor.

REFERENCES

AALEN, O., BORGAN, O. & GJESSING, H. 2008. *Survival and event history analysis: a process point of view*, Springer Science & Business Media.

AHMADI, A., MOSAVI-JARRAHI, A. & POURHOSEINGHOLI, M. A. 2015. Mortality determinants in colorectal cancer patients at different grades: a prospective, cohort study in Iran. *Asian Pacific Journal of Cancer Prevention,* 16**,** 1069-1072.

AKAIKE, H. (1974). A new look at the statistical model identification. Ieee Trans. Automat. Contr. AC-19:716-23. [Institute of Statistical Mathematics, Minato-ku, Tokyo, Japan].

ALYABSI, M., RAMADAN, M., ALGARNI, M., ALSHAMMARI, K. & JAZIEH, A. R. 2021. The effect of marital status on stage at diagnosis and survival in Saudis diagnosed with colorectal cancer: cancer registry analysis. *Scientific reports,* 11**,** 1-10.

ARNOLD, M., SIERRA, M. S., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2017. Global patterns and trends in colorectal cancer incidence and mortality. *Gut,* 66**,** 683-691.

ANG, C., BAKER, M. & NICE, E. 2017. Mass spectrometry-based analysis for the discovery and validation of potential colorectal cancer stool biomarkers. *Methods in enzymology,* 586**,** 247-274.

ANSA, B. E., COUGHLIN, S. S., ALEMA-MENSAH, E. & SMITH, S. A. 2018. Evaluation of colorectal cancer incidence trends in the United States (2000–2014). *Journal of Clinical Medicine,* 7**,** 22.

ARVELO, F., SOJO, F. & COTTE, C. 2015. Biology of colorectal cancer. *Ecancermedicalscience,* 9.

ATINAFU, B. T., BULTI, F. A. & DEMELEW, T. M. 2020. Survival Status and Predictors of Mortality Among Colorectal Cancer Patients in Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: A Retrospective Followup Study. *Journal of cancer prevention,* 25**,** 38.

AIHW:ACIM (Australian Cancer Incidence and Mortality) Books.http://www.aihw.gov.au/cancer/data/acim_books/index.cfm, Accessed 30 August 2009.

BANEGAS, M. P., YABROFF, K. R., O'KEEFFE-ROSETTI, M. C., RITZWOLLER, D. P., FISHMAN, P. A., SALLOUM, R. G., LAFATA, J. E. & HORNBROOK, M. C. 2018.

Medical care costs associated with cancer in integrated delivery systems. *Journal of the National Comprehensive Cancer Network,* 16**,** 402-410.

BENNETT, S. 1983. Log-logistic regression models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 32**,** 165-171.

BEKAII-SAAB, T., KIM, R., KIM, T. W., O'CONNOR, J. M., STRICKLER, J. H., MALKA, D., SARTORE-BIANCHI, A., BI, F., YAMAGUCHI, K. & YOSHINO, T. 2019. Third- or later-line therapy for metastatic colorectal cancer: reviewing best practice. *Clinical colorectal cancer,* 18**,** e117-e129.

BOAKYE, D., RILLMANN, B., WALTER, V., JANSEN, L., HOFFMEISTER, M. & BRENNER, H. 2018. Impact of comorbidity and frailty on prognosis in colorectal cancer patients: a systematic review and meta-analysis. *Cancer treatment reviews,* 64**,** 30-39.

BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians,* 68**,** 394-424.

Cancer fact sheets: Colorectal cancer. WHO International Agency for Research on Cancer. 2016.

COTTON, P., EASTMAN, P. & LE, B. H. Palliative care and colorectal cancer. Cancer Forum, 2014. The Cancer Council Australia, 66.

COX, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society: Series B (Methodological),* 34**,** 187-202.

COLLETT, D. 2015. *Modeling survival data in medical research*, CRC press.

COLUSSI, D., BRANDI, G., BAZZOLI, F. & RICCIARDIELLO, L. 2013. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *International journal of molecular sciences,* 14**,** 16365-16385.

COX, D. & SNELL, E. 1968. General distribution of residuals (with discussion). *Journal of the Royal Statistical Society, Series B,* 30**,** 248-275.

DOLATKHAH, R., SOMI, M. H., KERMANI, I. A., GHOJAZADEH, M., JAFARABADI, M. A., FARASSATI, F. & DASTGIRI, S. 2015. Increased colorectal cancer incidence in Iran: a systematic review and meta-analysis. *BMC public health,* 15**,** 1-14.

EFRON, B. 1977. The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association,* 72**,** 557-565.

ETISSA, E. K., ASSEFA, M. & AYELE, B. T. 2021. Prognosis of colorectal cancer in Tikur Anbessa Specialized Hospital, the only oncology center in Ethiopia. *PLoS One,* 16**,** e0246424.

Federal Ministry of Health; Ethiopia: National Cancer Control Plan 2016–2020 October 2015: Addis Ababa Ethiopia (https://www.iccp-portal.org/ sites/default/files/plans/NCCP%20Ethiopia%20Final%20261015.pdf. Accessed 2020.

FERLAY, J., SOERJOMATARAM, I., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D. M., FORMAN, D. & BRAY, F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer,* 136**,** E359-E386.

FERGUSSON, D. M., HORWOOD, L. J. & SHANNON, F. T. 1984. A proportional hazards model of family breakdown. *Journal of Marriage and the Family***,** 539-549.

FIOROT, A., POZZA, A., RUFFOLO, C., CARATOZZOLO, E., BONARIOL, L., D'AMICO, F. E., PADOAN, L., CALIA DI PINTO, F., SCARPA, M. & CASTORO, C. 2018. Colorectal cancer in the young: a possible role for immune surveillance? *Acta Chirurgica Belgica,* 118**,** 7-14.

GANDOMANI, H. S., AGHAJANI, M., MOHAMMADIAN-HAFSHEJANI, A., TARAZOJ, A. A., POUYESH, V. & SALEHINIYA, H. 2017. Colorectal cancer in the world: incidence, mortality and risk factors. *Biomedical Research and Therapy,* 4**,** 1656-1675.

GALVIN, A., DELVA, F., HELMER, C., RAINFRAY, M., BELLERA, C., RONDEAU, V., SOUBEYRAN, P., COUREAU, G. & MATHOULIN-PELISSIER, S. 2018. Sociodemographic, socioeconomic, and clinical determinants of survival in patients with cancer: a systematic review of the literature focused on the elderly. *Journal of geriatric oncology,* 9**,** 6-14.

GINSBERG, G. M., LAUER, J. A., ZELLE, S., BAETEN, S. & BALTUSSEN, R. 2012. Cost effectiveness of strategies to combat breast, cervical, and colorectal cancer in sub-Saharan Africa and South East Asia: mathematical modelling study. *Bmj,* 344.

GIOVANNUCCI, E. & WILLETT, W. C. 1994. Dietary factors and risk of colon cancer. *Annals of medicine,* 26**,** 443-452.

GRAHAM, A., DAVIES ADELOYE, L. G., THEODORATOU, E. & CAMPBELL, H. 2012. Estimating the incidence of colorectal cancer in Sub–Saharan Africa: A systematic analysis. *Journal of global health,* 2.

GUPTA, R. C., AKMAN, O. & LVIN, S. 1999. A study of log-logistic model in survival analysis. *Biometrical Journal: Journal of Mathematical Methods in Biosciences,* 41**,** 431-443.

HAGGAR, F. A. & BOUSHEY, R. P. 2009. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery,* 22**,** 191.

HASSAN, M. R. A., SUAN, M. A. M., SOELAR, S. A., MOHAMMED, N. S., ISMAIL, I. & AHMAD, F. 2016. Survival analysis and prognostic factors for colorectal cancer patients in Malaysia. *Asian Pacific Journal of Cancer Prevention,* 17**,** 3575-3581.

HOSMER, D. W. & LEMESHOW, S. 1999. *Applied survival analysis: Time-to-event*, Wiley-Interscience.

HOSMER, D. W., LEMESHOW, S. & MAY, S. 2008. Model development. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data***,** 132-168.

IBRAHIM, K., ANJORIN, A., AFOLAYAN, A. & BADMOS, K. 2011. Morphology of colorectal carcinoma among Nigerians: a 30-year review. *Nigerian journal of clinical practice,* 14**,** 432-435.

IRABOR, D., AROWOLO, A. & AFOLABI, A. 2010. Colon and rectal cancer in Ibadan, Nigeria: an update. *Colorectal Disease,* 12**,** e43-e49.

JAPUNTICH, S. J., KUMAR, P., PENDERGAST, J. F., JUAREZ CABALLERO, G. Y., MALIN, J. L., WALLACE, R. B., CHRISCHILLES, E. A., KEATING, N. L. & PARK, E. R. 2019. Smoking status and survival among a national cohort of lung and colorectal cancer patients. *Nicotine & tobacco research,* 21**,** 497-504.

JONES, W. F., AHNEN, D. J. & SCHROY III, P. C. 2020. Improving on-time colorectal cancer screening through lead time messaging. *Cancer,* 126**,** 247-252.

KAPLAN, E. L. & MEIER, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association,* 53**,** 457-481.

KARLITZ, J. J., OLIPHANT, A.-L. B., GREENWALD, D. A. & POCHAPIN, M. B. 2017. The American College of Gastroenterology and the 80% by 2018 colorectal cancer initiative:

a multifaceted approach to maximize screening rates. *American Journal of Gastroenterology,* 112**,** 1360-1362 .

KATALAMBULA, L. K., NTWENYA, J. E., NGOMA, T., BUZA, J., MPOLYA, E., MTUMWA, A. H. & PETRUCKA, P. 2016. Pattern and distribution of colorectal cancer in Tanzania: a retrospective chart audit at two national hospitals. *Journal of cancer epidemiology,* 2016.

KATSIDZIRA, L., GANGAIDZO, I., THOMSON, S., RUSAKANIKO, S., MATENGA, J. & RAMESAR, R. 2017. The shifting epidemiology of colorectal cancer in sub-Saharan Africa. *The Lancet Gastroenterology & Hepatology,* 2**,** 377-383.

KEKELIDZE, M., D'ERRICO, L., PANSINI, M., TYNDALL, A. & HOHMANN, J. 2013. Colorectal cancer: current imaging methods and future perspectives for the diagnosis, staging and therapeutic response evaluation. *World journal of gastroenterology: WJG,* 19**,** 8502.

KIM, M. S., PARK, E. J., KANG, J., MIN, B. S., LEE, K. Y., KIM, N. K. & BAIK, S. H. 2018. Prognostic factors predicting survival in incurable stage IV colorectal cancer patients who underwent palliative primary tumor resection. Retrospective cohort study. *International Journal of Surgery,* 49**,** 10-15.

KLEINBAUM, D. G. & KLEIN, M. 2005. Competing risks survival analysis. *Survival Analysis: A self-learning text***,** 391-461.

KLEIN, J. P. & MOESCHBERGER, M. L. 2003. *Survival analysis: techniques for censored and truncated data*, Springer.

KLEINBAUM, D.G. & KLEIN, M. 2011. Survival analysis a self-learning text (3rd Ed.). Springer, New York.

KIM, J.-M., KIM, H.-M., JUNG, B.-Y., PARK, E.-C., CHO, W.-H. & LEE, S.-G. 2012. The association between cancer incidence and family income: analysis of Korean National Health Insurance cancer registration data. *Asian Pacific Journal of Cancer Prevention,* 13**,** 1371-1376.

LAWLESS, J. 1982. Parametric regression models. *Statistical Models and Methods for Lifetime Data,* 6**,** 269-339.

LEUFKENS, A. M., VAN DUIJNHOVEN, F. J., BOSHUIZEN, H. C., SIERSEMA, P. D., KUNST, A. E., MOUW, T., TJØNNELAND, A., OLSEN, A., OVERVAD, K. &

BOUTRON-RUAULT, M. C. 2012. Educational level and risk of colorectal cancer in EPIC with specific reference to tumor location. *International journal of cancer,* 130**,** 622-630.

LOWERY, J. T., AHNEN, D. J., SCHROY III, P. C., HAMPEL, H., BAXTER, N., BOLAND, C. R., BURT, R. W., BUTTERLY, L., DOERR, M. & DOROSHENK, M. 2016. Understanding the contribution of family history to colorectal cancer risk and its clinical implications: a state-of-the-science review. *Cancer,* 122**,** 2633-2645.

LYALL, M. S., DUNDAS, S. R., CURRAN, S. & MURRAY, G. I. 2006. Profiling markers of prognosis in colorectal cancer. *Clinical Cancer Research,* 12**,** 1184-1191.

MAGAJI, B. A., MOY, F. M., ROSLANI, A. C. & LAW, C. W. 2017. Survival rates and predictors of survival among colorectal cancer patients in a Malaysian tertiary hospital. *BMC cancer,* 17**,** 1-8.

MANTEL, N. & HAENSZEL, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute,* 22**,** 719-748

MCCORMACK, V. & NEWTON, R. 2019. Research priorities for social inequalities in cancer in sub-Saharan Africa. *150 cours Albert Thomas, 69372 Lyon Cedex 08, France© International Agency for Research on Cancer, 2019 Distributed by WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland***,** 319.

MEETOO, D. 2008. Chronic diseases: the silent global epidemic. *British journal of nursing,* 17**,** 1320-1325.

MEMIRIE, S. T., HABTEMARIAM, M. K., ASEFA, M., DERESSA, B. T., ABAYNEH, G., TSEGAYE, B., ABRAHA, M. W., ABABI, G., JEMAL, A. & REBBECK, T. R. 2018. Estimates of cancer incidence in Ethiopia in 2015 using population-based registry data. *Journal of global oncology,* 4**,** 1-11.

MERRILL, R. M., HARRIS, J. D. & MERRILL, J. G. 2013. Differences in incidence rates and early detection of cancer among non-Hispanic and Hispanic Whites in the United States. *Ethnicity & disease,* 23**,** 349-355.

MEYER, J. E., NARANG, T., SCHNOLL-SUSSMAN, F. H., POCHAPIN, M. B., CHRISTOS, P. J. & SHERR, D. L. 2010. Increasing incidence of rectal cancer in patients aged younger than 40 years: an analysis of the surveillance, epidemiology, and end results database. *Cancer,* 116**,** 4354-4359.

MOAMER, S., BAGHESTANI, A., POURHOSEINGHOLI, M. A., HAJIZADEH, N., AHMADI, F. & NOROUZINIA, M. 2017. Evaluation of prognostic factors effect on survival time in patients with colorectal cancer, based on Weibull Competing-Risks Model. *Gastroenterology and Hepatology from bed to bench,* 10**,** 54.

MORGAN, S. P., LYE, D. N. & CONDRAN, G. A. 1988. Sons, daughters, and the risk of marital disruption. *American journal of sociology,* 94**,** 110-129.

MORISHIMA, T., MATSUMOTO, Y., KOEDA, N., SHIMADA, H., MARUHAMA, T., MATSUKI, D., NAKATA, K., ITO, Y., TABUCHI, T. & MIYASHIRO, I. 2018. Impact of comorbidities on survival in gastric, colorectal, and lung cancer patients. *Journal of epidemiology***,** JE20170241.

MORRIS, E., PENEGAR, S., WHITEHOUSE, L., QUIRKE, P., FINAN, P., BISHOP, D., WILKINSON, J. & HOULSTON, R. 2013. A retrospective observational study of the relationship between family history and survival from colorectal cancer. *British journal of cancer,* 108**,** 1502-1507.

MUNRO, A., BROWN, M., NIBLOCK, P., STEELE, R. & CAREY, F. 2015. Do Multidisciplinary Team (MDT) processes influence survival in patients with colorectal cancer? A population-based experience. *BMC cancer,* 15**,** 1-9.

NASAIF, H. A. & AL QALLAF, S. M. 2018. Knowledge of colorectal cancer symptoms and risk factors in the Kingdom of Bahrain: A cross-sectional study. *Asian Pacific journal of cancer prevention: APJCP,* 19**,** 2299.

National Cancer Control Plan of Ethiopia 2015.

NIKSIC, M., RACHET, B., DUFFY, S. W., QUARESMA, M., MØLLER, H. & FORBES, L. J. 2016. Is cancer survival associated with cancer symptom awareness and barriers to seeking medical help in England? An ecological study. *British journal of cancer,* 115**,** 876-886.

O'CONNELL, J. B., MAGGARD, M. A. & KO, C. Y. 2004. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *Journal of the National Cancer Institute,* 96**,** 1420-1425.

OLSEN, M. 2015. *Cancer in Sub-Saharan Africa: The need for new paradigms in global health*, Boston University Frederick S. Pardee Center for the Study of the Longer ….

PARÉS-BADELL, O., BANQUÉ, M., MACIÀ, F., CASTELLS, X. & SALA, M. 2017. Impact of comorbidity on survival by tumour location: Breast, colorectal and lung cancer (2000–2014). *Cancer epidemiology,* 49**,** 66-74.

PARK, J. S., HUH, J. W., PARK, Y. A., CHO, Y. B., YUN, S. H., KIM, H. C., LEE, W. Y. & CHUN, H.-K. 2015. Prognostic comparison between mucinous and nonmucinous adenocarcinoma in colorectal cancer. *Medicine,* 94.

PETERSE, E. F., MEESTER, R. G., SIEGEL, R. L., CHEN, J. C., DWYER, A., AHNEN, D. J., SMITH, R. A., ZAUBER, A. G. & LANSDORP-VOGELAAR, I. 2018. The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: microsimulation analysis I to inform the American Cancer Society colorectal cancer screening guideline. *Cancer,* 124**,** 2964-2973.

PIKE, M. 1966. A method of analysis of a certain class of experiments in carcinogenesis. *Biometrics,* 22**,** 142-161.

POURHOSEINGHOLI, M. A. & ZALI, M. R. 2012. Colorectal cancer screening: Time for action in Iran. *World journal of gastrointestinal oncology,* 4**,** 82.

QUACH, C., SANOFF, H. K., WILLIAMS, G. R., LYONS, J. C. & REEVE, B. B. 2015. Impact of colorectal cancer diagnosis and treatment on health-related quality of life among older Americans: A population-based, case-control study. *Cancer,* 121**,** 943-950.

QURESHI, U. F., ASLAM, M. N., ANSARI, M. N. & KHAN, M. 2018. role of asprin as prophylaxis ageinst colorectal cancer. *Pakistan Postgraduate Medical Journal,* 29**,** 16-19.

RASOOL, S., KADLA, S. A., RASOOL, V. & GANAI, B. A. 2013. A comparative overview of general risk factors associated with the incidence of colorectal cancer. *Tumor Biology,* 34**,** 2469-2476.

RASOULI, M. A., MORADI, G., ROSHANI, D., NIKKHOO, B., GHADERI, E. & GHAYTASI, B. 2017. Prognostic factors and survival of colorectal cancer in Kurdistan province, Iran: A population-based study (2009–2014). *Medicine,* 96.

SCHOEN, R. E., RAZZAK, A., KELLY, J. Y., BERNDT, S. I., FIRL, K., RILEY, T. L. & PINSKY, P. F. 2015. Incidence and mortality of colorectal cancer in individuals with a family history of colorectal cancer. *Gastroenterology,* 149**,** 1438-1445. e1.

SECRETAN, B., STRAIF, K., BAAN, R., GROSSE, Y., EL GHISSASSI, F., BOUVARD, V., BENBRAHIM-TALLAA, L., GUHA, N., FREEMAN, C. & GALICHET, L. 2009. A

review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *The Lancet. Oncology,* 10**,** 1033-1034.

SEMNANI, S., NOORAFKAN, Z., ARYAIE, M., SEDAGHAT, S. M., MOGHADDAMI, A., KAZEMNEJHAD, V., KHORASANINEJHAD, R., GHASEMI-KEBRIA, F. & ROSHANDEL, G. 2016. Determinants of healthcare utilisation and predictors of outcome in colorectal cancer patients from Northern Iran. *European journal of cancer care,* 25**,** 318-323.

SHARKAS, G. F., ARQOUB, K. H., KHADER, Y. S., TARAWNEH, M. R., NIMRI, O. F., AL-ZAGHAL, M. J. & SUBIH, H. S. 2017. Colorectal cancer in Jordan: survival rate and its related factors. *Journal of oncology,* 2017.

SIEGEL, R. 2018. KD Miller & A. Jemal. 2018. Cancer statistics. *CA Cancer J. Clin,* 68**,** 7-30.

SOLOMON, S. & MULUGETA, W. 2019. Diagnosis and Risk Factors of Advanced Cancers in Ethiopia. *Journal of cancer prevention,* 24**,** 163.

SAIDI, H., ABDIHAKIN, M., NJIHIA, B., JUMBA, G., KIARIE, G., GITHAIGA, J. & ABINYA, N. 2011. Clinical outcomes of colorectal cancer in Kenya. *Annals of African Surgery,* 7.

SCHOENFELD, D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika,* 69**,** 239-241.

THERNEAU, T. M., GRAMBSCH, P. M. & FLEMING, T. R. 1990. Martingale-based residuals for survival models. *Biometrika,* 77**,** 147-160

TANNENBAUM, S. L., HERNANDEZ, M., ZHENG, D. D., SUSSMAN, D. A. & LEE, D. J. 2014. Individual-and neighborhood-level predictors of mortality in Florida colorectal cancer patients. *PLoS one,* 9**,** e106322.

TASHIRO, J., YAMAGUCHI, S., ISHII, T., SUZUKI, A., KONDO, H., MORITA, Y., HARA, K. & KOYAMA, I. 2014. Inferior oncological prognosis of surgery without oral chemotherapy for stage III colon cancer in clinical settings. *World journal of surgical oncology,* 12**,** 1-7.

TEKA, M. A., YESUF, A., HUSSIEN, F. M. & HASSEN, H. Y. 2021. Histological characteristics, survival pattern and prognostic determinants among colorectal cancer patients in Ethiopia: A retrospective cohort study. *Heliyon,* 7**,** e06366.

TIMOTEWOS, G., SOLOMON, A., MATHEWOS, A., ADDISSIE, A., BOGALE, S., WONDEMAGEGNEHU, T., AYNALEM, A., AYALNESH, B., DAGNECHEW, H. & BIREDA, W. 2018. First data from a population based cancer registry in Ethiopia. *Cancer epidemiology,* 53**,** 93-98.

VAN EEGHEN, E. E., BAKKER, S. D., VAN BOCHOVE, A. & LOFFELD, R. J. 2015. Impact of age and comorbidity on survival in colorectal cancer. *Journal of gastrointestinal oncology,* 6**,** 605.

WALTER, V., JANSEN, L., KNEBEL, P., CHANG-CLAUDE, J., HOFFMEISTER, M. & BRENNER, H. 2017. Physical activity and survival of colorectal cancer patients: Population-based study from Germany. *International journal of cancer,* 140**,** 1985-1997.

WEI, L.-J. 1992. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in medicine,* 11**,** 1871-1879.

WHO. 2018. Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018. *International Agency for Research on Cancer. Geneva: World Health Organization*.

WOLDU, M., LEGESE, D., ABAMECHA, F. & BERHA, A. 2017. The prevalence of cancer and its associated risk factors among patients visiting the oncology unit, Tikur Anbessa specialized hospital, Addis Ababa-Ethiopia. *J Cancer Sci Ther,* 9**,** 414-21.

WOLF, A. M., FONTHAM, E. T., CHURCH, T. R., FLOWERS, C. R., GUERRA, C. E., LAMONTE, S. J., ETZIONI, R., MCKENNA, M. T., OEFFINGER, K. C. & SHIH, Y. C. T. 2018. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA: a cancer journal for clinicians,* 68**,** 250-281.

APPENDICES A

Table 4.7: Results of the univarite analysis of Cox PH model

Covariates	Category	HR	Std.err	p- value	75% CI(HR)
Age	>40(Ref)	-	-	-	-
	40-49	0.930	0.249	0.788	0.6835- 1.266
	50-59	1.182	0.299	0.506	0.8843-1.582
	>=60	1.233	0.328	0.429	0.9087- 1.6752
Sex	Male (Ref)	-	-	-	-
	Female	1.012	0.197	0.952	0.8082- 1.266
Family history	Yes (Ref)	-	-	-	-
	No	1.648	0 .370	0.026*	1.060- 2.560
Alcohol	Yes(Ref)	-	-	-	-
consumption	No	0.312	0.062	0.000***	0.2482- 0.3924
Marital status	Single (Ref)	-	-	-	-
	Married	0.701	0.174	0.153	0.5267- 0.9327
Smoking status	Smoking(Ref)	-	-	-	-
	non-smoking	0.766	0.156	0.190	0.6064- 0.9679
Residence	urban (Ref)	-	-	-	-
	Rural	0.894	0.175	0.569	0.7141- 1.1206
BMI	Underweight (Ref)	-	-	-	-
	Healthy weight	0.706	0.146	0.092	0.5576- 0.8960
	Overweight	0.850	0.246	0.576	0.6093- 1.1867
Comorbidity	Yes(Ref)	-	-	-	-
	No	0.590	0.118	0.008	0.4692- 0.7419
Physical Exercise	Yes(Ref)	-	-	-	-
	No	3.201	1.018	0.000	2.2212- 4.6156
Religion	Orthodox (Ref)	-	-	-	-
	Muslim	0.896	0.218	0.655	0.6772- 1.1872
	Protestant	0.778	0.211	0.356	0.5701- 1.0633
Primary site	Colon (Ref)	-	-	-	-
	Rectal	1.004	0.191	0.985	0.8063- 1.2493

Stage	Stage I and II(Ref)	-	-	-	-
	Stage III	2.859	1.037	0.004	1.883- 4.339
	Stage IV	3.887	1.342	0.000	2.6134- 5.782
Tumor grade	Well-differentiated(Ref)	-	-	-	-
	Moderately differentiated	0.960	0.264	0.882	0.6998- 1.3169
	poorly differentiated	1.953	0.427	0.002	1.5188- 2.5120
Histology type	Adenocarcinoma (Ref)	-	-	-	-
	mucinous /signet ring _cell carcinoma	1.188	0.2400	0.394	0.9414- 1.4986
	Radiotherapy alone (Ref)	-	-	-	-
	surgical treatment alone	0.2131	0.164	0.044	0.0876- 0.5153
	chemotherapy alone	1.846	0.627	0.071	1.248- 2.7282
treatment modality	surgery plus chemo	1.873	0.636	0.065	1.2673- 2.7686
	radiation+surgery+chemo	2.023	0.804	0.076	1.2805- 3.1953

Table 4.8 Results of multivariable cox regression analysis of colorectal cancer patients in Tikur Anbessa specialized hospital, Addis Ababa, Ethiopia.

Covariates	Category	HR	Std.err	p- value	95% CI(HR)
Family history	Yes(Ref)	-	-	-	-
	No	1.648	0 .370	0.026*	1.060- 2.560
Alcohol consumption	Yes(Ref)	-	-	-	-
	No	0.327	0.070	0.000***	0.2144- 0.4980
Marital status	Single (Ref)	-	-	-	-
	Married	0.503	0.135	0.011*	0.2968- 0.8515
Smoking status	Smoking(Ref)	-	-	-	-
	non-smoking	0.846	0 .189	0.454	0.5455- 1.3113
BMI	Under Weight (Ref)	-	-	-	-
	Healthy Weight	0.784	0.176	0.280	0.5046- 1.219
	Over Weight	1.016	0.3118	0.959	0.5567- 1.8542

Comorbidity	Yes(Ref)	-	-	-	-
	No	0.703	0 .148	0.094	0 .465- 1.062
Physical Exercise	Yes(Ref)	-	-	-	-
	No	3.027	1.000	0.001**	1.583- 5.786
Stage	Stage I and II(Ref)	-	-	-	-
	Stage III	1.777	0.671	0.128	0.848- 3.725
	Stage IV	2.459	0.883	0.012*	1.216- 4.973
Tumor grade	Well-differentiated(Ref)	-	-	-	-
	Moderately differentiated	1.246	0.358	0.443	0.710- 2.1875
	poorly differentiated	1.935	0.459	0.005*	1.216- 3.079
	Radiotherapy alone				
	surgical treatment alone	-	-	-	-
	chemotherapy alone	0.382	0.299	0.220	0.0821- 1.779
treatment modality	surgery plus chemo	2.051	0.711	0.038*	1.0392- 4.049
	radiation+surgery+chemo	2.000	0.717	0.053	0.9906- 4.0386
		1.673	0.680	0.206	0.7537- 3.7145

Table 4.9: Test of proportionality assumption by Schoenfeld residual

Covariates	Categories	Roh	chi2	Prob>chi^2
Family history	No			
	Yes	0.05531	0.37	0.5432
Alcohol consumption	Yes			
	No	0.17889	3.75	0.0527
	Single			
Marital status	Married	0.04912	0.30	0.5839
	Overweight			
BMI	Healthy weight	-0.09802	1.37	0.2419
	Overweight	0.11338	1.53	0.2167
Smoking status	Smoke			
	Non-smoker	0.30303	12.20	0.005
Comorbidity	Yes			
	No	0.04029	0.19	0.6634
Physical Exercise	Yes			
	No	-0.09055	0.94	0.3329
Stage	Stage I and II			
	Stage III	0.01401	0.02	0.8838
	Stage IV	0.07219	0.58	0.4471

Tumor grade	Well-differentiated			
	Mod. differentiated	-0.20676	5.29	0.0214
	Poorly differentiated	0.08677	1.03	0.3101
	Radiotherapy alone			
Treatment modality	surgical treatment alone	0.05282	0.30	0.5834
	chemotherapy alone	-0.05704	0.37	0.5455
	surgery plus chemo	0.02206	0.06	0.8121
	radiation+surgery+chemo	0.02206	0.58	0.4481
global test		NA	28.72	0.0259

Table 4.10 univariable analysis of log-logistic model

Covariates	Categories	Coef	SE	Φ	z	p-value	75 % CI
Age	>40(Ref)	-	-	-	-	-	-
	40-49	0.109707	0.1878686		0.58	0.559	-0.2585-0.4779
	50-59	-0.0608519	0.1773752		-0.34	0.732	-0.4085-0.2867
	>=60	-0.1712944	0.1738658		-0.99	0.325	-0.51206- 0.1695
Family history	Yes (Ref)	-	-	-	-	-	-
	No	-0.2977782	0.1319795		-2.26	0.024	-0.5564- -0.0391
Alcohol consumption	Yes (Ref)	-	-	-	-	-	-
	No	0.6094746	0.1239228		4.92	0.000	0.3665-0.8523
Marital status	Single (Ref)	-	-	-	-	-	-
	Married	0.4868904	0.1820788		2.67	0.007	0.1300-0.84375
Body mass index	Underweight (Ref)	-	-	-	-	-	-
	Healthy weight	0.060997	0.12875		0.47	0.636	-0.1914- 0.31335
	Overweight	0.0091908	0.1786506		0.05	0.959	-0.3409-0.35933
Comorbidity	Yes(Ref)	-	-	-	-	-	-
	No	0.2103361	0.1294421		1.62	0.104	-0.04336-0.46403
Physical exercise	Yes (Ref)	-	-	-	-	-	-
	No	-0.7501022	0.1817848		-4.13	0.000	-1.106- -0.3938
Smoking status	Smoker (Ref)	-	-	-	-	-	-
	non-smoker	0.2232406	0.1313569		1.70	0.089	-0.0342-0.48069
Stage of the diseases	Stage I and II (Ref)	-	-	-	-	-	-
	Stage III	-0.2469154	0.2134521		-1.16	0.247	-0.6652-0.17144
	Stage IV	-0.4598481	0.1969292		-2.34	0.020	-0.8458- -0.0738
Tumor grade	Well-						

	differentiated(Ref)	-	-	-	-	-	-
	Moderately differentiated	-0.0523748	0.1574285		-0.33	0.739	-0.3609-0.25617
	Poorly differentiated	-0.2406802	0.14031		-1.72	0.086	-0.51568-0.03432
Treatment Modality	Radiotherapy alone (Ref)						
	surgical treatment alone	-	-	-	-	-	-
	chemotherapy alone	0.5662594	0.3918911		1.44	0.148	-0.2018- 1.3343
	surgery + chemo	-0.4089032	0.1891695		-2.16	0.031	-0.7796- -0.0381
	radiation +surgery+ chemo	-0.4100433	0.1981217		-2.07	0.038	-0.7983- -0.0217
		-0.3184986	0.2345827		-1.36	0.175	-0.7782- 0.1412

APPENDICES B

Kaplan-Meier curves of categorical predictor variables