COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

MODELING TIME TO LOSS FOLLOW-UP AMONG WOMEN UNDER BREAST CANCER PATIENTS AT HAWASSA UNIVERSITY COMPREHENSIVE SPECIALIZED HOSPITAL: APPLICATION OF COMPETING RISK MODEL

BY: TSEGISHET TEMESGEN

MARCH, 2025

HAWASSA ETHIOPIA

MODELING TIME TO LOSS FOLLOW-UP AMONG WOMEN UNDER BREAST CANCER PATIENTS AT HAWASSA UNIVERSITY COMPREHENSIVE SPECIALIZED HOSPITAL: APPLICATION OF COMPETING RISK MODEL

BY TSEGISHET TEMESGEN

SPECIALIZATION: BIOSTATISTICS

ADVISOR: DEREJE DENBE (PhD)

A THESIS SUBMITTED TO HAWASSA UNIVERSITY, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES, DEPARTMENT OF STATISTICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

MARCH, 2025

HAWASSA, ETHIOPIA

ADVISORS' APPROVAL SHEET

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY ADVISORS' APPROVAL SHEET

This is to certify that the thesis entitled "MODELING TIME TO LOSS FOLLOW-UP AMONG WOMEN UNDER BREAST CANCER PATIENTS AT HAWASSA UNIVERSITY COMPREHENSIVE SPECIALIZED HOSPITAL: APPLICATION OF COMPETING RISK MODEL" submitted in partial fulfillment of the requirements for the degree of Master's with specialization in Biostatistics the Graduate Program of the Department of Statistics, and has been carried out by Tsegishet Temesgen,ID. No: GpBiStR/0007/15, under my supervision. Therefore I recommend that the student has fulfilled the requirements and hence hereby can submit the thesis proposal to the department.

Dereje Denbe (**PhD**)	_____	_____
Name of major advisor	Signature	Date

EXAMINER'S APPROVAL SHEET-1

HAWASSA UNIVERSITY SCHOOL OF GRADUATE STUDIES

EXAMINERS' APPROVAL SHEET-1

(Submission Sheet-2)

We, the undersigned, Members of the Boards of Examiners of the final Open Defense by Tsegishet Temesgen have read and evaluated this thesis entitled "Modeling time to loss follow-up among women under breast cancer patients at Hawassa university comprehensive specialized hospital: application of competing risk model'' and examined the candidate. This is therefore, to certify that the thesis has been accepted in partial fulfillment of the requirement for the Degree of Master of Science in Biostatistics.

_____	_____	_____
Name of major Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner-I	Signature	Date
_____	_____	_____
Name of Internal Examiner-II	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date
_____	_____	_____
SGS Approval	Signature	Date

DECLARATION

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been properly acknowledged.

Name: Tsegishet Temesgen

Signature: _____

This MSc thesis has been submitted for examination with my approval as thesis advisor.

Name: DEREJE DENBE (**PhD)**

Signature: _____

Place and Date of Submission: _____

Acknowledgement

First and foremost, I would like to extend my unshared thanks to the almighty God for providing me the opportunity for what I have achieved and for his mercy. I would like to express my deepest gratitude to my kind and dedicated advisor, Dereje Denbe (PhD) for his invaluable guidance throughout this journey. His generous support provision of essential resources and insightful feedback has been crucial to the successful completion of this work.

I am profoundly grateful to the Statistics staff members of Hawassa University for their unwavering commitment to academic excellence and for creating a conducive learning environment. I extend my heartfelt thanks to the staff of Statistics Anteneh Bezabih (MSc) for his valuable comments suggestion and cooperation. I would like to gratefully thank the" Hawassa university comprehensive specialized hospital oncology center" for giving me an opportunity to access their data set.

Finally and most importantly, my deepest appreciation goes to my family members**,** whose love, prayers, and encouragement have been my greatest source of motivation. Their unwavering belief in my potential has kept me going through challenging times. Thank you for always being by my side. I am thankful to Abenezer yohanis (MSc) and all my best friends for their unconditional love unwavering support, they have shared this academic journey with me, collaboration, and insightful discussions that have enriched my learning experience. Lastly, I would like to extend my gratitude to everyone who contributed to this work, whether directly or indirectly. Your encouragement, advice, and assistance have played a significant role in making this achievement possible. Thank you all for your support and for being part of this journey.

Abbreviation and Acronyms

ACS…….……...…….American Cancer statistics

AIDS…….…..…….... Acquired Immunodeficiency Syndrome

BC.....Breast Cancer

CDC……..…….....Center of Disease Control and prevention

CICR.....Cumulative Incidence Competing Risk

CIF..... Cumulative Incidence Function

HIV.....Human Immune Deficiency Virus

CPH..... Cox proportional Hazard

CSC ……..……..... Cancer Statistics Center

CSH..... Cause Specific Hazard Model

CSHR..... Cause Specific Hazard Ratio

ECA……..……..... Ethiopian Cancer Association

GCO……..…...… Global Cancer Observatory

GLOBOCAN……..……...…... Global Cancer Statistics and Estimates

HUCSH……..……..…. Hawassa University Comprehensive Specialized Hospital

IARC…….…..…. International Agency for Research on Cancer

KM..... Kaplan Meir

LMICs …..……...….. Low- and Middle-income Countries

LTFU.....Loss to follow up

MOH…….….. Minister of Health

NCD.....Non Communicable Disease

Pt.……..……Patients

RCC.....Regional cancer center

RD..... Relative Difference

SDH.....Sub-distribution Hazard

SDHR.....Sub-distribution Hazard Ratio

TASH…….…….…….....Tikur Anbesa Specialized Hospital

WHO.....World Health Organization

Table of Contents

List of tables Pages

List of figures Pages

Abstract

Incomplete breast cancer treatment poses a significant challenge for breast cancer programs. This study aims to assess risk factors for treatment discontinuation and loss to follow-up among breast cancer patients at Hawassa University Comprehensive Specialized Hospital, considering death as a competing risk event. This study used all four year data of 463 breast cancer patients under follow up between January 2020 and December 2023. Descriptive analysis and cumulative incidence survival curve evaluations were conducted in the presence of competing risks. Univariate and Multivariate cause-specific and sub-distribution hazard analyses examined the factors associated with loss to follow-up, with death as a competing event. Out of 463 breast cancer patients recorded, 96 (20.73%) were loss to follow up, 121(26.13%) were died and 246 (53.13%) censored during the follow up. The cause-specific hazard and sub-distribution hazard models revealed that age, place of residence, types of Treatment, Stage of cancer, has non communicable disease, performance status, marital status, distance, experienced chemotherapy, experienced radiotherapy, expected adverse effect, histologic grade were the significant risk factor associated with time to loss follow up treatment. Patients with non-communicable disease had 1.846 and 2.8589 times higher risk of losing follow up compared to patients without non communicable disease comorbidities (CSHR : 1.846(0.56,2.27859))and (SDHR2.8589(0.567,4.3006)) respectively. Distance patients greater than 100km have 1.954 and 201times high risk of lost follow up compare to distance less than 100km (CSHR:2.011(0.67,2.52991) and SDHR: 1.954(0.623,6.122)). The findings showed that the estimates of the covariates effects were different for the two hazard models. This study conclude that breast cancer patients with non-communicable diseases, those residing far from the treatment center, experiencing adverse effects during follow-up, having advanced cancer stages, and being aged 60 years and older are at a higher risk of loss to follow-up. These findings highlight the critical need for Targeted interventions are essential to enhance patient retention and prevent treatment discontinuation. Finally this study recommended that the hospital authorities to give attention to patients those who are distant from Hospital, had Has NCD Comorbidities, Experienced Advert Effect at follow up and older patients.

***Key Words**: Competing risks, Loss to follow up, Breast Cancer, survival analysis, fine gray model, sub-distribution hazard*

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

Discontinuation of breast cancer follow-up is a significant challenge for patient outcomes globally, particularly in Africa (Dairo et al., 2018). Breast cancer is a type of cancer that arises from uncontrollably altered cell proliferation or function in the breast tissue (Saudi Minister of Health (SMOH), 2021). Breast cancer comes in various forms, which breast cells develop into cancer determines the type of breast cancer (WHO; (Shaikh et al., 2021). Most cases of breast cancer start in the lobules or ducts. Blood and lymph vessels are two ways that breast cancer can travel outside of the breast (Saunders & Jassal, 2009). Breast cancer was the most prevalent cancer diagnosed in 2021 and the primary cause of cancer-related deaths among women worldwide (Lei et al., 2021; IARC, 2022). It affected an estimated 2.26 million people, accounting for 25.8% of all new cases of cancer and 685,000 fatalities (17%) in 2020 (Sung et al., 2021).

Loss to follow-up (LTFU) is a significant issue in breast cancer care, often hindering early detection of recurrence and timely interventions, which are critical for patient outcomes. LTFU is characterized by patients discontinuing or inconsistently attending post-treatment appointments, often due to socioeconomic factors, geographic challenges, or systemic healthcare barriers. Addressing the factors driving LTFU is thus essential for improving outcomes and managing the long-term burden of breast cancer (Baccolini et al., 2022; Jones et al., 2020). In 2020, there were over 2.3 million new cases and 670,000 deaths from breast cancer worldwide with approximately 84% of new cases and 88% of deaths occurring in low- and middle-income countries (Kumar et al., 2022). In America, approximately 1 in 8.2 women will be diagnosed with breast cancer in their lifetime and 1 in 39 was die (Giaquinto et al., 2022). About 60% of the new cases and deaths worldwide in 2020 occurred in low-and middle-income countries (Martinez et al., 2024). The intricate etiology of breast cancer is influenced by the interaction of various socio demographic and clinical variables (WHO, 2021; Nawaz et al.,2023).

In Africa, breast cancer represents 28% of all cancer cases and 20% of cancer-related deaths in women, with 186,598 cases and 85,787 deaths projected in 2020 (Sung et al., 2021). The IARC reported breast cancer incidence rates ranging from 26.8 per 100,000 women in Middle Africa to 38.9 per 100,000 in Southern Africa in 2012 (Ferlay et al., 2019). Breast cancer is the leading cause of cancer-related death among women in sub-Saharan Africa, it accounts for one out of every four diagnosed cases and one out of every five cancer-related fatalities. Loss to follow-up is a major challenge in the successful management of breast cancer patients in sub-Saharan Africa; true outcomes of patients lost to follow-up thus become difficult to assess. This Loss of follow-up due to economic, geographical, social, or patient-related factors, threatens the effectiveness of cancer care programs (Gradishar et al., 2022).

In Ethiopia, cancer is the fourth leading cause of death, with an estimated 51,865 deaths in 2020 (Afework et al., 2022). The most prevalent cancers in Ethiopia among the entire adult population are breast cancer (30.2%), cervical cancer (13.4%), and colorectal cancer (5.7%). According to a collaborative report (Global Cancer Statistics *2020*, WHO, 2022 and IARC), there are a research shortage in low- and middle-income countries (LMICs) (Ginsburg et al., 2017; Porter, 2008). Every year, there is an increase in the incidence of breast cancer (Deressa et al., 2019). Ethiopian women come from a varied range of backgrounds and lifestyles, with varying customs, economic standing, and approaches to childrearing and reproduction.

A study conducted in gonder from the total sample, 300 (62.2%) patients had confirmed (135 were alive, 165 have died), 182 (37.8%) patients were lost to follow-up (Tiruneh et al., 2021). Also a study conducted in addis ababa reported that 231 (21.6%) of breast cancer patients were lost to follow-up, while 285 (26.6%) experienced an event, and 554 (51.8%) were censored during the study period (Kantelhardt, Zerche, et al., 2014).Accordingly, the outcome of 245 (81.1%) patients was confirmed: 178 (58.9%) were alive, 67 (22.2%) were deceased, and 57 (18.9%) patients were LTF. Of the 302 BC patients, 252 (83.4%) presented with advanced stage at the first hospital visit: 161(53.3%) with stage III and 91(30.1%) with stage IV disease. Thirty-five (11.6%) patients were diagnosed with comorbidity, out of which 23 (7.6%) had hypertension and 5 (1.6%) had HIV (Shita et al., 2023). Breast cancer was found to be the most common cancer, accounting for 18.62% of all cancers registered at HUCSH (Shalamo et al., 2022).

Follow-up cancer care is important for patients who have received treatment but some patients discontinue their care and are lost to follow-up (LTFU) at the cancer center where they were treated. The benefits of follow-up in breast cancer patients include early detection of potential curable events, management of side effects, psychosocial care, support, physical exercise, weight reduction, adjuvant therapy reevaluation, and endocrine therapy monitoring. However, these benefits are missed by patients discontinuing follow-up (Gill et al., 2018).

1.2 Statement of the Problem

Breast cancer is the leading cause of cancer death among women in developing countries (Wilkinson & Gathani, 2022). Ethiopia plans to eliminate breast cancer as a public health problem by 2030, following the World Health Organization's call for action. However, a significant challenge in breast cancer management is loss to follow-up (Foerster, et al., 2020). The factors associated with LTFU were described in previous studies (Bojude et al., 2018; R. Feng et al., 2020; Gill et al., 2018; Ruddy et al., 2020). However, most of these studies were conducted in Western countries. The challenge of treatment discontinuation in breast cancer management is considerable, with many patients failing to complete therapy. This contributes to higher mortality and emphasizes the need for consistent follow-up efforts (Vidhubala et al., 2020). Despite recognition of LTFU as a global issue, it remains particularly pressing in Africa (Bojude et al., 2018) and requires greater attention in Ethiopia.

Many scholars were used logistic regressions to determine predictors of LTFU under breast cancer treatment program, Michelle from Nigeria, Ouyang from Guangzhou (china), R. Feng from Shanxi (central china) (R. Feng et al., 2020; Holmes et al., 2013; Michelle et al., 2023; Ouyang et al., 2021). Logistic regression does not account the censoring observations that mean it does not hold for time to event data (Vock et al., 2016). Logistic regression analysis leads to incorrect conclusions because the Regression model can't handle time to event of interest, competing event and censoring. However, survival analysis will be more appropriate than logistic frame work that takes censoring in to considerations (Singer & Willett, 1993). Therefore, there is a restriction on the events when we use logistic models for the follow up time which loss massive information (King & Zeng, 2001).

Most of the previous works done on the issue of lost follow up among breast cancer patients had applied the standard survival analysis methods without considering the occurrence of other events which could potentially alter the chance of the occurrence of the event of interest (Bojude et al., 2018; Dairo et al., 2018; Ng et al., 2015; Ruddy et al., 2020). Survival models like cox proportional and Kaplan-Meier treat competing events as censoring (Klein and Moeschberger, 2003). So ignoring the potential effects of competing events can easily lead to incomplete reporting, incorrect risk estimates, and potentially lead to wrong interpretation. Alternative methods are required that specifically designed for analyzing competing risk data (Noordzij et al., 2013).

Competing risk results are often documented as a multi-component endpoint, with the first event being an "event of interest" or a "competing risk event" (Wolbers et al., 2014). In LTFU treatment studies, mortality is considered a competing risk event, as the patient is no longer at risk of treatment. Hence, computationally efficient alternative methods are required that are specifically designed for analyzing competing risk (Krawczyk, 2016).

To the best of our knowledge, no studies in Ethiopia examined patients' loss to follow up breast cancer treatment by using competing risk model. Thus, this study aims to address this gap. Understanding the factors associated with loss to follow up treatment of breast cancer patients during follow-up treatment is essential in Ethiopian context. The lack of existing studies on this topic has been key motivation to our research. By filling this knowledge gap, our study aim to provide evidence-based insights to inform clinical decision-making, improve patient outcomes, and support the development of targeted interventions to reduce loss to follow-up in these patients. Recognizing these factors will aid in identifying the impact of loss to follow-up and strategies to control it, representing a significant step towards enhancing treatment outcomes and delivering comprehensive care for these patients.

Therefore, this study answers the following major research questions:

I. What were the risk factors associated with loss follow-up among breast cancer patients?
II. What were the risks of loss to follow up among patients in different group covariates?
III. Is there difference between covariates effects of cause-specific hazard and the sub-distribution hazard?

1.3 Objective of the Study

1.3.1 General Objective

The general objective of this study was to model the time to loss to follow-up among Breast cancer patients at Hawassa University Comprehensive Specialized Hospital using Competing risk model.

1.3.2 Specific Objective

The specific objectives of this study were:

- To identify factors associated with the risk of loss to follow-up among breast cancer patients.
- To evaluate the risk of loss to follow-up across different covariate groups among breast cancer patients.
- To compare the effects of covariates on outcomes using cause-specific hazard and sub-distribution hazards.

1.4 Significance of the Study

The significance of this study is to identify and give detailed information on the major prognostic factors that play a critical role in LTFU treatment among BC patients by accounting for death due to any cause as a competing risk event. For academicians or statisticians, it will direct to thoughts and genuine interest on the subject matter for further research, especially when multiple end points are occurring. The findings will aid healthcare policymakers, oncologists, and public health professionals in designing targeted interventions to reduce treatment discontinuation, ultimately improving survival outcomes for breast cancer patients. Furthermore, comparing cause-specific and subdistribution hazard models contributes to the methodological advancement of competing risk analysis in oncology research.

1.5 Scope of the study

The study aims to analyze factors influencing loss to follow-up (LTFU) among breast cancer patients at Hawassa University Comprehensive Specialized Hospital. Using cause-specific and subdistribution hazard models, it examines how demographic, clinical, and treatment-related factors affect cumulative incidence and the probability of LTFU versus other outcomes like disease progression or death. By comparing the effects of covariates in a competing risks framework, the study seeks to enhance understanding of LTFU risk factors and inform strategies for better patient follow-up and care retention. Factors contributing to loss to follow-up and their impact on patient outcomes are identified. Potential interventions to reduce loss to follow-up and improve outcomes are explored

1.6 **Operational definition**

Lost to Follow-up (LTFU): patients in a study or treatment program who discontinue participation failed to attend scheduled follow-up appointments or lost contact with the healthcare provider and not yet classified as ‘died’ or ‘transferred-out’.

Censoring is caused when a patient is referred to other hospitals, failure or end of the study. The type of censoring used in this study is random right censoring.

Drop out: when participants voluntarily withdraw from a study or treatment program before its completion.

Death: The occurrence of mortality, it often represents a final outcome for some participants.

Right censoring: When the event of interest has not occurred by the end of the study or follow-up period, leaving the exact time unknown.

Histologic grade: also known as Nottingham grade or Elston grade is a way to assess how abnormal breast cancer cells appear under a microscope compared to healthy breast cells. The individual scores are then added together to get a total score. The total score is used to determine the histologic grade:

Performance status (ECOG): A measure of the patient's functional ability and how cancer affects their daily living abilities often assessed using the Eastern Cooperative Oncology Group (ECOG) Performance Status Scale, , which ranges from 0 (fully active) to 5 (deceased), with higher scores indicating greater disability or illness.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Definition and Overview of Breast Cancer

Breast cancer originates from a nasty tumor that begins in the breast cells (Jahan et al., 2016). A collection of cancer cells that have the ability to develop into nearby tissues or travel (metastasize) to other parts of the body is called a malignant tumor (Breast Cancer Org; Feng et al., 2018). Men can also contract the condition, but women account for the majority of cases. The primary components of a typical female breast include ducts, which are tiny tubes that convey milk from the lobules to the nipple, stroma, which is connective tissue, fatty tissue, blood vessels, and lymphatic vessels, that surround the ducts and lobules (Rass, 2023).The cells lining the ducts are where most breast malignancies start (ductal tumors). Certain types of malignancies originate from cells lining the lobules, whereas a tiny percentage originates from other tissues (ACS, 2016).

Despite the availability of effective screening and treatment options, many women still face barriers to accessing healthcare services, particularly in low and middle income countries (LMICs) (Knaul et al., 2012). One of the major challenges in managing breast cancer is loss to follow-up, which refers to patients who do not return for scheduled appointments or fail to complete their treatment regimen (Knaul et al., 2012). Loss to follow-up (LTFU) and/or noncompliance to the surveillance plan are detrimental to the reliability of clinical research (B. D.Li et al., 2000). Furthermore, LTFU might increase the risk of non-adherence to endocrine therapy, which might compromise long-term diseases control. Thus, it is important to analyze the risk factors of LTFU. The burden of caring for these large numbers of patients in a low resource country is enormous (Bojude et al., 2018). Even though the number of cancer cases in Africa is rising, the continent still places a low priority on the disease due to a lack of funding and other urgent public health issues, such as communicable diseases like AIDS, malaria, and tuberculosis, as well as HIV/AIDS-related infections (Elsherif et al., 2012).

Most breast cancer is sporadic (90 to 95 percent), only 5–10% of individuals have a detectable genetic mutation, (Alkabban & Ferguson, 2024). The most prevalent genetic disorders linked to

breast cancer are BRCA 1 and 2. This subtype, which accounts for 1% to 2% of breast tumors, is microscopically defined by infiltrating cells of minimum type that form tiny glands and tubules (Roux et al., 2019). These aggressive, poorly differentiated tumors are more prevalent in younger patients and those with BRCA mutations (Cserni, 2020). In 2020, invasive breast cancer accounted for around 11.7% of new cases, making it the most frequent cancer among women globally (Sung et al., 2021). However, the incidence and mortality rates of breast cancer are still rising in several Asian and African nations, including South Korea, India, and SSA countries (Alkabban & Ferguson, 2024; Sung et al., 2021).

Breast cancer is a disease that often affects women over a mid-age. Age is the strongest predictor for the death and lost follow up of breast cancer patients (Huang, 2013). Due to ageing of the population, the number of older patients diagnosed with breast cancer is rapidly growing and observance was highest among women aged 50–59 years. With increasing life expectancy, older patients remain at the risk to follow up because of breast cancer over a longer time period, which can impact their treatment outcomes and overall health management (Santiá et al., 2022). Women under the age of 20 are rarely affected, whereas women over 65 account for more than 15% of diagnoses. However, cancer usually strikes women over 65 who were not regularly screened (Mueller, 2019).

The goal of the Global Breast Cancer Initiative (GBCI) is to prevent 2.5 million breast cancer deaths worldwide between 2020 and 2040 by reducing the global breast cancer mortality rate by 2.5 percent annually (WHO, 2024). 25% of breast cancer deaths among women age less than 70 would be prevented by 2030 and 40% by 2040 if the worldwide breast cancer mortality rate were reduced by 2.5 percent annually (WHO; 2024).

Developing nations account for over 60% of breast cancer-related mortality (da Costa Vieira et al., 2017) and Approximately 40% out of 1328 were lost follow up care (Foerster, et al. 2020). Breast cancer risk factors include alcohol, cigarette smoking, obesity, physical inactivity, diet, hormone therapy after menopause and family history of breast cancer according to the World Health Organization (WHO). Effective treatment initiation and follow-up care is essential for monitoring signs of recurrence, managing treatment-related side effects and providing psychosocial support (Ward et al., 2019).

A study (Ngo et al., 2024) conducted in Cameroon in 2024 and 500 patients of breast cancer patients was included. During the follow-up, 98 patients died, 168 were lost to follow-up, and 234 were alive at the closing date of the study. The independent factors associated with a reduction in survival were breast ulceration (HR = 3.23; p = 0.002), bilateral location of the tumor (HR = 9.2; p < 0.001) and clinical stage 3 (HR = 1.72; p = 0.010).

In Ethiopia a national cancer program follows the strategy and uses recognized international criteria for the diagnosis and treatment of breast cancer patients (Ayele et al., 2021). However, the treatment success rate of breast cancer in Ethiopia is generally lower than in developed countries due to the aforementioned factors (Tesfaw et al., 2020).

The success rate can vary depending on the stage of the cancer at diagnosis and the availability of treatment options. In general, early-stage breast cancer has a higher success rate than advanced-stage cancer. In Ethiopia, the success rate of breast cancer treatment varies depending on the region and the availability of resources (Birnbaum et al., 2018). According to a study conducted in Addis Ababa, Ethiopia, the 5-year survival rate for breast cancer patients was 38.6%. However, this study only included patients who received treatment at a single hospital, so the success rate may vary in other regions of the country (Zerche, et al., 2014).

2.2 Predictors of Breast Cancer

Various studies have explored the risk factors associated with loss to follow up among breast cancer patients on the world. A study in Nigeria found (Bojude et al., 2018) Out Of 504 patients in the program, 500 women met inclusion criteria, of whom 350 (69.5%) 5 year LTFU, 467 (92.6%) were 10 year lost to follow up (LTFU) and 154 (30.5%) , 37(7.3%)were alive and in care respectively. In adjusted analyses, patients with no documented disease stage and advanced stage at presentation were more likely to be early lost to follow up (LTFU) vs. patients with stage 1 and 2 when controlling for other factors (HR: 1.310, 95% CI 1.047, 1.639), Older Age (HR: 1.415, 95% CI 1.044, 1.917), marital status, with palliative care as type of treatment received and patients with non-communicable disease (NCD) (HR: 1.404, 95% CI 1.120, 1.760) were more likely to be late LTFU when controlling for other factors. Patients with metastasis were also likely to be late LTFU (HR: 1.793, 95% CI 1.396, 2.302). Also (Dairo et al., 2018)

shared that patients marital status, older age, having non-communicable disease (NCD) and had advanced stage were Significant factors for lost to follow-up of BC patients.

In a study conducted in china an examination was undertaken on the risk factors associated with loss to follow up of breast cancer patients. In order to analyze the data, the researcher used a logistic regression model combined with inferential statistics and descriptive statistical methods. Age (younger and older), lack of health insurance, distance from home to hospital, pathology (DCIS/Paget's), lymph node metastasis, lack of endocrine therapy, and fewer than five contact numbers were all significantly and independently linked to the risk of LTFU, according to the data analysis. Furthermore, there was a strong correlation between LTFU and not receiving adjuvant chemotherapy, radiation, and endocrine treatment compared to receiving them (Ouyang et al., 2021).

A study in India (Holmes et al., 2018) found that patients who lived farther from the treatment center(most rural countries), had lowest income, and advanced Cancer stage were more likely to be lost to follow-up. Also another study in India (Bhatt et al., 2018) found that patients who had financial constraints, lacked social support, and had poor knowledge about breast cancer were at higher risk of loss to follow-up. A study in Israel (Amat et al., 2022) found that patients who lived farther from the treatment center ,had a trainee as their primary provider, accessing primary care, age , and were more likely to be lost to follow-up.

To address this issue, researchers have proposed various strategies to improve retention of breast cancer patients in care. One approach is to use predictive models to identify patients who are at high risk of loss to follow-up. A study (Foerster, et al., 2020) conducted on losses follow up in sub-Saharan African by employing competing risk model found that, of several variables hypothesized to affect follow up care, like increasing age , HIV AIDS positive, residence(urban) was associated with a higher risk of being LTFU. A cumulative total of 19 women had become LTFU by the end of year 1 and 58 by the end of year 2, yielding LTFU percentages of 1.3% (95% CI: 0.8, 1.9) and 4.0% (95% CI: 3.1, 5.1), respectively. At 3 years, LTFU percentages were low in Nigeria (n = 3; LTFU = 0.8%, 95% CI: 0.2, 2.2) and Namibia (n = 10; LTFU = 2.2%, 95% CI: 1.1, 3.9) and slightly higher in Uganda (n = 21; LTFU = 5.6%, 95% CI: 3.6, 8.3).

A study (Feng et al., 2020) developed a risk score model based on patient demographics, clinical characteristics, and treatment factors to predict loss to follow-up among breast cancer patients. The model was able to accurately predict which patients were more likely to be lost to follow-up and could be used to target interventions to improve retention in care.

Overall, modeling time to loss to follow-up among breast cancer patients is an important area of study that can help identify high-risk patients and improve their clinical management (Ng et al., 2015) by understanding the factors associated with loss to follow-up and developing predictive models, healthcare providers can implement targeted interventions to improve retention in care and ultimately improve outcomes for breast cancer patients (Akl et al., 2012).

A study conducted Limbe Regional Hospital (Tchounzou et al., 2019) found One hundred twenty-six patients had symptoms suggestive of cancer, but only 110 (87.30%) could pay for biopsy, 29 (26.36%) of those did not collect their results, 17 (18.7%) denied their results, and 20 (19%) did not benefit from treatment. Only 44 of 110 patients were able to finish their cancer care treatment program. Reasons for discontinuing the cancer care included lack of financial means to pay for it, distance from the care center and belief in alternative treatments.

Previous studies on breast cancer loss to follow-up (LTFU) and death have provided valuable insights into patient survival and risk factors. Strong aspects include the use of survival models, competing risk analysis, and identification of key predictors such as disease stage and treatment type, including in Ethiopia. However, weaknesses include limited longitudinal data, high censoring rates, high rates of missing information and inadequate follow-up systems and a lack of research in low-resource settings like Ethiopia. In Ethiopia, studies are scarce, often relying on small sample sizes and hospital-based data, limiting generalizability. Additionally, few studies address socioeconomic and healthcare accessibility challenges influencing LTFU and some studies also fail to account for competing risks, leading to potential bias in estimates.

2.3 Summary of Competing Risk Models

The study has done on Second primary malignancies associated with radiation therapy in breast cancer patients diagnosed between 2010 and 2015, a population-based competing-risk study by (Bao et al., 2021) which indicates that classical survival analysis is not appropriate when

analyzing the association between radiotherapy and second malignancies in order to avoid risk estimation bias caused by conventional statistical methods which improperly process data of patients who die before the event of interest, who do not experience the event before the end of follow-up, or who are lost to follow-up.

A possibility for analyzing data in the presence of competing risks is the Fine-Gray sub-distribution hazard model of risk that has proved useful in the analysis of the factors associated with death for cases receiving treatment for breast cancer in this study.

A study conducted by (Cho et al., 2021) on Prognostic factors in African American and white American patients with breast cancer, found that traditional survival analysis, including the Kaplan–Meier method and Cox proportional hazards model, is typically used for single-event outcomes like death from breast cancer. However, many clinical trials involve multiple endpoints, necessitating methods that can handle competing risks, where events are mutually exclusive and impact one another.

Studies conducted by Yang et al., on competing-risks model for predicting the prognosis of penile cancer. The study shows that medical research generally has competing risks. When discussing specific causes of death, the traditional method may overestimate the cumulative incidence of each variable. It is therefore necessary to use the competing-risks model to deal with multiple end events (Yang et al., 2019).

In medical studies, competing risks data arise when patients encounter failure from multiple causes, making it difficult to observe failure from other causes. A semi-parametric proportional sub distribution risks model was proposed by Fine and Gray (Fine & Gray, 2015) to assess covariate effects on the cumulative incidence function (CIF). The Cox proportional hazards model has been expanded to include survival data analysis in various contexts, including competing risks, clustering survival, recurrent events, and interval censored (Logan et al., 2006).Competing events are not treated as typical censoring events without impact on the cumulative incidence function for the event of interest, in contrast to the Kaplan–Meier method's use.

However, in many clinical trials, the number of endpoints is not limited to one; they therefore have a “competing” trend. In situations with competing risks, failure from one cause limits the observation of other events. Fine and Gray’s semi-parametric model allows direct assessment of covariate effects on the cumulative incidence function (CIF). This is crucial in scenarios where participants may experience multiple events, affecting the visibility of the event of interest (J. Li et al., 2015). A competing risk situation happens when the occurrence of one type of event changes the ability to observe the event of interest (Pintilie, 2011). Cox proportional hazards model has been extended to various settings of survival data analysis, including competing risks data, clustered survival data, recurrent events data and interval censored data (Collett, 2023). The CICR accounts for all types of events; in the case of competing events, the cumulative incidence function is estimated both for the event of interest and for all competing events, and their estimates depend on each other (Palareti et al., 2010). Unlike in the application of the Kaplan–Meier method, competing events are not handled as regular censoring events without influence on the cumulative incidence function for the event of interest (Palareti et al., 2010)..

Fine and Gray proposed proportional SDHs model that links the covariate effect and cause-specific CIF and facilitates the direct modeling of marginal hazard of the event via the proposed “SDH” (Mozumder et al., 2021). There are two main approaches to statistical modeling of competing risks data are modeling CSHs and direct modeling of the cause-specific CIF (Varadhan et al., 2010). In the first approach, both the CSH function of the primary event of interest and any competing events must be modeled if an estimate of the cause-specific CIF is required. Thus, with one event of primary interest and one competing event, two CSH models are required. For the second approach of direct modeling of the cause specific CIF, the most common method is to use the proportional SDHs model first described by Fine and Gray. The model is semi-parametric in that the baseline sub-distribution hazard function is not directly estimated with the only parameters directly estimated being log-sub distribution hazard ratios (Lambert et al., 2017).

CHAPTER THREE

3. METHODOLOGY

3.1 Description of the Study Area

This study was conducted at Hawassa university comprehensive specialized hospital, which is one of the tertiary and teaching hospitals in the Sidama region, Hawassa town. Hawassa is located 275 km to the southeast of the capital city of Ethiopia, HUCSH is the first referral hospital established in the region serving as a teaching hospital for the College of Medicine and Health Science of Hawassa University, with a catchment population of 10-12 million. It serves about 43,384 patients of all types per year. The hospital is located in Hawassa city with its precise geographic coordinates of 7°03'33.2" North latitude and 38°28'28.4" East longitude, the hospital is easily accessible to patients from various parts of the city and surrounding areas. HUCSH Oncology Center is the center of care for cancer treatment in which chemotherapy treatment, and palliative care and other comprehensive care services are being delivered for cancer patients. Hawassa University Comprehensive Specialized Hospital serves as a major referral center for oncology care in the region, making it an ideal setting to study breast cancer patient follow-up. There is limited research on breast cancer follow-up in southern Ethiopia, and studying this location helps address the gap and provide region-specific recommendations.

3.2 Study Design

Retrospective study design was employed to identify factors associated with loss to follow-up among breast cancer patients. The data used for the study collected from medical records. The study used competing risk statistical methods and analyzed the data. This study used a model that predicted the time to loss to follow-up, while taking death into account as a competing risk event.

3.3 Study Population and period

The population of the study was all patients with Breast Cancer who had registered at HUCSH from 1st January 2020 to 30th December 2023. All the data reviewed carefully from the registration log book and patients' registration card; the data checked from the file if any

inadequate information counters and excluded from analysis if proven to be inadequate. The data was collected, reviewed and analyzed from April 2024 to February 2025.

3.4 Data Collection Procedure

This research utilized secondary data collected from a hospital setting by trained enumerators and the principal investigator. The data was carefully reviewed using the registration logbook and patients' registration cards. Any records with insufficient information were removed from the dataset and excluded from the analysis. The dataset included demographic information, cancer stages, treatments and other reasons for loss to follow-up. The collected data were analyzed using a competing risk model, which estimated the probability of loss to follow-up over time while considering death as a competing event.

3.4.1 Inclusion and Exclusion Criteria

Inclusion Criteria: - All patients who were under treatment of breast cancer at HUCSH during the study period and registered with full information including in the registration log book or in the patient's identification card were considered eligible for the study.

Exclusion Criteria: - Patients who had not started treatment for breast cancer and those with insufficient information on any vital variables in the registration book or on their patient card were excluded from the study. Additionally, breast cancer patients who began treatment after the start of the study period (midway through the study) were also excluded.

3.5 Data Structure

The table below illustrates the standard way of representing competing risks data for modeling time to LTFU treatment of BC as an event of interest and death due to any cause of BC patients as a competing risk event is illustrated in the following table.

Table 1: Data Structure

ID	Observation time	Status	Outcome(case)	age	Stage of Breast cancer
1	2	0	Censored	≤ 39	Stage 2
2	5	1	LTFU	40-49	Stage 1
3	9	2	Death	50-59	Stage 3
4	5	0	Censored	≥ 70	Stage 4

ID: - Patients

Observation time: - Time in months at which the event of interest, competing events, or censoring occurs.

Status: - 1 event of interest, 2 competing risk event and 0 censoring.

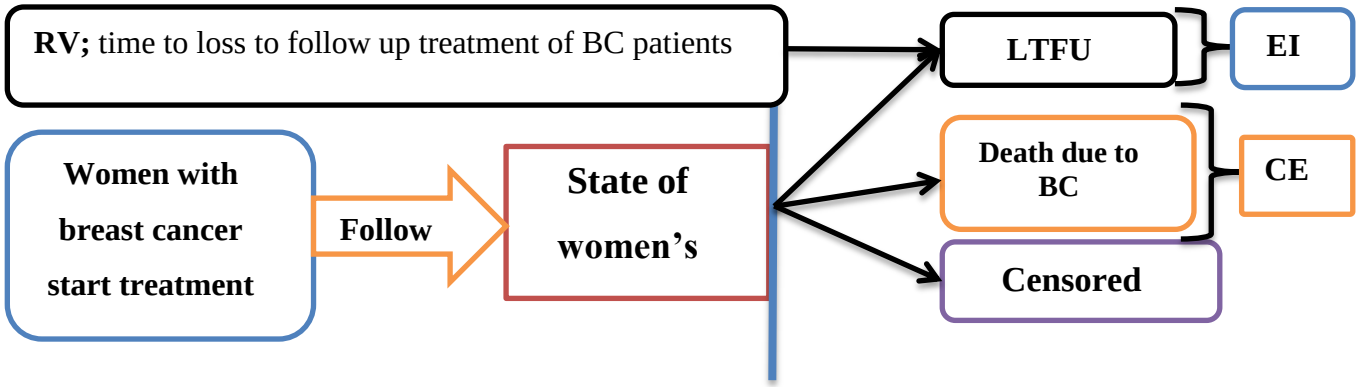
For each patient, the event of interest was the time from the month of the patient's registration at the hospital to the LTFU in months. The events of interest are loss to follow up treatment and competing risk event were death due to any cause of BC patients under follow up. The event of interest time is the time from when patients registered for BC treatment to the time when their treatment was interrupted for two consecutive months or more. Censoring was caused when a patient is referred to other hospitals or end of the study. The type of censoring used in this study was random right censoring (Collett, 1994). Age and Cancer stage were covariates.

Age: - Age of the patient at the time of diagnosis.
Stage of Cancer:-The stage of cancer at the time of diagnosis.

3.6 Variables in the Study

3.6.1 Response Variable

The response variable in this study was the time (months) to loss to follow up treatment of BC patients starting from the months of the patients registered at hospital. The status of the patient's is 1 if an interesting event occurred (loss to follow up treatment of BC patients), 2 death due to any cause of BC patient (competing risk event), 3 censored.



EI refers to the **event of interest, CE**: the **competing events** and **RV: response variable**

Figure 1: The layout of competing Risk Model

3.6.2 Independent Variables

The explanatory variables that were considered as the predictor factors associated with loss to follow up according to different literature sources (references) were described variables like Age, Distance from Hospital, Height, Weight, Has NCD commodities, Performance ECOG, Cancer stage at presentation, type of treatment received, Experienced chemotherapy toxicity, Experienced radiotherapy toxicity, marital status, educational level, histologic grade and Experienced adverse effect at follow up were used as independent Variables in this study.

Table 2: Description of Independent Variables

Variables	Categories	Codes
Age (Santiá et al., 2022)	≤ 39	0
	$40-49$	1
	$50-59$	2
	$60-69$	3
	≥ 70	4
Weight(Kimura et al., 2024)	$< 55KG$	0
	$\geq 55KG$	1
Residence	Urban	0
	Rural	1
Educational level (Dairo et al., 2018)	Primary and below	0
	Secondary and above	1
Distance from hospital (Ouyang et al., 2021)	$< 100KM$	0
	$\geq 100KM$	1
Has (non-communicable disease) NCD comorbidities	No	0
	Yes	1
Performance status(ECOG)	0-1	0

	2+	1
Marital Status (Dairo et al., 2018)	Single	0
	Married	1
	Divorce	2
	Widowed	3
Cancer stage (Manchester staging)	Stage 1	0
	Stage 2	1
	Stage 3	2
	Stage 4	3
Experienced Chemotherapy toxicity	No	0
	Yes	1
Experienced Radiotherapy toxicity	No	0
	Yes	1
Expected adverse effect as follow up	No	0
	Yes	1
Type of treatment they received	Both chemo and Radio	0
	Chemotherapy only	1
	Radiotherapy only	2
	No treatment recorded	3
Histologic grade	Grade I	0
	Grade II	1
	Grade III	2

3.6.3 Description of Variables

Age: The age of the patient at the time of data collection.

Cancer stage: The extent or severity of cancer in the patient's body, often classified into stages (e.g., Stage I, II, III, IV) based on factors like tumor size, lymph node involvement, and metastasis.

Lost to Follow-up (LTFU): Patients who were enrolled in the study but failed to attend scheduled follow-up appointments or lost contact with the healthcare provider and not yet classified as ‘died’ or ‘transferred-out’ (Assemie et al., 2018; Fufa et al., 2023; Ganta et al., 2024).

Censoring is caused when a patient is referred to other hospitals, failure or end of the study (Assemie et al., 2018; Ganta et al., 2024). The type of censoring used in this study is random right censoring (Fufa et al., 2023).

Type of treatment they received: The specific therapy or combination of therapies (e.g., surgery, chemotherapy, and radiation therapy) administered to the patient for their cancer treatment.

Performance status (ECOG): A measure of the patient's functional ability and how cancer affects their daily living abilities often assessed using the Eastern Cooperative Oncology Group (ECOG) Performance Status Scale, , which ranges from 0 (fully active) to 5 (deceased), with higher scores indicating greater disability or illness.

Expected adverse effect of lost follow-up: Predicted negative outcomes or consequences resulting from patients being lost to follow-up, such as disease progression, complications, or inability to monitor treatment response.

Experienced toxicity of Chemotherapy: any adverse reactions or side effects experienced by the patient as a result of chemotherapy treatment

Experienced toxicity of Radiotherapy: Any adverse reactions or side effects experienced by the patient as a result of radiotherapy treatment

Histologic grade: also known as Nottingham grade or Elston grade is a way to assess how abnormal breast cancer cells appear under a microscope compared to healthy breast cells. Doctors use this information, along with other factors like cancer stage and hormone receptor status, to determine prognosis and guide treatment decisions. The individual scores are then added together to get a total score. The total score is used to determine the histologic grade:

Grade 1 (well-differentiated): Score of 3, 4, or 5.These tumors are slow-growing and appear most similar to normal breast tissue.

Grade 2 (moderately differentiated): Score of 6 or 7. These tumors have some features that are abnormal, but they also have some features that resemble normal breast tissue.

Grade 3 (poorly differentiated): Score of 8 or 9. These tumors appear very abnormal and do not resemble normal breast tissue at all. They are typically the fastest-growing and most aggressive type of breast cancer.

3.7 The Statistical Models

A statistical model, specifically the competing risk model within survival analysis was employed to predict the time to loss to follow-up in among breast cancer patient data under follow up. This model incorporates a set of statistical assumptions that pertain to the generation

of a sample from a larger population. It allows for the prediction of a dependent variable (time to event) based on a set of independent variables or predictors, considering the nature of the survival time data in this context.

3.7.1 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 is employed in the analysis of time-to-event data that emerges in a number of practical domains, including epidemiology, public health, biology, medicine and demography (Aalen et al., 2008). Unlike other statistical methods, survival analysis takes into account a crucial analytical issue: censoring. Censoring basically happens when we know a little bit about a person's survival time but not the precise amount. Random right censoring is the censoring type employed in this investigation. When a research lasts until a certain point in time yet participants (patients) arrive and go at various periods, this is known as random right censoring.

The distribution of survival time is characterized by survivorship, probability density function and hazard function. Let T be a random variable associated with the survival times or in survival analysis time until a patient submits to a breast cancer and 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 survivor function $S(t)$, is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time t and so

The Survival Function takes on the following form:

$$S(t) = P(T > t) = 1 - F(t), \quad t \geq 0. \qquad \textbf{(1)}$$

The survivor function can, therefore, be used to represent the probability that a randomly selected subject survives from the time origin to some specified time beyond t.

Survival Time (T): Represents the time until a specific event occurs, in this case the time to loss to follow-up.

Survival Function ($S(t)$): models the probability that the time to loss to follow-up exceeds t, allowing us to understand follow-up duration patterns.

Cumulative Distribution Function($F(t)$): Represents the cumulative probability of experiencing the event (e.g. LTFU or death) by time *t*. The probability that the time to loss to follow-up is less than or equal to t and $1 - F(t)$) **:** Represents the probability of surviving beyond time t by subtracting the cumulative probability of experiencing the event up to that time from 1. Therefore, be used to represent the probability that, a randomly selected subject survives from the time origin to some specified time beyond t.

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

$$F(t) = P(T \leq t) = \int_0^t f(u)du, \quad t \geq 0 \qquad \textbf{(2)}$$

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

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\big(1 - S(t)\big) = \frac{-d}{dt}S(t). \qquad \textbf{(3)}$$

The hazard function is the instantaneous probability of having an event at time t (per unit time) given that one has survived up to time t (Klein & Moeschberger, 2003). The hazard function $\boldsymbol{\lambda}(t) \geq 0$ is given as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \qquad \textbf{(4)}$$

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

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{-d}{dt} lnS(t). \qquad \textbf{(5)}$$

The cumulative hazard function is defined a

$$\Lambda(t) = \int_0^t \lambda(u)du = -lnS(t). \qquad \textbf{(6)}$$

Thus

$$S(t) = e^{-\Lambda(t)} . \quad \textbf{(7)}$$

3.8 Competing Risk Analysis

Competing risks occur when the occurrence of one event prevents or alters the probability of another event of interest (Austin et al., 2016). In studies involving multiple mutually exclusive events, neglecting the effects of competing events can lead to inaccurate risk estimates and misinterpretations (Schuster et al., 2020). Traditional survival analyses, such as Kaplan-Meier and Cox regression, may not be appropriate in the presence of competing risks, as they can censor data, resulting in incomplete analyses (Schober & Vetter, 2018).

To address these challenges, two primary methods are used for analyzing competing risks: the Cause-Specific Hazard (CSH) model and the Fine-Gray Proportional Sub-distribution Hazard (SDH) model (P. C. Austin & Fine, 2017). The CSH model involves fitting separate Cox models for each competing event, focusing solely on the event of interest. In contrast, the Fine-Gray model introduces a sub-distribution hazard that accounts for patients who experience competing events, thereby adjusting the risk set and resolving interpretation issues associated with the CSH approach technique (Fine & Gray, 2015). The CSH represents the instantaneous rate of occurrence for a specific cause among event-free individuals, while the SDH correlates with cumulative incidence and retains patients with competing events in the risk set. This distinction is crucial, as it allows for a more accurate understanding of event probabilities in the presence of competing risks. By employing these frameworks, researchers can enhance their analyses of events of interest and better interpret the complexities involved when competing risks are present. This is in contrast to the cause-specific hazards approach which censors such patients at the time of occurrence of the competing event. However, just like standard survival analysis and the cause-specific hazard approach, the sub-distribution hazard (SDH) can be modeled in a proportional hazards framework using The CIF for the k^{th} failure or cause.

Cumulative Incidence Function

The cumulative incidence function $F_k(t)$, estimating the probability of failing from cause k (loss to follow-up or another competing event) before a given time t , is used instead to provide information for a certain population or to compare a discrete number of subgroups descriptively. The cumulative incidence function can be denoted as:

$$CIF(t) = F_k(t) = P(T \leq t, D = k), for\ t > 0. \quad \textbf{(8)}$$

Where D: is the type of event that happened from the possible event at survival time t (in our case lost follow up or death). In standard survival analysis, we know that the survival function is given by equation (1) and the incidence of the event throughout follow up is given as equation (2). This shows that, in contrast to equation (2). The function $F_k(t)$ denotes the probability of experiencing the k^{th} event before time t (k =1 lost follow up, k =2 death due to BC) and before the occurrence of a different type of event. The cumulative incidence function (CIF) enables estimation of the incidence of an event's occurrence (LTFU) while accounting for competing risk. This makes it possible to calculate incidence in a population where clinical decision-making needs to take into account all competing occurrences.

3.9 Competing Risks Regression model

In survival analysis with competing risks, two regression modeling approaches depend on the above mentioned hazard functions: the cause-specific hazards model and the Sub-distribution hazard model also called the Fine-Gray model. The cause-specific hazard model is used to estimate the effect of the covariates on the rate of occurrence of the outcome in those subjects who are currently event free. In contrast, the Fine-Gray Model allows us to estimate the effect of covariates on the absolute risk of the outcome over time Fine-Gray (Fine & Gray, 2015).

3.9.1 Cause-Specific Hazard Model

The Kaplan-Meier estimator is a non-parametric estimator used to estimate the survival function with censoring, which is not based on the actual observed event and censoring times, but rather on the order in which events and censored observations occur. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored

times (Kaplan & Meier, 1958). Therefore, the Kaplan-Meier estimate of the survival function at time t , $S\ (t),$ is given by:

$$\hat{S}(t) = \prod_{j:r_{j\leq t}} (1 - \frac{d_j}{r_j}). \tag{19}$$

$\hat{S}(t)$ denote the **estimated survival probability** at time t, r_j is the number of subjects at risk group (patients of breast cancer under follow up and at risk of experiencing LTFU or other competing event were at risk group) just before time t_j (the j^{th} ordered survival time) and d_j denotes the number who were lost follow up at time t_j .

The Kaplan-Meier plots are used to see whether there is a difference in survival times or not between groups of covariates under investigation. But, the plot cannot be used to decide whether the survival time of loss to follow up breast cancer patients in each covariate is different or not, and the Log Rank test was used for this purpose(Mantel & Haenszel, 1959). Test statistic for the Log Rank test is given by:

$$X^2{}_{Logrank} = \frac{\left(\Sigma\left((d_{0j} - r_{0j}\frac{d_j}{r_j})\right)^2\right)}{\left(\Sigma\ \frac{r_{0j}r_{1j}d_j(r_j - d_j)}{r_j^2(1-r_j)}\right)} \sim X^2{}_{(1)} \tag{10}$$

Where: $X^2{}_{Logrank}$ represents the log-rank test statistic, which is used to compare the survival distributions of two groups. d_{0j} is the Observed number of events (LTFU or death) in j^{th} time of 1^{st} group, d_{1j} is the number of event in j^{th} time of 2^{nd} group, d_j is the total number of event(LTFU ,death) across all groups in j^{th}time $(d_{0j} + d_{1j})$, r_{0j} is the number at risk at j^{th} time of 1^{st} group, r_{1j} is the number at risk at j^{th}time of 2^{nd} group and r_j is the total number of at risk cross all groups at j^{th} time $(r_{0j} + r_{1j})$ in our case at risk group were breast cancer patients under follow up who are more likely to experience the event of interest (LTFU) at Hawassa university oncology center.

The hypotheses to be tested are:

H_0 : - There is no difference between the survival curves.

H_1 : - There is difference between the survival curves.

Indicates there is a significant difference in the time to loss to follow-up among breast cancer patients when accounting for competing risks, such as death from breast cancer or other causes.

In the absence of competing risks the cumulative incidence function equals $1 - S(t)$, where $S(t)$ is the survivor function, which can be derived by the Kaplan-Meier estimator. In the presence of competing risks, the cumulative incidence function in a population or in subgroups of interest can be estimated as

(11)

$$\hat{F}_k(t) = \sum_{i:t_i \leq t} = \hat{\lambda}_k(t)\, \hat{S}(t_i - 1)$$

In standard survival analysis, we know that the survival function is given by equation (1) and the incidence of the event throughout follow up is given as equation (2). From this, we can see that the cumulative incidence function is distinct from equation (2) which allows estimation of the incidence of the occurrence of an event while taking competing risk into account. This allows one to estimate incidence in a population where all competing events must be accounted for in clinical decision making. One key idea is that, given a system with competing risks, only one kind of event is allowed to happen and the occurrence of that type of event prevents the occurrence of other types of events in the future (Austin et al., 2016).

Here $\widehat{\boldsymbol{S}}(\boldsymbol{t})$ is the estimator for the overall survivor function at time $\boldsymbol{t}$ including all types of event and t_i denotes the i^{th}ordered event time $\hat{S}(t) - 1$ the estimated survival probability just before time t_i . $\boldsymbol{\lambda_k(t)}$ Is defined as the cause-specific hazard rate which is given as

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}, k = 1, 2, \ldots, D \qquad \textbf{(12)}$$

And can be estimated by

$$\lambda_k(t) = \frac{d_{ki}}{n_i}, \qquad \textbf{(13)}$$

Where $\boldsymbol{d_{ki}}$ is the number of failures from type *k* at time $\boldsymbol{t_i}$ and $\boldsymbol{n_i}$ the risk set at time $\boldsymbol{t_i}$, i.e. the number of BC patients who were not censored and have not failed from any cause up to time $\boldsymbol{t_i}$.

The cause-specific hazard (CSH) model assesses how different covariates affect the rate at which an event occurs in individuals who have not yet experienced that event. When dealing with continuous covariates or interested in the simultaneous effects of multiple covariates on cause-specific failure, the competing risks variant of the Cox proportional hazards model is typically the best option. Prentice et al., proposed a method to estimate the effects of covariates on the cause-specific hazard rate. In this modeling approach, each cause-specific hazard is analyzed independently, treating individuals who fail from other causes as censored observations. This method is appropriate for identifying which factors influence the event occurrence rate. Traditional regression analysis of competing risks applies a Cox proportional hazards (CPH) model (Prentice et al., 1978) for each cause-specific hazard (CSH). The CSH of cause *k* for a subject with covariate vector $\boldsymbol{X}$ is modeled as follows:

$$\lambda_k(t|X) = \lambda_{k,0}(t)exp(\beta_k^T X)\,,. \quad k = 1,\ldots,D \qquad \textbf{(14)}$$

Where X is a $p \times 1$ vector of covariates, β_k is a $p \times 1$ vector of unknown parameters to be estimated for each outcome variable, k is the cause of failure at time t, and $\boldsymbol{\lambda_{k,0}(t)}$ is the baseline CSH of cause *k*, β_k is a vector of unknown parameters to be estimated for each outcome variable, k is the cause of failure at time t. Examination of conflicting hazards With statistically standard software packages, one may do Cox regression on data based on the CSH by applying classical Cox regression, which treats failures from the cause of interest as occurrences and failures from other causes as censored observations.

3.9.2 Sub-distribution Hazard Regression-the Fine-Gray Model

Gray's test is an extension of the Log Rank test which is used to evaluate the CIF similarity between two or more groups. The concept of Gray's test is to compare hazard sub-distribution functions.

The Fine and Gray model allows us to estimate the effect of covariates on the absolute risk of the outcome over time (Fine & Gray, 2015). To overcome some of the problems occurring with CSH regression as described above, Fine and Gray developed a regression model that directly links the regression coefficients with the CIF (Fine & Gray, 2015). In the Fine and Gray model

the association between the so called SDH introduced earlier by Gray (Gray, R. J. 1988) and covariates of interest is assessed infinitesimal small time interval Δt, given the subject experienced no event until time t or experienced an event other than k before time t.

$$\lambda_k^*(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t, D = k | T \ge t \cup (T \le t \cap D \neq k))}{\Delta t} \quad \textbf{(15)}$$

Individuals failing before time *t*, but not from the cause of interest, remain in the risk set for all future failure times. When modeling the likelihood of a certain event occurring in individuals who have not yet experienced it, the cumulative incidence function is employed. It represents the immediate failure risk resulting from the k^{th} event in individuals who have not yet encountered a type *k* event. It represents the immediate failure risk resulting from the k^{th} event in those who have not yet encountered a type *k* event. The risk sets are the fundamental distinction between the two.

The Cox PH model is frequently used when analyzing data on competing hazards. Fine and Gray (Fine & Gray, 2015) created a CIF based proportional hazard model to assess survival data in the context of competing hazards, as this technique has significant shortcomings. A hazard function in the presence of covariates is taken into consideration in the competing risks setup for each cause, the occurrence of an event of interest considered (Mohammad et al., 2017). Knowing which factors influence an event's likelihood of happening over time is helpful. Additionally, the sub-distribution hazard ratio derived from the Fine-Gray model shows how factors affect the cumulative incidence function or sub-distribution hazard function in relation to one another (P. C. Austin & Fine, 2017). The sub-distribution hazard model is also used to determine factors associated with the incidence of a given cause. This method of analysis does not treat individuals failing from other causes as censored observations. In the presence of competing risks, the choice of competing risk model depends on the objective of the study.

Under the proportional hazard, Fine-Gray Model can be specified (Leoce, 2016) as

$$\lambda_k^*(t|X) = \lambda_{k,0}^*(t) exp(\beta_k^{*T} X), \quad \textbf{(14b)}$$

Where λ_k^*(t) denotes the baseline SDH function (*i.e*. the sub-distribution hazard for a (fictitious) individual with all covariates set to zero). SDH rates are assumed to be proportional for the included covariates.

The CIF for the event of interest can be estimated directly from the regression coefficients obtained by a Fine and Gray model without explicit consideration of the covariate effects on competing events. In medical research, regression models for the CSH and the SDH are most commonly utilized for the study of competing risk data. Beyersmann et al., 2009 compare competing risk analyses using CSH and SDH in a real data example (Beyersmann et al., 2009; Lakshmanan, 2015). In the presence of two possible types of failure, the relationship can be

$$\lambda_1(t/X) = \left(1 + \frac{F_2(t/X)}{S(t/X)}\right).\lambda_1^*(\frac{t}{X}), \quad \textbf{(16)}$$

Where $S(t|X)$ denotes the probability of being free of any event up to time *t* given X. For each variable the relative difference (RD) of the hazard ratios was computed as(Van Der Pas et al., 2018):

$$RD = \frac{(SDH_{ratio} - CSH_{ratio})}{CSH_{ratio}} \times 100\% \quad (\textbf{17})$$

Therefore, the advantages of the sub-distribution hazard model is to maintain individuals who develop the competing event in the risk set, while in the cause-specific hazard model it is considered as censored. Furthermore, in the presence of competing risks, the choice of a competing risk model depends on the objective of the study. Scholars suggested that use the sub-distribution hazard model when the focus is on estimating incidence or predicting prognosis in the presence of competing risks and use the cause-specific hazard model when the focus is on addressing etiologic questions (Austin et al., 2016). . In this thesis, we have applied the CSH model and Fine-Gray modeling approaches to identify the potential factors of the loss to follow up treatment of BC patients when death from any cause as a competing risk event. Using the relationship between the survival, hazard and cumulative incidence function (Leoce, 2016) in equation (5). The sub-distribution hazard (hazard of the cumulative incidence) for each cause supposed as the hazard for an individual who either fails from cause k or does not, can be written as

$$\lambda_k{}^*(t;\ X)\ = \frac{f_k(t)}{1 - F_k(t)}\ \ . \qquad \textbf{(18)}$$

Under the proportional hazard, Fine-Gray Model can be specified (Leoce, 2016) as

$$\lambda_k^*(t/X)\ =\ \lambda_{k,0}^*(t)exp(X\beta_k). \qquad \textbf{(19)}$$

Where $\lambda_{0k}{}^*(t)$ is the baseline sub-distribution hazard for the cause of *k* and e^{β_k} is the relative risk probability of k^{th} cause associated with the given covariates.

3.10 Methods of Parameter Estimation

In order to have an estimate of the coefficient vector, we build the partial likelihood, using the same procedure available in the univariate Cox proportional hazard model. Let $0 < t_{1k} < t_{2k} <$ ... t_{uk} be ordered distinct time points at which failures of any causes occur for the risk *k*. Assume that only one failure can happen at each failure time. *I.e*. there are no tied failure times in the data. The partial likelihood for specific hazard k is given by(Cox, 1975).

$$L_k(\beta)_k = \prod_{k=1}^{n_k} \frac{\exp(\beta_k^T X_{ik}(t_{ik}))}{\sum_{I \in R} \exp(\beta_k^T X_{ik}(t_{ik}))} \qquad (20)$$

Where n_k is the number of individuals in specific hazard *k*, X_{ik} is a vector of covariates for individuals ***i*** specific to *k*-type risk at time t, the vector β_k represents the regression coefficients of cause *k* to be estimated. Since the same variables could have different effects on the different risks, it is reasonable to assume that for each *k*, β_k is independent of each other and $R(t_{ik}) = \{\frac{l}{t_{lk}} \geq t_{ik}\}$ is the set of individuals at risk at timet_{ik}. The overall partial likelihood function of CSH model is:

$$L(\beta) = \prod_{k=1}^{K} L_k(\beta_k) \qquad (21)$$

Estimation of parameters in the Fine-Gray model uses the partial likelihood approach similar to the standard Cox model since a proportional hazard assumption is imposed on the sub-distribution hazards (Pintilie, 2006) . However, in this model, the parameters are estimated by incorporating weights in the partial likelihood. The partial likelihood for the Fine-Gray model is given (Kuk et al., 2013) as:

$$L(\beta) = \prod_{i=1}^{n} \frac{\boldsymbol{exp}(\boldsymbol{\beta}^T \boldsymbol{X_i})}{\sum_{\boldsymbol{i} \in \boldsymbol{R}^*(\boldsymbol{t_k}} w_{il} \boldsymbol{exp}(\boldsymbol{\beta}^T \boldsymbol{X_i})} \quad \textbf{(22)}$$

Where: $\hat{\beta}_1\ \hat{\beta}_2 \ldots \hat{\beta}_p$: regression parameter estimation event type *k*

$X_{i1}\ X_{i2} \ldots X_{ip}$: Predictor variable object ***i*** event type ***k***

A collection of risks $R_{(ti)}$ defined by adding weights, which is based on non-censored objects It can be concluded that collection of risks is formed based on the event of interest object with the following definition: $R(t_i) = l;\ \mathrm{T}_l \geq \mathrm{t}$ or $(\mathrm{T}_l \leq t)$; and an individual in a competing risk w_{il} = the weight of the object *i* in the event of interest, the weight can be defined as follows:

$$W_{ki} = \frac{\hat{G}(t_k)}{\hat{G}(min(t_i, t_k))} \quad . \quad \textbf{(23)}$$

Where: $\widehat{\boldsymbol{G}}(\boldsymbol{t_i})$ = estimated of survival function from survival time ***i*** object

$\widehat{\boldsymbol{G}}(\boldsymbol{min}(\boldsymbol{t_i}, \boldsymbol{t_l})$ = estimation of the survival function from the minimum value between the survival time of the object ***i*** and the object on the event ***l*** or event of interest. The weight were 1 if the object has a survival time $t_i < t_l$ and has a value ≤ 1 if $t_l > t_i$ (Pintilie, 2006).

As a result, individuals who experience a competing event at time t_i do not participate fully in the partial likelihood; the further the time point (t_k) is from the time of the competing event (t_i) the smaller the weight. When there is only one event of interest, the weights are all equal to 1, and the risk set contains only those at risk at the specified time point (Kuk et al., 2013). Once the likelihood is formulated, the goal is to choose the values of parameters that maximize the likelihood.

3.11 Model Diagnosis

The main assumption when modeling survival data is the proportionality of hazards. When the Fine-Gray model is used, the hazards of the CIF must be proportional whereas, in the Cox proportional hazard model, it is the cause-specific hazards that need to be proportional (Pintilie, 2006). The proportionality assumption is the most common in competing risk regression model, which considers the sub-distribution with covariates $\boldsymbol{X}$ is a constant shift on the complementary log-log scale from a baseline sub-distribution function. If the curves do not cross with each other then we say that the model does not violate the assumption of proportionality (Zhang, 2017).

3.11.1 Proportionality of the cause-specific hazards

A visual inspection of the plot of *log (−log(S))* vs. *log (time)* can give a hint as to whether the cause-specific hazards can be assumed to be proportional. In this case, S is the Kaplan–Meier estimate when the only event considered in the event of interest; both the observations without an event, as well as the competing risks, are censored. Such a plot can be drawn for each of the levels of a covariate (Pintilie, 2006).

3.11.2 Proportionality of the hazards of the CIF

To investigate the proportionality assumption for the competing risk regression, *log(−log(1−CIF))* can be plotted against log (time) where F is the CIF for the event of interest (Pintilie, 2006). In this study, the researcher is going to apply the Fine-Gray model and check the assumption of proportionality assumption of the CIF. In this thesis, the CSH model and the SDH-Fine and Gray model were applied. We have checked the assumption of proportionality for both models.

Statistical Software Used

The competing risk data was analyzed using the survival package and $cmprsk$ package of the R statistical software (version 4.3.3).

CHAPTER FOUR
RESULT AND DISCUSSION

4.1. Descriptive Results of the Data

4.1.1 Socio-demographic characteristics of the study participants

The data utilized in this research was extracted from a cohort of 463 individuals who underwent Breast Cancer treatment from January 1, 2020, to December 30, 2023 at HUCSH, located in Hawassa, Ethiopia. Detailed accounts of the demographic and clinical attributes of the patients included in the breast cancer dataset can be found in Tables 3 and 4.

From a total of 463 patients included in the study 96 (20.73%) lost to follow up, 121 (26.13%) death and 246 (53.13%) censored at the end of the study. The median time to LTFU was 5 months, while the minimum and maximum observed events time were 3 and 6 months, respectively. From these BC patients 181 (39.09%) and 282(60.91%) receded Urban and Rural respectively.

Looking at the age of patients, more than one-third of BC patients fall under the age category 50 to 59, while less than 10% were between ages of 0-39 years. About 44(9.5%) of patients were aged ≤ 39, 105(22.68%) were age Between 40 to 49, 157(33.91%) patients were at the age of 50 to 59 years, 116(25.05%) were at the age of 60 to 69 and 41(8.86%) were aged 70 or older. Median time to loss to follow-up was 4 for age groups. Regarding hospital distance, 164 patients (35.42%) lived within 100 km, while 299 (64.58%) lived 100 km or more away. The median time to loss to follow-up was shorter 3 months for those 100 km or farther, compared to 5 months for those closer than 100 km. The mean and median age of patients was 52.74 and 53 years (Interquartile range: 50-59) respectively. Of the 463 study participants, 182 (39.31%) and were married, 110(23.76%) were divorced, 98(21.17%) were widowed and 73(15.77%) were single. Proportions are used to express the prevalence or incidence, the proportion of patients under age 39 in the censored group was 0.48, indicating that 48% of this group falls within this age category.

Table 3: Descriptive Result of Socio-demographic characteristics of the study participants

Variables	Categories	Censored	LTFU	Death	Total
Age	≤ 39	21(47.7%,0.48)	11(25%,0.25)	12(27.3%,0.27)	44(9.5%)
	$40-49$	61(58.1%,0.58)	20(19.1%,0.19)	24(22.8%,0.23)	105(22.68%)
	$50-59$	79(50.32%,0.5)	32(20.4%,0.20)	46(29.3%,0.29)	157(33.91%)
	$60-69$	68(58.62%0.59)	26(22.4%,0.22)	22(18.9%,0.19)	116(25.05%)
	≥ 70	17(41.5%,0.41)	7(17.1%,0.17)	17(41.5%,0.41)	41(8.86%)
Residence	Urban	126(69.61%,0.7)	31(17.1%,0.17)	24(13.3%,0.13)	181(39.09%)
	Rural	120(52.5%,0.43)	65(23.1%,0.23)	97(34.4%,0.34)	282(60.91%)
Educational level	Second and above(Ref)	68(44.75%,0.45)	37(24.3%,0.24)	47(30.9%,0.31)	152(32.83%)
	Primary and below	178(57.2%,0.57)	59(19%,0.19)	74(23.8%0.24)	311(67.17%)
Distance from hospital	$\geq 100KM$	154(51.5%,0.52)	63(21.1%,0.21)	82(27.4%,0.27)	299(64.58%)
	$< 100KM$	92(56.10%,0.56)	33(20.12%,0.2)	39(23.8%,0.24)	164(35.42%)
Marital Status	Single	33(45.2%,0.45)	17(23.3%,0.23)	23(31.5%,0.32)	73(15.77%)
	Married	122(67.1%,0.67)	29(15.9%,0.16)	31(17.0%,0.17)	182(39.31%)
	Divorce	46(41.8%,0.42)	25(22.7%,0.23)	39(35.5%,0.35)	110(23.76%)
	Widowed	45(45.9%,0.46)	25(25.5%,0.26)	28(28.6%,0.29)	98(21.17%)

4.1.2 Baseline clinic-pathologic and treatment characteristics of the study participants

From the total study participants, 68 (14.69%), 118 (25.49%), 174 (37.58%) and 103 (22.25%) of patients were at stage1, stage2, stage3 and stage4 respectively. The median time to loss to follow up among patients who were at stage 1 was 6 months and for stage 2 cancer patients while it was 5 months for stage 3 and 4 breast cancer stage. Furthermore, 179 (38.66%) experienced weight loss, and 197 patients (42.55%) had non-communicable diseases (NCDs) chronic illnesses that are not transmissible. The median time to loss to follow-up for these groups was 6 months. In terms of histologic grading, 153 patients (33.05%) had moderately differentiated tumors (grade II), while 182 (39.31%) had high-grade (grade III) tumors at diagnosis. For treatment types, 154 patients (33.26%) received both curative chemotherapy and radiotherapy aimed at completely eradicating the cancer, 116 (25.05%) underwent curative chemotherapy only, 106 (22.89%) received curative radiotherapy alone, and 87 patients

(18.79%) were either without recorded curative treatment or receiving palliative care to improve quality of life in advanced stages.

Among 357(77.1%) Experienced chemotherapy patients 184 (51.5%) were censored, 86 (24.1%) were lost to follow up and 87 (24.4%) were death with median time to loss to follow up 5 months and 6 months for patients who do not Experienced chemotherapy toxicity and who experienced chemotherapy toxicity respectively. Similarly, 373 (80.56%) of patients Experienced Radiotherapy among those 206 (55.2%) were censored, 68 (18.2%) were lost to follow up while 99 (26.5%) were death and their median time to loss to follow up was 4 and 6 months for those patients who do not Experienced radiotherapy toxicity and those patients who Experienced Radiotherapy toxicity respectively. As well incase 216(46.65%)of experienced advert effect of BC patients 106(49.1%) were censored, 62(28.7%) were loss to follow up and 48(22.2%) were death and their median time to loss to follow up was 6 months for both Experienced Advert affect at follow up and who did not Experienced Advert effect at follow up.

For the performance ECOG (a measure used by healthcare providers to assess a patient's functional ability and overall well-being in relation to their cancer diagnosis and treatment) 226(48.81%) patients were categorized under 0 to 1, 237(51.19%) were categorized under ≥ 2 and their median time to loss to follow up was 5 months and 6 months for those patients whose performance status is 0 to 1 and ≥ 2 respectively. The minimum and the maximum time to loss to follow up BC treatment were 3 and 6 months respectively. Proportion refers to the relationship between two quantities, often expressed as a fraction or percentage. A proportion value of performance ECOG which is censored is 0.81 indicates that one quantity is 81% of another, suggesting a significant presence or association.

Table 4: Descriptive Results of clinical and treatment characteristics

Variables	**Categories**	**Censored**	**LTFU**	**Death**	**Total**
Weight	$< 55KG$	95(53.1%,0.53)	39(21.2%,0.22)	45(25.1%,0.25)	179(38.66%)
	$\geq 55KG$	151(53.2%,0.53)	57(20.1%0.20)	76(26.7%,0.27)	284(61.34%)
Performance	0-1	184(81.4%,0.81)	15(6.6%,0.07)	27(12%,0.12)	226(48.81%)
status(ECOG)	2+	62(26.2%,0.26)	81(34.2%,0.34)	94(39.7%,0.39)	237(51.19%)

Has NCD	No	134(50.4%,0.50)	58(21.8%,0.22)	74(27.8%,0.28)	266(57.45%)
comorbidities	Yes	112(56.85%,0.57)	38(19.3%,0.19)	47(23.9%,0.24)	197(42.55%)
Cancer stages	Stage 1	40(58.8%,0.59)	16(23.5%,0.24)	12(17.7%,0.18)	68(14.69%)
	Stage 2	62(52.5%,0.53)	26(22.0%,0.22)	30(25.5%,0.25)	118(25.49%)
	Stage 3	86(49.4%,0.49)	37(21.3%,0.21)	51(29.3%,0.29)	174(37.58%)
	Stage 4	58(56.3%,0.56)	17(16.5%,0.17)	28(27.2%,0.27)	103(22.25%)
Type of	Both(C&R)	84(54.55%,0.55)	33(21.4%,0.21)	37(24.0%,0.24)	154(33.26%)
treatment	Chemo only	59(50.86%,0.51)	33(28.5%,0.28)	24(20.7%,0.21)	116(25.05%)
	Radio only	75(70.75%,0.71)	12(11.3%,0.11)	19(17.9%,0.18)	106(22.89%)
	No treat rec	28(32.2%,0.32)	18(20.7%,0.21)	41(47.1%,0.47)	87(18.79%)
Experienced	No	62(58.5%,0.58)	10(9.40%,0.22)	34(32.1%,0.32)	106(22.89%)
Chemotherapy	Yes	184(51.5%,0.52)	86(24.1%,0.24)	87(24.4%,0.24)	357(77.11%)
toxicity					
Experienced	No	40(44.44%,0.44)	28(31.1%,0.31)	22(24.5%,0.24)	90(19.44%)
Radiotherapy	Yes	206(55.2%,0.55)	68(18.2%,0.18)	99(26.5%,0.27)	373(80.56%)
toxicity					
Expected	No	140(56.7%,0.57)	34(13.8%,0.14)	73(29.5%,0.30)	247(53.35%)
adverse effect	Yes	106(49.1%,0.49)	62(28.7%,0.29)	48(22.2%,0.22)	216(46.65%)
Histologic	Grade I	79(61.7%,0.62)	23(18%,0.18)	26(20.3%,0.20)	128(27.65%)
grade	Grade II	82(53.6%,0.54)	30(19.6%,0.20)	41(26.8%,0.27)	153(33.05%)
	Grade III	85(46.7%,0.47)	43(23.6%,0.24)	54(29.7%,0.30)	182(39.31%)

4.2 Factors associated with Lost to follow up among patients

To statistically characterize the survival data with competing risk, CIFs can be used for several reasons why a study may fail. Since the competing risk event is treated as censored in scenarios with competing risks, the KM approach may produce estimates that are biased in such cases. Therefore, in order to determine the statistical significance of the categorical factors linked to LTFU, the Chi-square (Log rank) test was employed first. The results are given in **Table 5** and the KM estimate survival curves are given in **Figure 6**. The Chi-square analysis revealed that age, place of residence, distance from Hospital, marital status, cancer stage, having non communicable disease, performance status (ECOG), experienced chemotherapy, experienced

radiotherapy, type of treatment, expected adverse effect and histologic grade were significantly associated with LTFU (P-value <0.05) according to the study.

Table 5 : Log rank test and univariate analysis of competing risk model (Gray's test)

Variables	Categories	Log-rank	P-value	Gray's test	P-value	df
Age	≤39	18.4	0.001	21.43	0.0002	4
	40-49					
	50-59					
	60-69					
	≥70					
Weight	<55kg	8.1	0.1	9.6	0.5	1
	≥55kg					
Residence	Urban	12.1	0.0005	38.07	<0.0001	1
	Rural					
Distance	<100KM	5.5	0.02	15.9	0.00006	1
	≥100KM					
Marital status	Single	8.7	0.03	9.32	0.025	3
	Married					
	Divorce					
	Widowed					
Educational status	Secondary& above(Ref)	12.6	0. 4	19.31	0.09	1
	Primary and below					
Cancer stage	Stage I	8.1	0.00001	10.3	0.00001	4
	Stage II					
	Stage III					
	Stage IV					
Performance (ECOG)	0-1	12.1	0.0005	7.75	0.005	1
	2+					
Has NCD	No	12.3	0.0004	8.2	0.004	1
	Yes					
Exp. Chemo	No	9.4	0.002	7.5	0.005	1
	Yes					
Exp. Radio	No	6.6	0.01	4.01	0.04	1
	Yes					
Treatment type	No treat	26.7	0.008	8.2	0.04	3
	Chemo only					
	Radio only					
	chemo and radio					
Exp. adverse effect	No	11.1	0.0009	7.2	0.007	1
	Yes					
Histologic	Grade I					

grade	Grade II Grade III	13.1	0.004	12.9	0.002	2

In addition, the connection between each putative prognostic factor and the study's outcome variable was evaluated using Gray's test, commonly known as the modified chi-square test. The findings are shown in ***Table 5***. It is essential to examine the cumulative incidence curves between various groups using Gray's test as the CIFs for various causes of failure offer more information about the available survival data.

Thus, the Gray's test results showed difference between the cumulative incidence of groups such as age, place of residence, types of treatment, stage of cancer, has NCD, performance status (ECOG), marital status ,distance from hospital, experienced chemotherapy, experienced radiotherapy, expected adverse effect and histologic grade with LTFU. Weight and educational status were not significant.

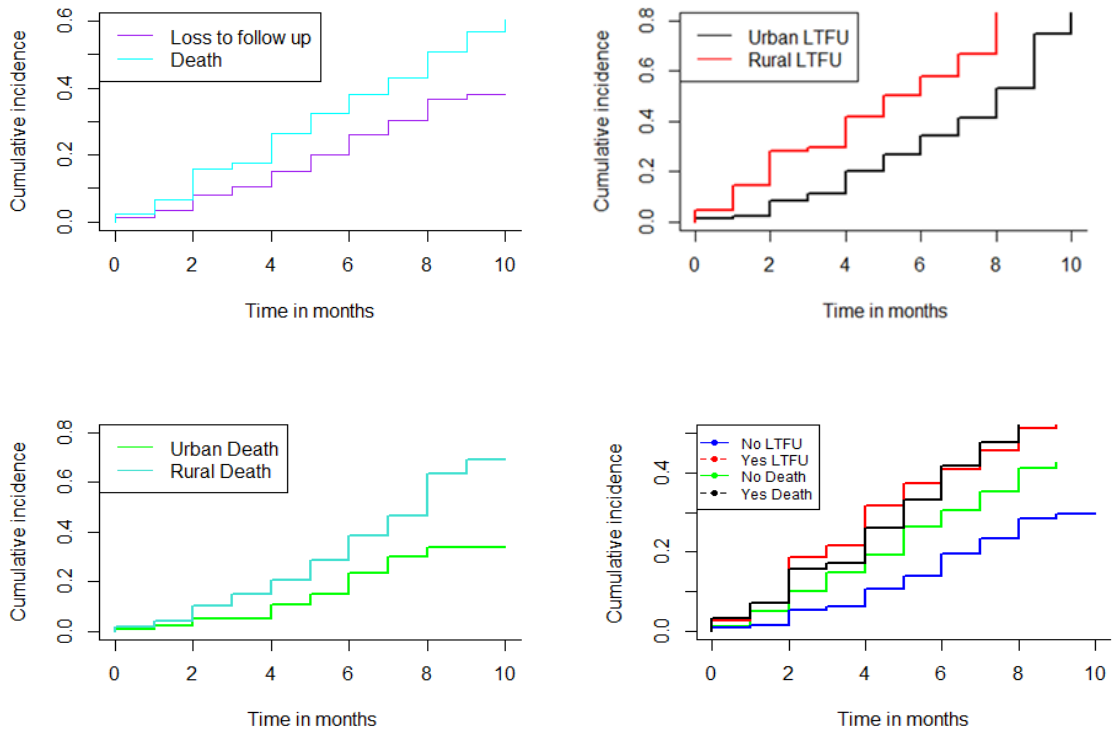


Figure 2: Cumulative incidence functions for LTFU and death (a), LTFU residence differences (b), death incidences (c) and has non communicable disease with LTFU and death (d).

Furthermore, non-parametric estimates of the cumulative incidence functions for LTFU and death events for the whole study population are given in **Figure 2** for the outcome variables LTFU and death in (a), comparison of urban and rural LTFU patients in (b) and comparison of urban and rural death incidences in (c) and comparison of patient has NCD with LTFU and death in (d). From the plots in **Figure 2**(a), the estimated probability of LTFU patients in the first ten months was below 40%, while the estimated probability of dying in the first ten months was below 30%.

Considering the residence categories from **Figure 2** (b), the estimated probability of LTFU in the first ten months is higher for rural patients when compared with urban patients. Similarly, the estimated probability of dying in the first ten months is higher for rural patients when compared with urban patients **Figure 2** (c), the estimated probability of LTFU and death with NCD in the first ten months is higher for NCD when compared with NOT having NCD patients.

Figure 3 illustrates the non-parametric cumulative incidence curves for death and loss to follow-up (LTFU), treating them as competing risk events. These curves are particularly useful for comparing variable categories, including patient residence, presence of comorbidities such as NCDs, distance from hospital, and experience with chemotherapy. By visually analyzing these incidence curves, we can gain insights into how these factors potentially influence patient outcomes within the framework of competing events.

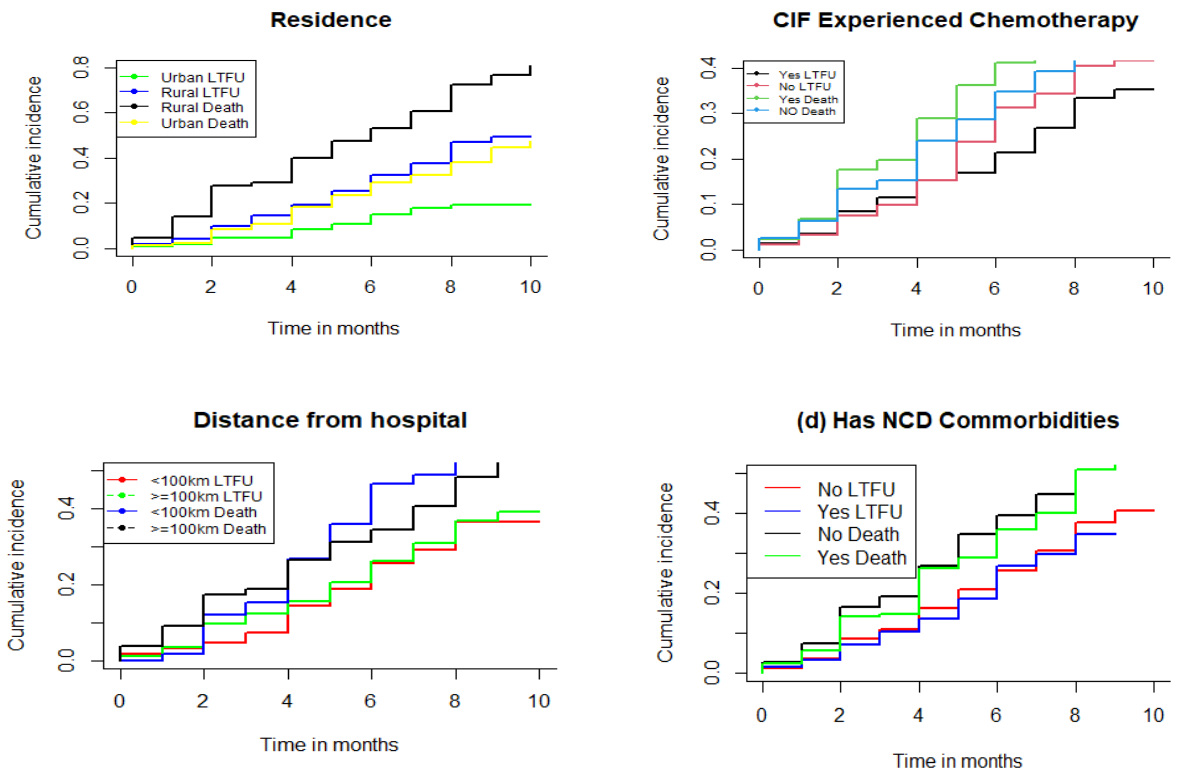


Figure 3: The non-parametric estimates of the cumulative incidence curves with LTFU and death as competing events are given in this figure and similarly the plots can be used to compare the categories of selected variables, namely, Residence, Has NCD Comorbidities and Distance from hospital and Experienced Chemotherapy

4.3 Estimation of the risk factors for Lost to follow up

The two presented competing risk methods were applied to a dataset using univariate and multivariable analysis. A cumulative incidence function was used for univariate analysis of each prognostic factor and Gray's test was used to determine the extent of a significant association between the outcome variable and the factors. Gray's test compares the weighted averages of the sub-distribution hazards across groups for the event of interest (Fine & Gray, 2015). Then, all the potential covariates that were significant at 5% level of significance in the univariate analysis obtained from Gray's test were added to the competing-risks model for the multi-variable analysis. From the multivariable analysis, variables with a P-value < 0.05 were selected as significant covariates.

Table 6: Test of assumptions in the cause-specific PH model for LTFU

Variables	Chi-square	Degree of freedom	P-value
Age	**0.2427**	4	0.99
Residence	**0.1795**	2	0.91
Distance	**1.9010**	1	0.17
Performance (EOCG)	**0.1154**	2	0.94
Cancer Stage	**3.5313**	3	0.32
Has NCD	**1.5107**	1	0.22
Type of treatment received	**4.2343**	4	0.38
Experienced chemotherapy toxicity	**0.0217**	1	0.88
Experienced radiotherapy toxicity	**0.1236**	1	0.73
Expected Adverse effect at Follow up	**0.0176**	1	0.89
Marital Status	**6.6300**	3	0.084
Histologic Grade	**0.134**	2	0.91
GLOBAL TEST	**16.1145**	**25**	0.95

4.3.1 Cause Specific Hazard Model

The effect of each variable on the cause-specific hazard adjusted for other variables was analyzed. The function $\boldsymbol{cox.zph}()$ in the survival package provides a convenient solution to test the proportional hazards assumption for each covariate included in a cause-specific hazard model fit. The $\boldsymbol{cox.zph}()$ investigation is not statistically significant for any of the covariates, and the global test is also not statistically significant (see **Table 6**). Furthermore, the plots of $log\ (-log\ (S))\ vs\ log\ (time)$ were also made to check the assumption of cause-specific PH for the event of interest, and the plots are presented as **Figure 4 and Figure 7**. The assumption revealed no evidence against proportionality for LTFU.

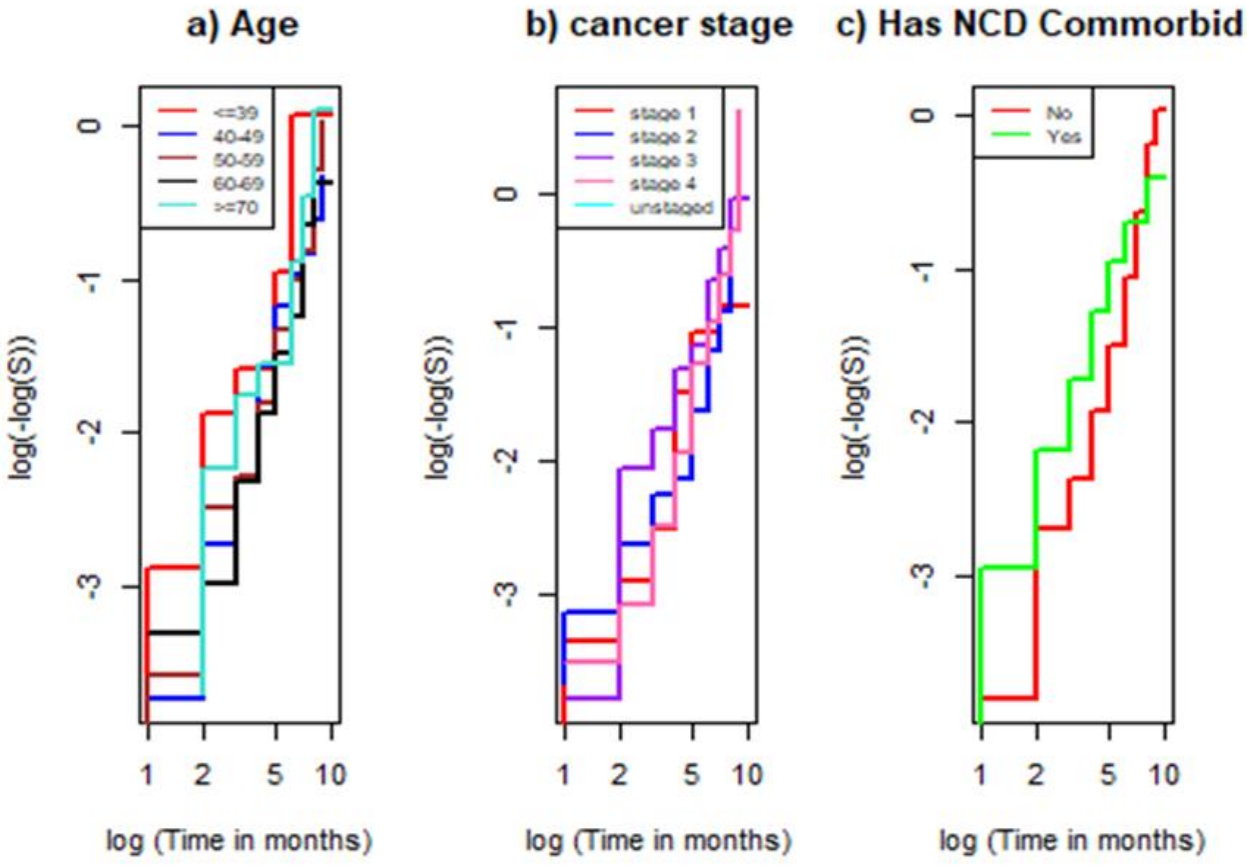


Figure 4: Cause-specific hazard proportionality for Age, Cancer stage, and has NCD comorbidities

The factors that were statistically significant in the Gray's test were added to the CSH model for the multi-variable analyses. In order to quantify the impact of the variables on the CSHs using the R function ***coxph*** from the library survival, the final model under the CSH using the Cox regression model was fitted using multi-variable for LTFU. ***Table 7*** presents the findings for the multi-variable CSH model. The multi-variable CSH model indicated that Age category 50-59, 60-69 and 70 and older, residence, distance, cancer stage 3, cancer stage 4, performance ECOG, has NCD comorbidities, marital status: married and divorce, experienced chemotherapy, experienced Radiotherapy, experienced advert effect at follow up, treatment type curative chemotherapy and radiotherapy, radiotherapy only and histologic grade III had a significant effect on the cause-specific hazards for LTFU. However, according to the CSH model Age category 40-49, cancer stage 2, marital status (widowed), experienced

chemotherapy and treatment type, curative chemotherapy only and histologic grade II were not significant at the 5% level of significance.

In order to quantify the impact of the variables on the CSHs using the R function ***coxph*** from the library survival, the final model under the CSH using the Cox regression model was fitted using multi-variable for LTFU. Table 7 presents the findings for the multi-variable CSH model. A distance CSH ratio between the < 100KM and ≥ 100KM Distances equal to 1.9540 (95% CI: 0.623955 to 6.1221) was observed for LTFU. In this case, the hazard ratio of 1.95446 indicates that patients who were at the distance of more than 100KM away from follow up center had a roughly 2 times higher hazard risk of LTFU than patients who were from follow up center less than 100KM. A CSH ratio between the rural and urban residences equal to 1.842 (95% CI: 1.143 to 2.9682) was observed for LTFU. This suggests that patients in rural areas had a roughly 1.842 times higher risk of LTFU than patients in urban areas. Furthermore, experienced radiotherapy CSH ratio between the groups experienced radiotherapy and not experienced radiotherapy equal to 3.9866 (95% CI: 1.637 to 4.3763) was observed for LTFU. In this case, the hazard ratio of 3.9866 indicates that patients who experienced radiotherapy had a roughly 4 times higher hazard risk of LTFU than those who did not experience radiotherapy.

Furthermore, patients who had poor performance have a roughly 1.52 higher risk of LTFU from breast cancer than healthy performance patients (CSHR: 1.521 (95% CI: 0.3280, 1.8304)). The patients who have Non Communicable Disease Comorbidities (CSHR: 2.8589; 95% CI: 0.567, 4.3). This CSHR value of 2.8589 indicates that the hazard rate of LTFU among patients with Non Communicable Disease comorbidities is approximately 2.86 times the hazard rate among patients without Non Communicable Disease comorbidities. This suggests that patients with Non Communicable Disease comorbidities are at a higher risk of experiencing LTFU compared to those without Non Communicable Disease comorbidities. Keeping other variables constant, the CSH of LTFU is 9.72 and 7.68 times higher among the age groups 60–69 years and ≥ 70 years, respectively, when compared with the age group < 39 years and younger. The CSH of LTFU is 2.76 and 3.52 times higher among patients with a Cancer stage 3 and cancer stage 4, respectively, when compared with cancer stage one. Furthermore The CSH of LTFU is 2.01 and 1.45 times higher among the treatment type Curative chemotherapy and Curative radiotherapy treatment, respectively, when compared with no treatment received.

Table 7 : Multivariable analysis of the cause-specific hazard regression model

Variables	Categories	Estimate	SE	P-value	CSHR (95%)
Age	≤39(Ref)				
	40-49	-0.4534	0.39926	0.24724	0.6350[0.2940,1.3717]
	50-59	-0.5384	0.3648	0.05060	3.5837[0.2850,5.1931]
	60-69	0.6490	0.3962	0.03120	9.7225[0.2400,10.360]
	≥70	-0.4128	0.4300	0.04598	7.6820[4.2830,9.5289]
Residence	Urban (Ref)				
	Rural	0.6109	0.2434	0.01353	1.8420[1.1430,2.9682]
Distance	<100KM(Ref)				
	≥100KM	0.2098	1.2330	0.04105	1.9540[0.6230,6.1220]
Marital status	Single (Ref)				
	Married	1.458	1.7530	0.00082	4.72655[0.5288,40.264]
	Divorce	1.223	0.0379	0.00500	12.1107[1.3987,104.78]
	Widowed	1.534	0.589	0.23510	8.77595[1.0032,76.462]
Cancer stage	Stage I(Ref)				
	Stage II	0.3829	0.3393	0.05630	1.4670[0.7544,2.8520]
	Stage III	0.5697	0.3056	0.00850	2.7680[0.9710,3.2174]
	Stage IV	0.4179	0.3377	0.0002	3.5190[0.7830,3.9443]
Performance (ECOG)	0-1 (Ref)				
	2+	-0.6503	0.2728	0.01560	1.5210[0.3280,1.8304]
Has NCD	No (Ref)				
	Yes	-0.1520	0.2117	0.00220	2.8589[0.5670,4.3006]
Exp. Chemo	No (Ref)				
	Yes	0.2822	0.2023	0.22648	1.3960[0.8920,1.9713]
Exp. Radio	No (Ref)				
	Yes	-0.6985	0.1964	0.00488	3.9866[1.637,4.3763]
Treatment type	No treat (Ref)				
	Chemo only	0.6985	0.2850	0.22570	2.0110[0.745,3.5155]
	Radio only	0.3728	0.7707	0.02390	1.4520[1.150,6.5755]
	Chemo & radio	0.3885	0.3102	0.02744	1.4750[0.803,2.7087]
Exp. adverse effect	No (Ref)				
	Yes	-0.2710	0.1950	0.02280	2.7626[1.520,3.1176]
Histologic grade	Grade I(Ref)				
	Grade II	0.3203	0.2370	0.07600	1.3770[0.865,2.1919]
	Grade III	1.2010	0.2486	0.00670	1.1280[0.692,1.8355]

4.3.2 Sub-distribution Hazard Regression-Fine and Gray model

An investigation using the plot of $log\ (-log\ (1-CIF))\ vs\ log\ (time)$ was made to check the assumption of sub-distribution hazard for LTFU. As shown from Figure 5 and Figure 7, the plot of log (– log (1– CIF)) vs log (time) revealed no evidence against the assumption of proportionality for LTFU. To formally check the assumptions for each covariate, we can add an interaction term with time. In the Fine-Gray model with time-varying covariates, Table 8 shows no statistical significance (p value >0.05), indicating the effect of all covariates is time-constant. The assumption revealed no evidence against proportionality for LTFU.

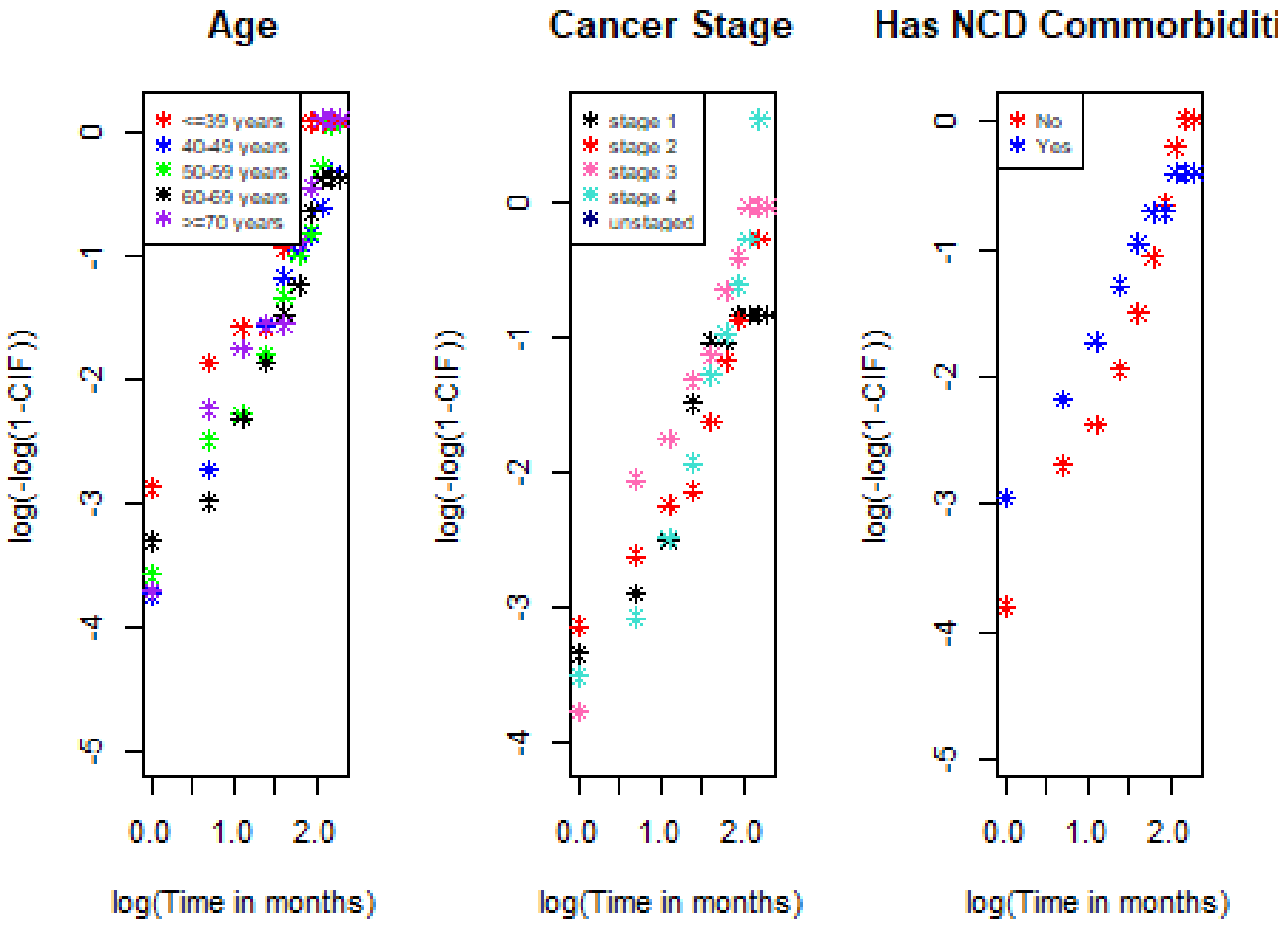


Figure 5: Plot of the proportionality of the hazard of the CIF for Age, Cancer Stage and has NCD

Then, the SDH regression model was fitted using univariate analysis (Gray's test) and the results are depicted in Table 5. The covariates such as age, place of residence, types of Treatment, Stage of cancer, has NCD, performance(ECOG), marital status ,distance from hospital, experienced chemotherapy, experienced radiotherapy, expected adverse effect, histologic grade were found to be significant and included in the multivariable analysis. The final model based on the SDH model, was fitted using multi-variable for LTFU to estimate the effect of the covariates on the cumulative incidence function using the $\boldsymbol{Crr}$ R function from the library $Cmprsk$. The results based on the multivariable analysis are given in Table 8. The covariates Age category 50-59, 60-69 and 70 and older, Residence, Distance, cancer stage 3, Cancer stage 4, Performance ECOG, Has NCD Comorbidities, marital status married and divorce, Experienced Radiotherapy, Experienced Advert Effect at follow up, Treatment type Curative Chemotherapy and radiotherapy and Radiotherapy type, Histologic grade 3 are the major risk factors that leads to LTFU of BC treatment by considering death due to BC cause in a BC patient as a competing risk event.

However, according to the SDH model, age less than 39 and 40-49, experienced chemotherapy, marital status widowed, treatment type curative chemotherapy only, histologic grade II and cancer stage II were not significant at the 5% level of significance. In our study, keeping other variables constant, the sub hazard of LTFU is 2.011 times higher among patients their distance were greater than 100km from hospital as compared to patients their distance were less than 100km from hospital (SDHR = 2.011; 95% CI: 0.671,2.52991). Looking at place of residence, the sub hazard of LTFU is 2.1 times higher among patients whose place of residence is rural compared to those whose residence is urban (SDHR = 2.050; 95% CI: 1.240,3.38482). The sub hazard of LTFU is 1.38 times higher among patients who are experienced chemotherapy compared to those who are not experienced chemotherapy (SDHR = 1.38; 95% CI: 0.939, 2.01628).

The sub hazard of LTFU is 1.84 for patients with NCD comorbidities (SDHR = 1.84; 95% CI: 0.560, 2.27859), than those of patients with no NCD. The age variable category 60-69 years old has SDHR = 8.653 with CI: 7.316, 10.4678. It means that the risk of LTFU in the age group of 60-69 years is 8.653 times higher than those patients in the age group 39 years and younger. The sub hazard of LTFU is 5.857 times higher among patients who are ≥70 years compared to

39 years and younger age groups (SDHR = 5.857; 95% CI: 3.384,7.91185). The sub hazard of LTFU is 2.73 times higher among patients who are 50-59 compared to 39 years and younger age groups (SDHR = 2.725; 95% CI: 0.374, 4.40563).

Keeping other variables constant, BC patients with performance (ECOG=2+) had a 2.404 times higher risk of LTFU than those with a Performance (ECOG=1 (SDHR = 2.404; 95% CI: 0.257, 4.25687). Breast cancer patients with chemotherapy treatment type had a 1.29 higher risk of LTFU than those with no treatment (SDHR =1.29; 95% CI: 1.02, 3.071961). Similarly, patients with a curative radiotherapy had a 1.43 higher risk of LTFU than those with no treatment (SDHR =1.43; 95% CI: 0.451, 4.53328).

Table 8: Results of the sub-distribution hazard regression (Fine and Gray) models for multivariable analysis

Variables	Categories	Estimate	SE	P-value	SDHR (95%)
Age	≤39(Ref)				
	40-49	-0.4026	0.367	0.21120	0.669[0.325,1.37349]
	50-59	-0.3222	0.338	0.04860	2.725[0.374,4.40563]
	60-69	-0..4264	0.369	0.02554	8.653[7.316,10.4678]
	≥70	-0.1540	0.409	0.00007	5.857[3.384,7.91185]
Residence	Urban (Ref)				
	Rural	0.7155	0.257	0.01300	2.050[1.240,3.38482]
Distance	<100KM(Ref)				
	≥100KM	-10.2339	1.074	0.04685	2.011[0.671,2.52991]
Marital status	Single (Ref)				
	Married	-0.29626	0.278	0.029000	2.744[0.4310,4.2800]
	Divorce	0.07131	0.289	0.008000	7.070[3.6220,9.8500]
	Widowed	-020647	0.289	0.470000	2.813[0.462,3.88300]
Cancer stage	Stage I(Ref)				
	Stage II	0.3705	0.339	0.06590	1.45[0.7450,2.81423]
	Stage III	0.6198	0.307	0.00448	1.86[1.020,3.389460]
	Stage IV	0.3211	0.348	0.00022	2.38[0.6970,2.72557]
Performance (ECOG)	0-1 (Ref)				
	2+	-0.9071	0.859	0.01360	2.404[0.257,4.25687]
Has NCD	No (Ref)				
	Yes	-0.1669	0.211	0.00162	1.846[0.560,2.27859]

Exp. Chemo	No (Ref)				
	Yes	0.3189	0.195	0.22544	1.38[0.9390,2.01628]
Exp. Radio	No (Ref)				
	Yes	-0.0261	0.197	0.00401	3.974[1.662,4.43318]
Treatment type	No treat (Ref)				
	Chemo only	0.5866	0.273	0.36730	1.290[1.02,3.071961]
	Radio only	0.3580	0.588	0.02310	1.430[0.451,4.53328]
	Chemo &radio	0.2526	0.255	0.02999	1.190[0.653,2.15775]
Exp. adverse effect	No (Ref)				
	Yes	-0.1174	0.184	0.00010	0.898[0.602,1.27573]
Histologic grade	Grade I(Ref)				
	Grade II	0.1610	0.229	0.53000	1.170[0.749,1.84146]
	Grade III	-0.0236	0.234	0.000047	0.977[0.618,1.54404]

4.3.3 Comparison of Cause-specific and Sub-distribution Hazard Ratios

4.3.3.1 Identifying the difference effects of covariates between cause-specific hazard and sub-distribution hazard model

When modeling time-to-event outcomes in the presence of competing risks, it is recommended to report both cause-specific Cox and Fine-Gray regression results side by side (Grambauer et al., 2010; Latouche et al., 2013). The hazard ratios provide insight into the relative impact of covariates in cause-specific and sub-distribution hazard models, enabling a comparison of the magnitude and direction of these effects across different models (Baccolini et al., 2021 ;Haller, 2014) .

In this study, the cause-specific hazard (CSH) ratio was higher than the sub-distribution hazard (SDH) ratio for certain covariates, including the age categories of 50-59, 60-69, and ≥70 years, advanced cancer stages (Stage 2, Stage 3, and Stage 4), those who had experienced chemotherapy or radiotherapy, marital status categories (Married, Divorced, Widowed), and treatment types involving curative chemotherapy only, curative radiotherapy only, or a combination of both. This suggests that these factors are more strongly associated with the immediate risk of loss to follow-up or death relative to the probability of experiencing either event over time. Conversely, the CSH ratio was lower than the SDH ratio in younger age

categories, specifically for patients aged 39 and below and those aged 40-49, suggesting they may experience better survival distributions compared to their cumulative hazards. Lower CSH ratios were also observed for factors such as residence, distance to care, performance status, and experienced adverse effects, indicating that these may correlate with more favorable survival outcomes.

When the CSH ratio exceeds the SDH ratio for a given covariate, it indicates a higher hazard of loss to follow-up when considering death as a competing event. This suggests that the likelihood of losing patients to follow-up is influenced by the occurrence of death. Conversely, a higher SDH ratio in the competing risks model suggests that death affects the likelihood of loss to follow-up during breast cancer treatment. Notably, the maximum relative difference in hazard ratios observed in this study was 75.3%, as detailed in Table 9.

4.4 Discussion

In this thesis, the two different hazards regression modeling approaches, under a proportional hazards assumption were used. There have been a number of commentaries and illustrations on how these regression models differ from each other and why one needs to be careful in choosing one from another(Andersen et al., 2012; A. Austin et al., 2016; Dignam et al., 2012; Fine & Gray, 1999; Kalbfleisch & Prentice, 2011).

The choice depends on the research question and the specific goals of the analysis (Haller, 2014). Our primary research focus was to identify the risk factors and prognosis of a specific event. Another interest was in describing the impact of covariates on the cumulative incidence of group of covariates, then we used sub-distribution hazard (Fine & Gray, 1999). Therefore a more comprehensive approach may also be considered which involved fitting both models rather than choosing one model over the other for this study (West et al., 2012). Fitting both models would permit a more detailed and complete understanding on the effect of the intervention on the modeling time to LTFU among Breast cancer patients.

According to the results obtained from the CSH and SDH models analysis, distance is significantly associated with time to LTFU. The CSH and SDH for distant patients were 1.2 and 2.01, respectively. These values indicate that the CSH and SDH of time to LTFU breast

cancer treatment for patients whose distance is greater than 100 km was greater than those patients whose distance is less than 100 km. This means, patients that lived closer to HUCSH or distant < 100km were more likely to continue the follow-up while patients from distant places had significantly higher rate of loss to follow-up when compared to patients residing < 100km (Lamont et al., 2023). Longer distance means difficulty in transportation, financial loss, as patient's family member may have to take more time off from work, may be one of the criteria responsible for discouraging patients from attending the follow-up visits for distantly resided patient (Kaku et al., 2008; Lamont et al., 2023; Syed et al., 2013).

Patient who lived closer to HUCSH were more likely to continue the follow-up. Three studies used national data sets to explore transportation barriers to health care access in minorities, the National Health Interview Survey, Medical Expenditure Panel Survey, and Bureau of Transportation Statistics found that 3.6 million people cannot access medical care due to transportation barriers. Those Research, primarily relying on self-reported data, has demonstrated the connection between transportation issues and missed appointments and poor medication adherence and definitive relation to health outcomes (Kaku et al., 2018; Syed et al., 2013; Wallace et al., 2015). Additionally, long travel distances can impose significant financial burdens, including lost income and high travel and accommodation costs, which may discourage patients from routine follow-up care (Bojude et al., 2018).

From disease factors, cancer stage also influenced patients attending for follow-up. In the present analysis, patients with higher stages (i.e. stages III and IV) were more likely to be LTFU. Patients who had cancer stage 3 or 4 at presentation were more likely to be LTFU after the first visit compared to patients who were in stage 1 or 2, although this difference was only borderline statistically significant (p-value: 0.07) (Habinshuti et al., 2020). As the patient grows more dependent, as the disease progresses, more familial support is necessary to keep up with the follow up care. Thus, a sicker patient will need more financial and physical support and resources as well as a more comfortable transportation to the facility to attend the follow-up care. Our finding was consistent with studies conducted in patient and disease related factors associated with lost-to follow-up/drop-outs of Breast Cancer Patients and Also a study at a major Cancer Hospital in South India (Ali et al., 2008; Bojude et al., 2018; Paul et al., 2010).

Similarly, patients with cervical cancer stage 3 or 4 were more likely to be early LTFU (OR: 2.49, 95% CI 0.93, 6.65) as compared to patients with stage 1 or 2 (Habinshuti et al., 2020). Many studies have investigated the association between late stage disease at presentation with LTFU and death (Ali et al., 2008; Bojude et al., 2018; Kantelhardt, Moelle, et al., 2014; Maranga et al., 2013; Mosha et al., 2009; Wabinga et al., 2003).

According to the results obtained based on the CSH and SDH models analysis, performance ECOG is significantly associated with time to LTFU. Our finding was consistent with studies conducted in India that showed patients with worse pre-treatment performance status were at increased risk of LTFU and inconsistent with the study conducted in Nigeria Patients with reduced performance ECOG 2+ were less likely to be LTFU compared to patients with ECOG 0 or 1 (Bojude et al., 2018). The study found that a patient's performance status before treatment significantly influenced their follow-up behavior. Patients with better performance status were more likely and better able to adhere to follow-up care, whereas those with poorer performance status (ECOG 2+ had a higher likelihood of being lost to follow-up (OR=2.4; 95% CI: 1.2-5.0). As mentioned above, a patient with a deteriorating and poor health condition may need more financial and familial support to keep up the follow-up care. Poorer performance status of the patients before treatment increased the likelihood of LFU (OR=2.4; 95% CI: 1.2-5.0) (Paul et al., 2010).

After the treatment began, patients were more likely to be lost to follow up if they were married, divorced or unmarried. Similar trends were also observed in studies where marital status influenced patient delayed reporting of the cervical or breast cancer at RCC (Ali et al., 2018; Kaku et al., 2008). This trend may exist due to lack of support of a partner, didn’t get time for treatment, where a woman of lower economic bracket may be more likely to be dependent on her partner for financial and psychological support (Vallikad, 2006).

According to the CSH and SDH patients with NCD comorbidities have 2.86 and 1.8 times higher LTFU Breast cancer treatment than those with no NCD comorbidities respectively. This is similar with the findings of research undertaken by (Memirie et al., 2018) factors associated with loss-to-follow-up among Breast cancer patients in Rwanda. Further, the CSH and SDH of patients aged 60–69 years are 9.7 and 8.6, respectively, indicating the risks of LTFU for

patients aged 60–69 years are greater risk of LTFU than those ≤ 39 years and younger. Being aged 70 years and older is also associated with an increase in the rate of LTFU (CSHR: 7.682; 95% CI:4.283, 9.5289).These results are in line with the result displayed by cumulative incidence analyses and facilitates the interpretation of the impact of treatment on the cumulative incidence of LTFU (SDHR: 5.857; 95% CI: 3.384, 7.91185). Indeed, a higher rate of LTFU in older patients was associated with a reduced rate of death for 70 years and older age group at the end of the study period. Similar studies conducted by Paul et al., 2010 and other (Ali et al., 2008; Amat et al., 2022; Bojude et al., 2018; Ouyang et al., 2021; Puvanesarajah et al., 2019) showed that Older patients greater than 60 years of age were more likely to be lost to follow up which is consistent with our study.

This study also found that older patients are more likely to discontinue follow-up care early. This is because older patients may be more dependent on others; they are less likely to speak up as death is inevitable at such a fragile age and would like to avoid further painful treatment. However, younger patients may be more capable and can attend some of the visits on their own without depending on other family members. Older age as a predictor of discontinuation among breast cancer patients on follow-up therapy was also identified by (Bojude et al., 2018; Fink et al., 2004; Kahn et al., 2007)

The current study revealed that the risk of LTFU Breast Cancer treatment is over four times higher among patients who are Experienced Radiotherapy toxicity compared to non-Experienced Radiotherapy toxicity for both cause-specific and sub hazard models. The study revealed that the risk of LTFU Breast Cancer treatment is over two times higher among patients who are Experienced Advert Effect at follow up compared to not Experienced Advert Effect at follow up for both cause-specific and sub hazard models.

The finding of this study has also showed that treatment type is significantly associated with the time to LTFU Breast cancer treatment patients. For the treatment type of curative chemotherapy only and radiotherapy only, the sub hazard ratio were 1.3 and 1.43 respectively, and the CSHR were 2.01 and 1.45, respectively which implies the patients with curative chemotherapy only and radiotherapy treatment were more likely to be LTFU compared to patients who were receiving no treatment. Limited studies have examined the relationship

between palliative or curative care and LTFU among cancer patients. Patients on palliative care may be less likely to initiate or continue to treatment because of the sense of hopelessness to be cured. This result is supported by a study conducted by (Habinshuti et al., 2020; Memirie et al., 2018; Ruddy et al., 2020) found that among patients with breast cancer, LTFU was substantially connected with older age, non-white race, prior receipt of radiation, chemotherapy, and endocrine treatment, as well as increasing time after surgery.

Overall, when comparing the CSH and SDH models, the current results revealed that higher covariate effects (hazard ratios) were observed for all covariates except age category ≥ 39 years and older, age 40-49, Residence, distance, cancer stage (stage 1), performance and histologic grade (I and III) in multivariable analyses for the CSH model.

Waisberg et al., (Waisberg et al., 2003) compared rival risk estimates using CSH and SDH models using an example of real data. The outcomes of proportional SDH regression models are examined by Latouche et al., (Latouche et al., 2013) when the proportionality assumptions for a CSHs model are satisfied. Their final finding was that the degree of variation in impact estimates between the two models depends on the effects of competing event(s) and cause-specific covariate effects on the event of interest.

Also, Dignam & Kocherginsky, extensively examined and reported the differences between CSH and SDH regression based on several simulated situations (Dignam & Kocherginsky, 2008). Their research showed that the results of a two group comparison including CSH or SDH ratios may vary dramatically when conflicting events are present. These models focus on different metrics, which might lead to widely differing results and interpretations.

Limitations of the Study

The study was limited to data retrieved from log books and patients' individual cards due to the use of a retrospective cohort study design. Most clinical variables without full information were not included, and the study was only limited to the variables mentioned in the methodological part. Likewise, the data used in this study came from Hawassa university referral Hospital, which findings from a single hospital may not represent the national situation and other settings due to specific factors. Caution should be taken when generalizing results.

CHAPTER FIVE

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

In this study, the competing risk models were employed to modeling time to LTFU among breast cancer patients. These models are more appropriate statistical models to modeling time to LTFU treatment breast cancer patients which well describes the time to LTFU patients who were diagnosed in HUCSH, death was used as a competing event. This study concluded that the factors such as Age, Residence, Distance, Cancer stage, Performance ECOG, has NCD comorbidities, Marital status, Experienced Chemotherapy, Experienced Radiotherapy, Experienced Advert Effect at follow up, Histologic grade and treatment type covariates were found to be statistically significant prognostic factors associated to time to LTFU of breast Cancer patients under both CSH and SDH models.

To sum up, the differences of the covariates effects between CSH and SDH models were identified. The hazard ratios were compared to identify the distinctions. Comparing cause-specific and sub-distribution hazard models revealed variations in the impact of covariates, emphasizing the need for tailored interventions to reduce LTFU and improve patient retention in treatment programs.

5.2 Recommendations

The study recommends to reduce the risk of loss to follow-up among breast cancer patients, healthcare systems should implement strong patient tracking mechanisms, expand accessibility to care, and provide financial or transportation support for those in remote areas. Comprehensive management of treatment-related adverse effects is crucial to encourage continued care. High-risk groups, such as older patients, those with advanced cancer stages, and individuals with NCD comorbidities, require targeted interventions through personalized care plans. Raising awareness among patients and caregivers about the importance of treatment adherence can also help mitigate LTFU. Additionally, strengthening national cancer care policies and improving institutional support will enhance patient retention and overall treatment outcomes.

REFERENCES

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. https://books.google.com

Afework, T., Seid, B., Anteneh, A., Ayele, W., Gebreyesus, S. H., & Endris, B. S. (2022). Burden of mortality from cancer among adults in Addis Ababa, Ethiopia, using verbal autopsy, 2007–2017. *Ecancermedicalscience*, *16*.

Akl, E. A., Briel, M., You, J. J., Mulla, S., Lamontagne, F., Bassler, D., Mills, E. J., Walter, S. D., & Cook, D. J. (2012). *Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review*. *2809*(May), 1–12. https://doi.org/10.1136/bmj.e2809

Ali, R., Mathew, A., & Rajan, B. (2008). Effects of socio-economic and demographic factors in delayed reporting and late-stage presentation among patients with breast cancer in a major cancer hospital in South India. *Asian Pac J Cancer Prev*, *9*(4), 703–707.

Alkabban, F. M., & Ferguson, T. (2024). *Continuing Education Activity*.

Amat, M., Duralde, E., Masutani, R., Glassman, R., Shen, C., & Graham, K. L. (2022). "Patient Lost to Follow-up": Opportunities and Challenges in Delivering Primary Care in Academic Medical Centers. *Journal of General Internal Medicine*, *37*(11), 2678–2683. https://doi.org/10.1007/s11606-021-07216-3

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Assemie, M. A., Muchie, K. F., & Ayele, T. A. (2018). Incidence and predictors of loss to follow up among HIV-infected adults at Pawi General Hospital, northwest Ethiopia: competing risk regression model. *BMC Research Notes*, *11*, 1–6.

Austin, A., Craig, S. L., Alessi, E. J., Wagaman, M. A., Paceley, M. S., Dziengel, L., & Balestrery, J. E. (2016). Guidelines for transgender and gender nonconforming (TGNC) affirmative education. *Alexandria, VA: Council on Social Work Education*.

Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting F ine-G ray model analyses for competing risk data. *Statistics in Medicine*, *36*(27), 4391–4400.

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

Austin, P. C., Lee, D. S., & Fine, J. P. (2016b). Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation*, *133*(6), 601–609. https://doi.org/10.1161/CIRCULATIONAHA.115.017719

Ayele, W., Addissie, A., Wienke, A., Unverzagt, S., Jemal, A., Taylor, L., & Kantelhardt, E. J.

(2021). Breast awareness, self-reported abnormalities, and breast cancer in rural Ethiopia: a survey of 7,573 women and predictions of the national burden. *The Oncologist*, *26*(6), e1009–e1017.

Baccolini, V., Isonne, C., Salerno, C., Giffi, M., Migliara, G., Mazzalai, E., Turatto, F., Sinopoli, A., Rosso, A., & De Vito, C. (2022). The association between adherence to cancer screening programs and health literacy: A systematic review and meta-analysis. *Preventive Medicine*, *155*, 106927.

Baccolini, V., Migliara, G., Isonne, C., Dorelli, B., Barone, L. C., Giannini, D., Marotta, D., Marte, M., Mazzalai, E., & Alessandri, F. (2021). The impact of the COVID-19 pandemic on healthcare-associated infections in intensive care unit patients: a retrospective cohort study. *Antimicrobial Resistance & Infection Control*, *10*(1), 87.

Bao, S., Jiang, M., Wang, X., Hua, Y., Zeng, T., Yang, Y., Yang, F., Yan, X., Sun, C., & Yang, M. (2021). Nonmetastatic breast cancer patients subsequently developing second primary malignancy: A population-based study. *Cancer Medicine*, *10*(23), 8662–8672.

Beyersmann, J., Latouche, A., Buchholz, A., & Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, *28*(6), 956–971.

Bhatt, S., Isaac, R., Finkel, M., Evans, J., Grant, L., Paul, B., & Weller, D. (2018). Mobile technology and cancer screening: Lessons from rural India. *Journal of Global Health*, *8*(2).

Birnbaum, J. K., Duggan, C., Anderson, B. O., & Etzioni, R. (2018). Early detection and treatment strategies for breast cancer in low-income and upper middle-income countries: a modelling study. *The Lancet Global Health*, *6*(8), e885–e893.

Bojude, A. D., Dairo, D. M., Ali-Gombe, M., Evaristus, O., Malami, U. A., Ntekim, A., Elnafaty, U. A., Campbell, O. B., & Ayeni, O. (2018). Predictors and Characteristics of Patients Lost to Breast Cancer Follow-up Care in Southeastern Nigeria. *GOMBE MEDICAL JOURNAL*, *1*(1).

Bray, F., Ren, J., Masuyer, E., & Ferlay, J. (2013). Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International Journal of Cancer*, *132*(5), 1133–1145.

Cho, B., Han, Y., Lian, M., Colditz, G. A., Weber, J. D., Ma, C., & Liu, Y. (2021). Evaluation of racial/ethnic differences in treatment and mortality among women with triple-negative breast cancer. *JAMA Oncology*, *7*(7), 1016–1023.

Collett, D. (1994). Modelling survival data. In *Modelling survival data in medical research* (pp. 53–106). Springer.

Collett, D. (2023). *Modelling survival data in medical research*. Chapman and Hall/CRC.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Cserni, G. (2020). Histological type and typing of breast carcinomas and the WHO classification changes over time. *Pathologica*, *112*(1), 25.

da Costa Vieira, R. A., Biller, G., Uemura, G., Ruiz, C. A., & Curado, M. P. (2017). Breast cancer screening in developing countries. *Clinics*, *72*, 244–253.

Dairo, M. D., Adamu, D. B., Onimode, Y. A., Ntekim, A., & Ayeni, O. (2018). Characteristics and Determinants of Patients Discontinuation of Breast Cancer Follow-Up Care at the Radiation Oncology Department, University College Hospital, Ibadan, Nigeria. *International Journal of Breast Cancer*, *2018*(1), 1597964.

Deressa, B. T., Cihoric, N., Badra, E. V., Tsikkinis, A., & Rauch, D. (2019). Breast cancer care in northern Ethiopia–cross-sectional analysis. *BMC Cancer*, *19*, 1–6.

Dignam, J. J., & Kocherginsky, M. N. (2008). Choice and interpretation of statistical tests used when competing risks are present. *Journal of Clinical Oncology*, *26*(24), 4027–4034.

Dignam, J. J., Zhang, Q., & Kocherginsky, M. (2012). The use and interpretation of competing risks regression models. *Clinical Cancer Research*, *18*(8), 2301–2308.

Elsherif, A. R., Ahmedin, A., Ding, Z., & Liu, X. (2012). Adaptive precoding for femtocell interference mitigation. *2012 IEEE International Conference on Communications (ICC)*, 4315–4320.

Ettinger, R. L., Qian, F., Xie, X.-J., & Watkins, C. A. (2004). Evaluation and characteristics of "dropouts" in a longitudinal clinical study. *Clinical Oral Investigations*, *8*, 18–24.

Feng, R., Jing, J., Zhang, X., Li, M., & Gao, J. (2020). Adherence to post-surgery follow-up assessment and its association with sociodemographic and disease characteristics in patients with breast cancer in Central China. *BMC Cancer*, *20*(1), 1–8. https://doi.org/10.1186/s12885-020-07600-y

Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., & Luo, W. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & Diseases*, *5*(2), 77–106.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, *144*(8), 1941–1953. https://doi.org/10.1002/IJC.31937

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509.

Fine, J. P., & Gray, R. J. (2015). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, *94*(446), 496–509.

https://doi.org/10.1080/01621459.1999.10474144

Fink, A. K., Gurwitz, J., Rakowski, W., Guadagnoli, E., & Silliman, R. A. (2004). Patient beliefs and tamoxifen discontinuance in older women with estrogen receptor—positive breast cancer. *Journal of Clinical Oncology*, *22*(16), 3309–3315.

Foerster, M., Anele, A., Adisa, C., Galukande, M., Parham, G., Zietsman, A., Oyamienlen, C. S., Iwuoha, K. C., Ezeigbo, E., Bakengesa, E., Kaggwa, A., Naamala, A., Nteziryayo, A., Nakazibwe, T., Pinder, L. F., Walubita, E., Pontac, J., Anderson, B. O., McKenzie, F., … McCormack, V. (2020b). Few Losses to Follow-up in a Sub-Saharan African Cancer Cohort via Active Mobile Health Follow-up. *American Journal of Epidemiology*, *189*(10), 1185–1196. https://doi.org/10.1093/aje/kwaa070

Fufa, D. B., Diriba, T. A., Dame, K. T., & Debusho, L. K. (2023). Competing risk models to evaluate the factors for time to loss to follow-up among tuberculosis patients at Ambo General Hospital. *Archives of Public Health*, *81*(1), 117.

Ganta, A. G., Wabeto, E., Minuta, W. M., Wegi, C., Berheto, T., Samuel, S., & Assele, D. D. (2024). Predictors of loss to follow up among adults on antiretroviral therapy before and after the start of treat-all strategy in public health facilities of Hawassa city, Ethiopia: a competing risk regression. *Plos One*, *19*(3), e0299505.

Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A., & Siegel, R. L. (2022). Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(6), 524–541. https://doi.org/10.3322/caac.21754

Gill, A., Gosain, R., Bhandari, S., Gosain, R., Gill, G., Abraham, J., & Miller, K. (2018). "Lost to follow-up" among adult cancer survivors. *American Journal of Clinical Oncology*, *41*(10), 1024–1027.

Ginsburg, O., Bray, F., Coleman, M. P., Vanderpuye, V., Eniu, A., Kotha, S. R., Sarker, M., Huong, T. T., Allemani, C., & Dvaladze, A. (2017). The global burden of women's cancers: a grand challenge in global health. *The Lancet*, *389*(10071), 847–860.

Gradishar, W. J., Moran, M. S., Abraham, J., Aft, R., Agnese, D., Allison, K. H., Anderson, B., Burstein, H. J., Chew, H., & Dang, C. (2022). Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, *20*(6), 691–722.

Grambauer, N., Schumacher, M., & Beyersmann, J. (2010). Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified. *Statistics in Medicine*, *29*(7-8), 875–884.

Habinshuti, P., Hagenimana, M., Nguyen, C., Park, P. H., Mpunga, T., Shulman, L. N., Fehr, A., Rukundo, G., Bigirimana, J. B., & Teeple, S. (2020). Factors associated with loss to follow-up among cervical cancer patients in Rwanda. *Annals of Global Health*, *86*(1).

Haller, B. (2014). *The analysis of competing risks data with a focus on estimation of cause-specific and subdistribution hazard ratios from a mixture model*. lmu.

Holmes, J. A., Rachidi, S., Zellars, R. C., & Lautenschlaeger, T. (2013). Demographic Predictors of Missed Follow-Up among Breast Cancer Patients Treated at a Large Multi-Site Academic Institution. *Radiation Oncology Biology*, *102*(3), e408. https://doi.org/10.1016/j.ijrobp.2018.07.1200

Holmes, J. A., Rachidi, S., Zellars, R. C., & Lautenschlaeger, T. (2018). Demographic Predictors of Missed Follow-Up among Breast Cancer Patients Treated at a Large Multi-Site Academic Institution. *International Journal of Radiation Oncology, Biology, Physics*, *102*(3), e408.

Huang, Y. (2013). *Estimating the Impact of Lost to Follow-up for Breast Cancer Patients' Disease-free Survival*. *May*. http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1036&context=statsp

Jahan, S., Al-saigul, A. M., & Abdelgadir, M. H. (2016). Breast Cancer Breast Cancer. *Journal of the Royal Society of Medicine*, *70*(8), 515–517.

Jones, R. H., Casbard, A., Carucci, M., Cox, C., Butler, R., Alchami, F., Madden, T.-A., Bale, C., Bezecny, P., & Joffe, J. (2020). Fulvestrant plus capivasertib versus placebo after relapse or progression on an aromatase inhibitor in metastatic, oestrogen receptor-positive breast cancer (FAKTION): a multicentre, randomised, controlled, phase 2 trial. *The Lancet Oncology*, *21*(3), 345–357.

Kahn, K. L., Schneider, E. C., Malin, J. L., Adams, J. L., & Epstein, A. M. (2007). Patient centered experiences in breast cancer: predicting long-term adherence to tamoxifen use. *Medical Care*, *45*(5), 431–439.

Kaku, M., Mathew, A., & Rajan, B. (2008). Impact of socio-economic factors in delayed reporting and late-stage presentation among patients with cervix cancer in a major cancer hospital in South India. *Asian Pac J Cancer Prev*, *9*(4), 589–594.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kantelhardt, E. J., Moelle, U., Begoihn, M., Addissie, A., Trocchi, P., Yonas, B., Hezkiel, P., Stang, A., Thomssen, C., & Vordermark, D. (2014). Cervical cancer in Ethiopia: survival of 1,059 patients who received oncologic therapy. *The Oncologist*, *19*(7), 727–734.

Kantelhardt, E. J., Zerche, P., Mathewos, A., Trocchi, P., Addissie, A., Aynalem, A., Wondemagegnehu, T., Ersumo, T., Reeler, A., & Yonas, B. (2014). Breast cancer survival in Ethiopia: a cohort study of 1,070 women. *International Journal of Cancer*, *135*(3), 702–709.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations.

Journal of the American Statistical Association, *53*(282), 457–481.

Kimura, M., Yamada, S., Go, M., Yasuda, S., Toyoda, H., & Usami, E. (2024). Identifying risk factors of dose reduction or treatment discontinuation due to fatigue or gastrointestinal symptoms in patients receiving lenvatinib treatment for hepatocellular carcinoma. *Journal of Oncology Pharmacy Practice*, 10781552241281900.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*(2), 137–163.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (Vol. 1230). Springer.

Knaul, F. M., Bhadelia, A., Gralow, J., Arreola-Ornelas, H., Langer, A., & Frenk, J. (2012). Meeting the emerging challenge of breast and cervical cancer in low-and middle-income countries. *International Journal of Gynecology & Obstetrics*, *119*, S85–S88.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232.

Kuk, D., medicine, R. V.-S. in, & 2013, undefined. (2013). Model selection in competing risks regression. *Wiley Online LibraryD Kuk, R VaradhanStatistics in Medicine, 2013•Wiley Online Library*, *32*(18), 3077–3088. https://doi.org/10.1002/sim.5762

Kumar, P., Bird, C., Holland, D., Joshi, S. B., & Volkin, D. B. (2022). Current and next-generation formulation strategies for inactivated polio vaccines to lower costs, increase coverage, and facilitate polio eradication. *Human Vaccines & Immunotherapeutics*, *18*(7), 2154100.

Lakshmanan, B. C. (2015). *Manuscript Info*. *3*(6), 583–588.

Lambert, P. C., Wilkes, S. R., & Crowther, M. J. (2017). Flexible parametric modelling of the cause-specific cumulative incidence function. *Statistics in Medicine*, *36*(9), 1429–1446.

Lamont, E. B., Hayreh, D., Pickett, K. E., Dignam, J. J., List, M. A., Stenson, K. M., Haraf, D. J., Brockstein, B. E., Sellergren, S. A., & Vokes, E. E. (2003). Is patient travel distance associated with survival on phase II clinical trials in oncology? *Journal of the National Cancer Institute*, *95*(18), 1370–1375.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., & Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology*, *66*(6), 648–653.

Lei, S., Zheng, R., Zhang, S., Wang, S., Chen, R., Sun, K., Zeng, H., Zhou, J., & Wei, W. (2021). Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Communications*, *41*(11), 1183–1194.

Leoce, N. (2016). *Prognostic Modeling in the Presence of Competing Risks: An Application to Cardiovascular and Cancer Mortality in Breast Cancer Survivors*. https://search.proquest.com/18750

Li, B. D. L., Brown, W. A., Ampil, F. L., Burton, G. V, Yu, H., & McDonald, J. C. (2000). Patient compliance is critical for equivalent clinical outcomes for breast cancer treated by breast-conservation therapy. *Annals of Surgery*, *231*(6), 883–889.

Li, J., Scheike, T. H., & Zhang, M.-J. (2015). Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals. *Lifetime Data Analysis*, *21*(2), 197–217.

Logan, B. R., Zhang, M.-J., & Klein, J. P. (2006). Regression models for hazard rates versus cumulative incidence probabilities in hematopoietic cell transplantation data. *Biology of Blood and Marrow Transplantation*, *12*(1), 107–112.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719–748.

Maranga, I. O., Hampson, L., Oliver, A. W., Gamal, A., Gichangi, P., Opiyo, A., Holland, C. M., & Hampson, I. N. (2013). Analysis of factors contributing to the low survival of cervical cancer patients undergoing radiotherapy in Kenya. *PloS One*, *8*(10), e78411.

Martinez, M. E., Schmeler, K. M., Lajous, M., & Newman, L. A. (2024). Cancer Screening in Low-and Middle-Income Countries. *American Society of Clinical Oncology Educational Book*, *44*(3), e431272.

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.

Michelle, T., Koh, R. Y. V., & Lee, E. S. (2023). A scoping review on the factors associated with the lost to follow-up (LTFU) amongst patients with chronic disease in ambulatory care of high-income countries (HIC). *BMC Health Services Research*, *23*(1), 1–21. https://doi.org/10.1186/s12913-023-09863-0

Mohammad, K. A., Fatima-Tuz-Zahura, M., & Bari, W. (2017). Fine and Gray competing risk regression model to study the cause-specific under-five child mortality in Bangladesh. *BMC International Health and Human Rights*, *17*(1), 1–8. https://doi.org/10.1186/S12914-017-0112-8

Mosha, D., Mahande, M., Ahaz, J., Mosha, M., Njau, B., Kitalp, B., & Obure, J. (2009). Factors associated with management of cervical cancer patients at KCMC Hospital, Tanzania: a retrospective cross-sectional study. *Tanzania Journal of Health Research*, *11*(2).

Mozumder, S. I., Rutherford, M. J., & Lambert, P. C. (2021). Estimating restricted mean

survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models. *BMC Medical Research Methodology*, *21*, 1–20.

Mueller, N. (1995). Overview: viral agents and cancer. *Environmental Health Perspectives*, *103*(suppl 8), 259–261.

Nawaz, Y., Nazar, I., Ali, U., Ahmad, H., Iqbal, N., & kaleem Ullah, M. (2023). *Research Journal of Applied Medical Sciences*.

Ng, T., Toh, M. R., Cheung, Y. T., & Chan, A. (2015). *Follow-up care practices and barriers to breast cancer survivorship : perspectives from Asian oncology practitioners*. https://doi.org/10.1007/s00520-015-2700-2

Ngo, U. E. M., Ntsama, J. A. M., Motolouze, K., Schouame, N. M. L. N., Atenguena, O. E., Ndoua, C. C. N., & Mbu, R. E. (2024). Determinants of Early Survival of Breast Cancer Patients in Yaoundé-Cameroon. *Open Journal of Obstetrics and Gynecology*, *14*(3), 487–501.

Noordzij, M., Leffondré, K., van Stralen, K. J., Zoccali, C., Dekker, F. W., & Jager, K. J. (2013). When do we need competing risks methods for survival analysis in nephrology? *Nephrology Dialysis Transplantation*, *28*(11), 2670–2677.

Ouyang, Q., Li, S., Gao, M., Zhu, L., Xu, S., Meng, S., Wu, S., Huang, L., Su, F., & Ren, Z. (2021). Risk factors associated with loss to follow-up of breast cancer patients: A retrospective analysis. *The Breast*, *57*, 36–42.

Palareti, G., Cosmi, B., Lessiani, G., Rodorigo, G., Guazzaloca, G., Brusi, C., Valdré, L., Conti, E., Sartori, M., & Legnani, C. (2010). Evolution of untreated calf deep-vein thrombosis in high risk symptomatic outpatients: the blind, prospective CALTHRO study. *Thrombosis and Haemostasis*, *104*(11), 1063–1070.

Paul, M., George, P. S., & Mathew, A. (2010). Patient and disease related factors associated with lost-to follow-up/drop-outs of cervical cancer patients: a study at a Major Cancer Hospital in South India. *Asian Pac J Cancer Prev*, *11*(6), 1529–1534.

Pintilie, M. (2006). *Competing risks: a practical perspective*. https://books.google.ep78

Pintilie, M. (2011). An introduction to competing risks analysis. *Revista Española de Cardiología (English Edition)*, *64*(7), 599–605.

Porter, P. (2008). ``Westernizing''Women's Risks? Breast Cancer in Lower-Income Countries. *New England Journal of Medicine*, *358*(3), 213.

Prentice, R., Kalbfleisch, J., Biometrics, A. P. J.-, & 1978, undefined. (n.d.). The analysis of failure times in the presence of competing risks. *JSTORRL Prentice, JD Kalbfleisch, AV Peterson Jr, N Flournoy, VT Farewell, NE BreslowBiometrics, 1978•JSTOR*. Retrieved January 29, 2024, from https://www.jstor.org/stable/2530374

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V, Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541–554.

Puvanesarajah, S., Gapstur, S. M., Patel, A. V, Sherman, M. E., Flanders, W. D., Gansler, T., Troester, M. A., & Gaudet, M. M. (2019). Mode of detection and breast cancer mortality by follow-up time and tumor characteristics among screened women in Cancer Prevention Study-II. *Breast Cancer Research and Treatment*, *177*, 679–689.

Rass, A. M. R. (2023). *The tumour microenvironment: the effect of breast cancer cell conditioned medium on the endothelium*.

Roux, P., Knight, S., Cohen, M., Classe, J. M., Mazouni, C., Chauvet, M.-P., Reyal, F., Colombo, P.-E., Jouve, E., & Chopin, N. (2019). Tubular and mucinous breast cancer: results of a cohort of 917 patients. *Tumori Journal*, *105*(1), 55–62.

Ruddy, K. J., Herrin, J., Sangaralingham, L., Freedman, R. A., Jemal, A., Haddad, T. C., Allen, S. V, Hieken, T., Boughey, J. C., & Ganz, P. A. (2020). Follow-up care for breast cancer survivors. *JNCI: Journal of the National Cancer Institute*, *112*(1), 111–113.

Santiá, P., Jansana, A., Del Cura, I., Padilla-Ruiz, M., Domingo, L., Louro, J., Comas, M., Sanz, T., Duarte-Salles, T., & Redondo, M. (2022). Adherence of long-term breast cancer survivors to follow-up care guidelines: a study based on real-world data from the SURBCAN cohort. *Breast Cancer Research and Treatment*, *193*(2), 455–465.

Saunders, C., & Jassal, S. (2009). *Breast cancer*. OUP Oxford.

Schober, P., & Vetter, T. R. (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia & Analgesia*, *127*(3), 792–798.

Schuster, N. A., Hoogendijk, E. O., Kok, A. A. L., Twisk, J. W. R., & Heymans, M. W. (2020). Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *Journal of Clinical Epidemiology*, *122*, 42–48.

Shaikh, K., Krishnan, S., Thanki, R., Shaikh, K., Krishnan, S., & Thanki, R. (2021). An introduction to breast cancer. *Artificial Intelligence in Breast Cancer Early Detection and Diagnosis*, 1–20.

Shalamo, T., Geleta, D., & Yilikal, S. (2022). PREDICTORS OF BREAST CANCER AMONG WOMEN ATTENDI, ETHIOPIA: A HOSPITAL-BASED UNMATCHED CASE CONTROL STUDY. In *HU* (Issue 8.5.2017). www.aging-us.com

Shita, A., Yalew, A. W., Seife, E., Afework, T., Tesfaw, A., Gufue, Z. H., Rabe, F., Taylor, L., Kantelhardt, E. J., & Getachew, S. (2023). Survival and predictors of breast cancer mortality in South Ethiopia: A retrospective cohort study. *Plos One*, *18*(3), e0282746.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, *18*(2), 155–195.

statistics, R. G.-T. A. of, & 1988, undefined. (n.d.). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *JSTORRJ GrayThe Annals of Statistics, 1988•JSTOR*. Retrieved January 29, 2024, from https://www.jstor.org/stable/2241622

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021a). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021b). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Syed, S. T., Gerber, B. S., & Sharp, L. K. (2013). Traveling towards disease: transportation barriers to health care access. *Journal of Community Health*, *38*, 976–993.

Tchounzou, R., Simo Wambo, A. G., Njamen, T. N., Ilick, I. O., Neng, H. T., Dadao, F., & Sone, A. M. (2019). Patients lost to follow-up for cervical cancer in the Limbe Regional Hospital. *Journal of Global Oncology*, *4*, 1–5.

Tesfaw, A., Alebachew, W., & Tiruneh, M. (2020). Why women with breast cancer presented late to health care facility in North-west Ethiopia? A qualitative study. *PloS One*, *15*(12), e0243551.

Tiruneh, M., Tesfaw, A., & Tesfa, D. (2021). Survival and Predictors of Mortality among Breast Cancer Patients in Northwest Ethiopia: A Retrospective Cohort Study. *Cancer Management and Research*, *13*, 9225–9234. https://doi.org/10.2147/CMAR.S339988

Uematsu, T. (2024). Rethinking screening mammography in Japan: next-generation breast cancer screening through breast awareness and supplemental ultrasonography. *Breast Cancer*, *31*(1), 24–30. https://doi.org/10.1007/s12282-023-01506-w

Vallikad, E. (2006). Cervical cancer: the Indian perspective. *International Journal of Gynecology & Obstetrics*, *95*, S215–S233.

Van Der Pas, S., Nelissen, R., & Fiocco, M. (2018). Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events: illustrated with 138,234 hip (124,560 patients) and 139,070 knee (125,213 patients) replacements from the Dutch Arth. *Acta Orthopaedica*, *89*(2), 145–151.

Varadhan, R., Weiss, C. O., Segal, J. B., Wu, A. W., Scharfstein, D., & Boyd, C. (2010). Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Medical Care*, *48*(6), S96–S105.

Vidhubala, E., Shewade, H. D., Niraimathi, K., Dongre, A. R., Gomathi, R., Ramkumar, S., & Sankar, M. B. (2020). Loss to follow-up after initial screening for cervical cancer: A qualitative exploration of barriers in Southern India. *Cancer Research, Statistics, and Treatment*, *3*(4), 700–707.

Vock, D. M., Wolfson, J., Bandyopadhyay, S., Adomavicius, G., Johnson, P. E., Vazquez-Benitez, G., & O'Connor, P. J. (2016). Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting. *Journal of Biomedical Informatics*, *61*, 119–131.

Wabinga, H., Ramanakumar, A. V, Banura, C., Luwaga, A., Nambooze, S., & Parkin, D. M. (2003). Survival of cervix cancer patients in Kampala, Uganda: 1995–1997. *British Journal of Cancer*, *89*(1), 65–69.

Waisberg, M., Joseph, P., Hale, B., & Beyersmann, D. (2003). Molecular and cellular mechanisms of cadmium carcinogenesis. *Toxicology*, *192*(2–3), 95–117.

Wallace, R., Hughes-Cromwick, P., Mull, H., & Khasnabis, S. (2005). Access to health care and nonemergency medical transportation: two missing links. *Transportation Research Record*, *1924*(1), 76–84.

Ward, S. E., Holmes, G. R., Ring, A., Richards, P. D., Morgan, J. L., Broggio, J. W., Collins, K., Reed, M. W. R., & Wyld, L. (2019). Adjuvant chemotherapy for breast cancer in older women: an analysis of retrospective English cancer registration data. *Clinical Oncology*, *31*(7), 444–452.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. *Handbook of Structural Equation Modeling*, *1*, 209–231.

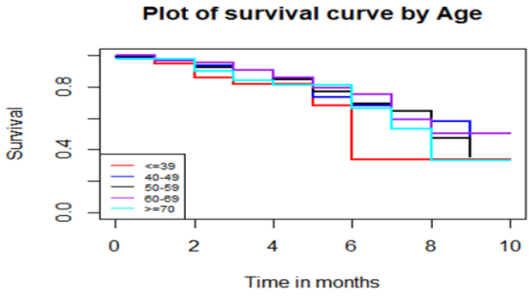
Wilkinson, L., & Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British Journal of Radiology*, *95*(1130), 20211033.

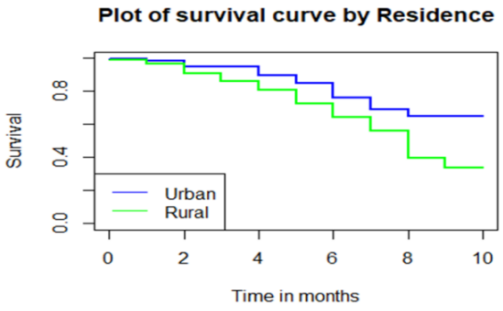
Wolbers, M., Koller, M. T., Stel, V. S., Schaer, B., Jager, K. J., Leffondre, K., & Heinze, G. (2014). Competing risks analyses: objectives and approaches. *European Heart Journal*, *35*(42), 2936–2941.

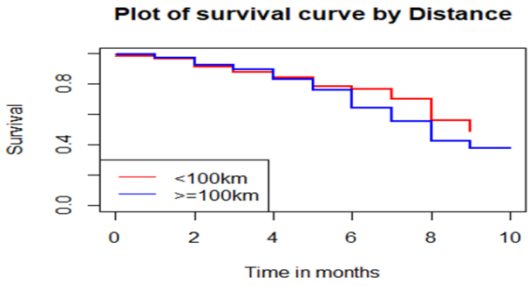
Yang, J., Pan, Z., He, Y., Zhao, F., Feng, X., Liu, Q., & Lyu, J. (2019). Competing-risks model for predicting the prognosis of penile cancer based on the SEER database. *Cancer Medicine*, *8*(18), 7881–7889.

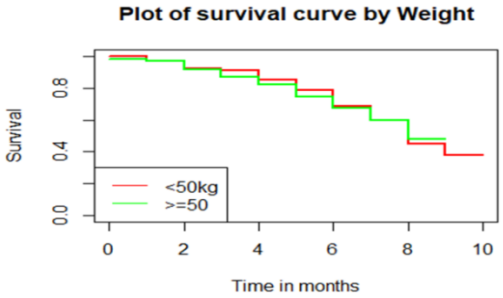
Zhang, Z. (2017). Survival analysis in the presence of competing risks. *Annals of Translational Medicine*, *5*(3).

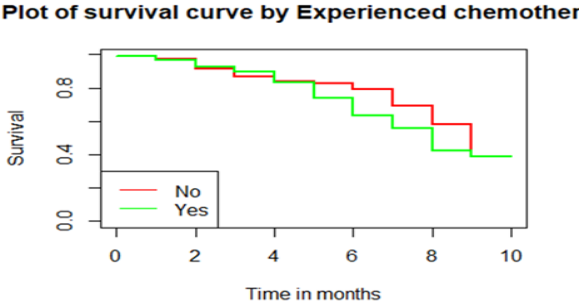
Appendices -Appendix-I Kaplan-Meir estimate survival curve for LTFU

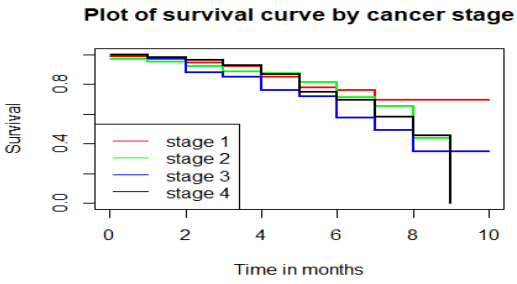


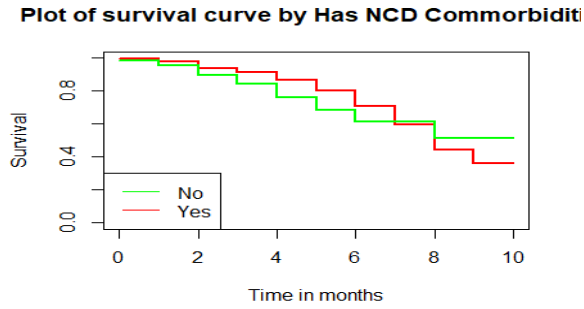


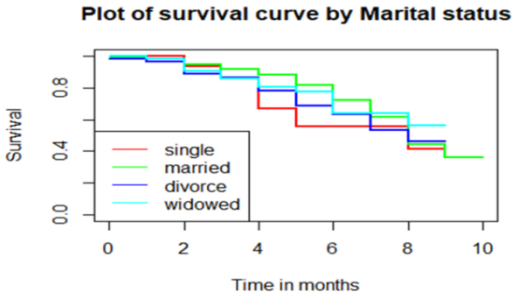


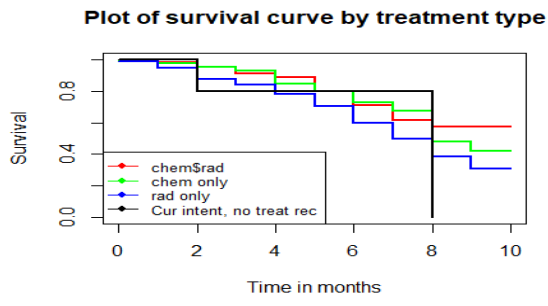












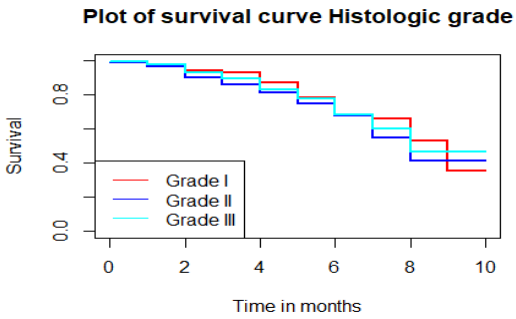


Figure 6 : Non-parametric estimates of Kaplan-Meir survival curves with LTFU for all variables.

The survival curve of residence shows the probability of surviving for a certain amount of time. In the graph, the survival curve for people who live in urban areas is higher than the survival curve for people who live in rural areas. This means that people who live in urban areas are more likely to survive for a given amount of time than people who live in rural areas.

Survival curves of cancer stage are used to estimate the probability of surviving for a certain amount of time after a diagnosis. In the graph, the survival curve for stage 1 cancer is the highest, followed by stage 2, stage 3, and stage 4. The survival curve for stage 4 cancer is the lowest. This means that people with stage 1 cancer are more likely to survive for a given amount of time than people with other stages of cancer. Survival curves are used to estimate the probability of surviving for a certain amount of time after a diagnosis.

In the graph, the survival curve for people who do not have NCD comorbidity (the green line) is higher than the survival curve for people who do have NCD comorbidity (the red line). This means that people who do not have NCD comorbidity are more likely to survive for a given amount of time than people who do have NCD comorbidity.

Appendix-II Proportionality of the hazard of CSPH for LTFU

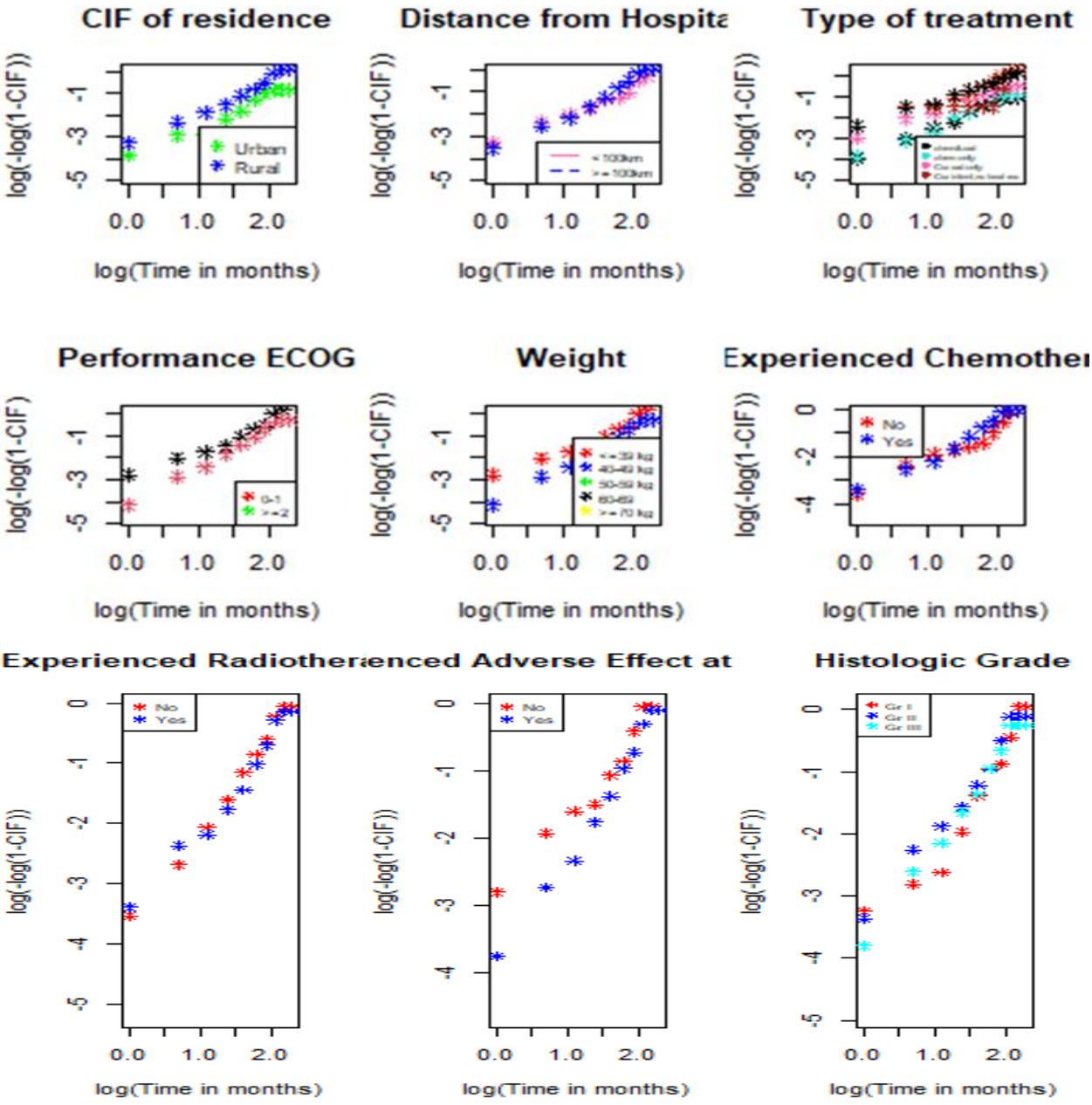


Figure 7 : Proportionality of the hazard of CIF for residence, distance, weight, treatment type, Histologic grade, adverse effect, and baseline weights

Appendix-III Proportionality assumption in Fine-Gray model

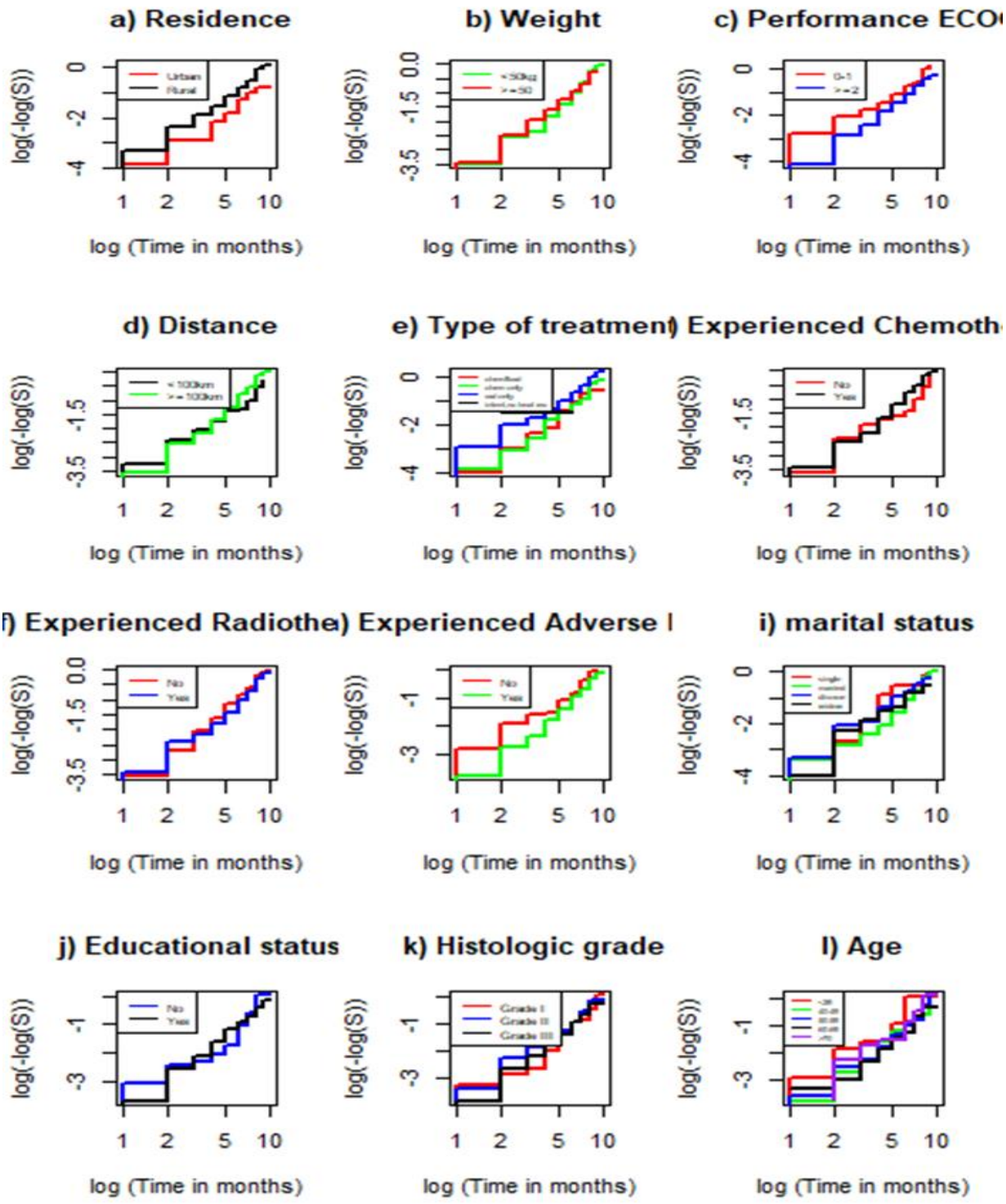


Figure 8 : Proportionality of the hazard of CIF for all variables

Appendix-IV Cumulative incidence curves

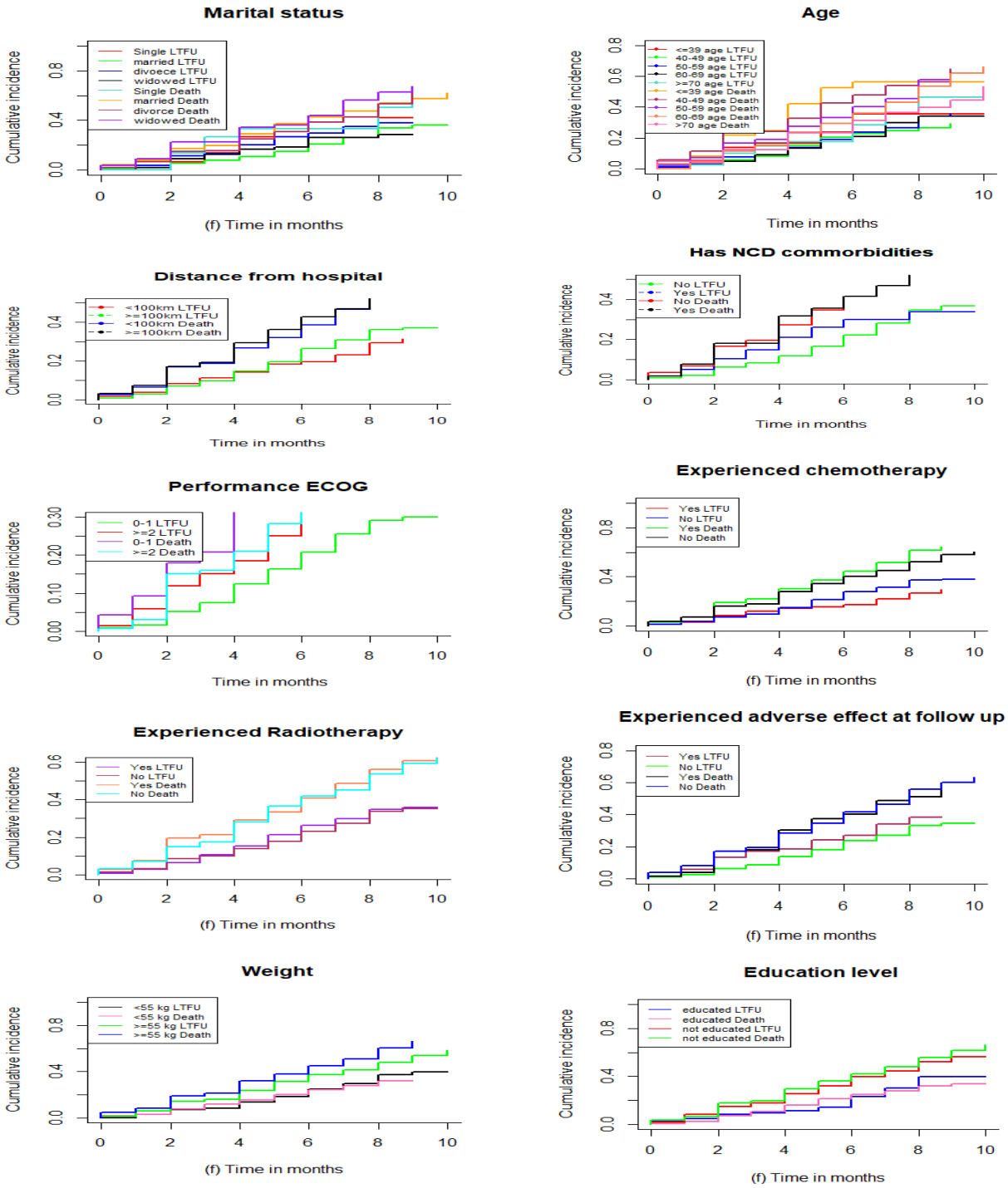


Figure 9 : Non-parametric estimates of cumulative incidence curves with LTFU and death as a competing event for selected variables categories.

Appendix V: Median survival (follow up) time

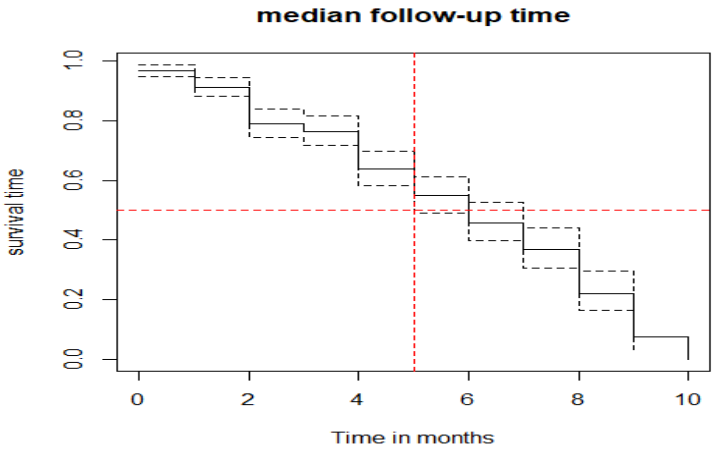


Figure 10 : Median survival (follow up) time graph

Table 9 : Relative difference of the hazard ratios

Variables	Categories	SDHR (95%CI)	CSHR(95%CI)	RD
Age	<=39(Ref)			
	40-49	0.669[0.325,1.37349]	0.6350[0.2940,1.3717]	5.35
	50-59	2.725[0.374,4.40563]	3.5837[0.2850,5.1931]	23.9
	60-69	8.653[7.316,10.4678]	9.7225[0.2400,10.360]	11.0
	≥70	5.857[3.384,7.91185]	7.6820[4.2830,9.5289]	23.7
Residence	Urban (Ref)			
	Rural	2.050[1.240,3.38482]	1.8420[1.1430,2.9682]	11.3
Distance	$< 100\ km$			
	$\geq 100\ km$	2.011[0.671,2.52991]	1.2330[0.8000,2.9015]	63.1
Marital status	Single (Ref)			
	Married	2.744[0.4310,4.2800]	4.72655[0.5288,40.264]	41.9
	Divorce	7.070[3.6220,9.8500]	12.1107[1.3987,104.78]	41.6
	widowed	2.813[0.4620,3.8830]	8.77595[1.0032,76.462]	67.9
Cancer stage	Stage I (Ref)			
	Stage II	1.45[0.7450,2.81423]	1.4670[0.7544,2.8520]	1.2
	Stage III	1.86[1.020,3.389460]	2.7680[0.9710,3.2174]	32.8
	Stage Iv	2.38[0.6970,2.72557]	3.5190[0.7830,3.9443]	32.4
Performance (ECOG)	0-1 (Ref)			
	2+	2.404[0.257,4.25687]	1.5210[0.3280,1.8304]	58.1
Has NCD	No (Ref)			
	Yes	1.846[0.560,2.27859]	2.8589[0.5670,4.3006]	35.4
Exp. Chemo	No (Ref)			
	Yes	1.38[0.9390,2.01628]	1.3960[0.8920,1.9713]	1.2
Exp. Radio	No (Ref)			
	Yes	3.974[1.662,4.43318]	3.9866[1.637,4.3763]	0.3
Treatment type	No treat (Ref)			
	Chemo only	1.290[1.02,3.071961]	2.0110[0.745,3.5155]	35.8
	Radio only	1.430[0.451,4.53328]	1.4520[1.150,6.5755]	1.52
	Both chemo & radio	1.190[0.653,2.15775]	1.4750[0.803,2.7087]	19.3
Exp. adverse effect	No (Ref)			
	Yes	0.898[0.602,1.27573]	2.7626[1.520,3.1176]	67.5
Histologic grade	Grade I(Ref)			
	Grade II	1.170[0.749,1.84146]	1.3770[0.865,2.1919]	15.03
	Grade III	1.977[0.618,2.54404]	1.1280[0.692,1.8355]	75.3

Appendix-VI Information sheet and data extraction form

Information sheet

Introduction:- This information sheet is prepared for Hawassa University Comprehensive Specialized Hospital, Hawassa, Ethiopia. The aim of the form is to make clear about the purpose of the thesis, data collection procedures and to get permission for data collection.

Objective:- The aim of this study is to identify factors affecting survival time of loss to follow up treatment among Breast cancer patients in HUCSH, Hawassa, Ethiopia using a competing risk model.

Data collection procedure:- In order to achieve the above objective, information, which is necessary for the study, will be taken from the registration log book and patients registration card; if any inadequate information is countered it is checked from the file and excluded from analysis if proven to be inadequate. In order to come up with the above mentioned findings, a total document of program clients enrolled during first January, 2020 up to December 2023 will be seen and a review of the required information from the records are made by using the checklist.

Confidentiality: - To ensure confidentiality the data on the chart will be collected by the researcher and those individuals who are working in a Hospital unit nurse and information will be collected without the name of the clients. The information collected from this thesis will be kept confidential and will be stored in a card. In addition, it will not be revealed to anyone except the investigator.

5.3 Ethical Consideration

The research ethics review board of Hawassa University will provide an ethical clearance for the study. The data will be collected from Hawassa University Oncology Center after permission is obtained from the Hospital and to get the permission of data collection Department of Statistics write an official cooperation letter to the Hospital. The study will conducted without individual informed consent because it relied on retrospective data. Confidentiality of any information related to the patients and their clinical history will be maintained by keeping both the hard-copy and soft-copy of every collected data in a locked cabinet and password secured computer. Only the researcher will access to the identified data that has been kept in a secure place. All data will be coded with numbers. All analysis will be on identified and coded data obtained from hospital. During the study, there is no contact between the patients and the researcher. The study is non-invasive and without any harm to the patients. In this research, the information obtained from log book and patients' card will kept secured.

5.4 Data collection tools or format Table

Table 10: Data collection tools or format

No.	Categories	Options	Response
1	Smoking history of pt.	0. Yes 1. No	
2	Drinking History of pt.	0. Yes 1. No	
3	Has NCD Comorbidities of pt.	0. Yes 1. No	
4	Experienced chemotherapy toxicity	0. Yes 1. No	
5	Experienced Radiotherapy Toxicity	0. Yes 1. No	
6	Experienced adverse effect at follow up	0. Yes 1. No	
7	Weight of pt.	Weight in KG	
8	Residence of pt.	0. Urban 1. Rural	
9	Distance of pt. from hospital	Distance in KM	
10	Educational Status of pt.	0.Secondary& above(Ref) 1. Primary and below	
11	HIV Status of pt.	0. Negative 1. Positive	
12	Age of pt.	Age at baseline	
13	Performance status(EGOG) of pt.	0. 0−1 1. 2+	
14	Marital status of pt.	Single Married Divorce widowed	
15	Cancer stage of pt.	Stage I Stage II Stage III Stage Iv	
16	Type of treatment received of pt.	No treat (Ref) Chemo only Radio only Both chemo & radio	
17	Histologic grade of pt.	Grade I Grade II Grade III	