JOINT MODELLING OF RECURRENT EVENT IN PROSTATE CANCER AND TIME TO TERMINATE OF PATIENTS IN TIKUR ANBESSA SPECIALIZED HOSPITAL

MSC THESIS

BY IBSA NURU

HAWASSA UNIVERSITY,

Hawassa, Ethiopia

October 30, 2023

JOINT MODELING OF A RECURRENT EVENT IN PROSTATE CANCER AND TIME TO TERMINATE OF PATIENTS IN TIKUR ANBESSA SPECIALIZED HOSPITAL

BY IBSA NURU

A THESIS SUBMITTED TO THE SCHOOL OF STATISTICAL SCIENCE

HAWASSA UNIVERSITY,

HAWASSA, ETHIOPIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE IN STATISTICS

(SPECIALIZATION: BIOSTATISTICS)

October 30, 2023

APPROVAL SHEET -1

This is to certify that the thesis entitled “Joint Modelling Of Recurrent Event In Prostate Cancer And Time To Terminate Of Patients In Tikur Anbessa Specialized Hospital” submitted in partial fulfillment of the requirements for the degree of master of science in statistics with specialization in Bio-Statistics of the graduate program of the College Of Natural and Computational Science, Hawassa University, and is a record of original research carried out by Ibsa Nuru ID.No: _____, under my supervision, and no part of the thesis has been submitted for any other degree or diploma. The guidance and assistance received during the course this investigation have all been duel acknowledged. Therefore, I recommend that it would be accepted fulfilling the thesis requirements.

_____	_____	_____
Name of Advisor	Signature	Date

APPROVAL SHEET -2

We the undersigned, member of the board of examiners of the final open defense by Ibsa Nuru have read and evaluated his thesis entitled "Joint Modelling of Recurrent Event in Prostate Cancer and Time to Terminate of Patients in Tikur Anbessa Specialized Hospital" and examined the candidate. This is, therefore, to certify that thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Statistics (Specialization: Biostatistics).

_____	_____	_____
Name of Department Head	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date
_____	_____	_____
Name of Internal Examiner I	Signature	Date
_____	_____	_____
Name of Internal Examiner II	Signature	Date
_____	_____	_____
School of Graduate committee	Signature	Date

Table of Contents

Acronyms

ACS..... American Cancer Society

AJCC.....American Joint Committee on Cancer

Cox-PH.....Cox Proportional Hazard model

GLOBOCAN.....Global Burden of Cancer

LCV.....Likelihood cross-validation criterion

IARC.....International Agency for Research on cancer

SSA.....Sub-Saharan Africa

NCI.....National Cancer Institute

TASH.....Tikur Anbessa Specialized Hospital

WHO.....World Health Organization

Abstract

Background*: The prostate gland, an organ found in the male reproductive system, is where prostate cancer typically originates. Prostate cancer ranks fourth among malignancies diagnosed in men behind lung, colorectal, and stomach cancers worldwide. Recurring occurrences of the same or different kinds of events for specific people or units across time are referred to as recurrent events such as prostate cancer which is an important clinical indicator and the leading cause of prostate cancer mortality. The major aim of this study was to investigating predictors of prostate cancer recurrence and terminal events (death) due to prostate recurrence.*

Methodology*: to reach the aim, 222 prostate cancer patients, between the study period January 1, 2018 to January 30, 2021, who were registered with detailed, comprehensive personal and medical information were include. The retrospective longitudinal study design was applied and the data were analyzed using joint frailty model on prostate CA and death that are recorded in the oncology department of Tikur Anbessa specialized hospital. To investigate determining factors of prostate cancer recurrence and death, a joint frailty model of the recurrent event and terminal event proposed by Liu and others i.e., the joint frailty proportional hazards model was used alongside reduced models for prostate CA recurrence and terminal event (death).*

Result*: From the total of 432 recurrent observation, about 210 (61.1%) of them experienced recurrence of prostate cancer, 192 of the experienced death (terminated) event and 222 (38.9%) were censored. The shared gamma joint frailty model was chosen as the best fit for the prostate cancer data set based on the value of Likelihood cross validation criterion. From the result of shared gamma joint frailty model smoking, stage of prostate cancer, distance metastasize and Gleason score were significantly associated with recurrence of prostate cancer and death.*

Conclusion and recommendation*: The result of shared gamma joint frailty model shows that the stage (III, IV), smoking, distance metastasize (metastasized tumor) and Gleason score were significantly increases the risk of recurrence of prostate cancer and death. While, quitting smoking may improve patients overall prognosis. Timely detection and management of metastasis in prostate cancer patients are crucial, necessitating focused treatment and surveillance. It is recommended that policy maker, ministry of health and Tikur Anbessa Specialized Hospital are expected to make intervention to improve the management and care of prostate cancer patients, ultimately enhancing their quality of life and prognosis.*

Key words: prostate cancer, Counting Approach, Recurrent events, Shared frailty model, joint model

1. **Introduction**

1.1 **Background of the study**

The prostate gland, a tiny, walnut-shaped organ found in the male reproductive system, is where prostate cancer typically originates. When cells in the prostate gland start to multiply and expand uncontrollably, a tumor is created, which is known as prostate cancer (prostate CA) (American Cancer Society, 2022a). The actual etiology of prostate cancer is unknown, but several variables may play a role in its emergence, according to the WHO. Given that prostate cancer incidence rises with age, age is the most important risk factor. Men are also more likely to have the disease if they have a family history of prostate cancer or certain hereditary gene mutations, like those in the BRCA1 and BRCA2 genes. Prostate cancer risk can also be increased by environmental and lifestyle factors, including smoking, being obese, eating a diet heavy in red meat and low in fruits and vegetables, and being exposed to certain chemicals and toxins(WHO, 2021).

Prostate cancer frequently has no indications at the outset, making periodic screening crucial for early detection. Yet, if cancer spreads, it could result in symptoms like frequent urinating, particularly after dark, difficulty urinating, either starting or stopping, insufficient or irregular urine flow, discomfort or burning during urinating, Having blood in your pee or sperm, discomfort in the thighs, hips, or lower back and erection problem (National Cancer Institute, 2023) s.

Prostate cancer, which ranks fourth among malignancies diagnosed in men behind lung, colorectal, and stomach cancers, is predicted to have 1.4 million new cases worldwide in 2020, according to the International Institute for Research on Cancer (IARC). It is also predicted to have a mortality rate of 375,304 deaths globally in 2020, making it the fifth greatest cause of cancer death in males, after lung, colorectal, stomach, and liver cancers (Sung et al., 2021a).

Prostate cancer incidence varies greatly by geographic location, with Asia and Africa reporting the lowest rates while North America, Europe, and Australia have the highest rates. It's possible that genetic variations, lifestyle factors, and screening procedures are to blame for this variation in incidence rates (Sung et al., 2021b).

Prostate cancer accounts for 15% of all instances of male cancer and is the most often diagnosed malignancy among men in Africa. Prostate cancer is anticipated to cause 89,604 new cases and 50,773 fatalities in Africa in 2020(WHO Africa, 2020). Prostate cancer occurs at a rate of 26.6 per

100,000 people in Africa, and there are 70,000 new instances of the disease each year in sub-Saharan Africa (Bray et al., 2022).

Prostate cancer is a leading cause of cancer-related mortality in African men. The GCO reports that Zimbabwe, South Africa, and Namibia had the highest age-standardized mortality rates from prostate cancer in Africa, with rates of 37.1, 24.3, and 23.8 per 100,000, respectively. In contrast, North Africa has the lowest death rates, with age-standardized rates ranging from 1.0 to 4.6 per 100,000 people (Global Cancer Observatory, 2020). Prostate cancer has quietly increased in prevalence in Ethiopian men recently. Leukemia and colon cancer, the third most prevalent cancer in men, are next (Solomon & Mulugeta, 2019).

There were a projected 2,364 new instances of prostate cancer in Ethiopia in 2020, according to the most recent GCO data, with an age-standardized incidence rate of 12.9 per 100,000 people. According to estimates, there would be 1,740 prostate cancer fatalities in Ethiopia in 2020, or 8.9 deaths for every 100,000 people (Ferlay et al., 2019a).

Recurring occurrences of the same or different kinds of events for specific people or units across time are referred to as recurrent events. There is a recurrent pattern in the outcome variable of interest in many scientific investigations. Many different fields, including business, social research, public health, insurance, and reliability, frequently use recurrent event data (Cook & Lawless, 2007).

It makes sense to use this type of data to examine the relationship between significant predictors and the rate at which events occur while accounting for many events per subject (David G. Kleinbaum, 2011). Recurring incidents are also reported by prostate cancer patients. The recurrence of prostate cancer is an important clinical indicator and the leading cause of prostate cancer mortality. It occurs when cancer cells are discovered following the first surgery, radiation treatment, or chemotherapy. The severity of the illness, the initial course of treatment, and the pati-

ent's functional state and co-morbidities can all influence whether the disease returns. The simplest method of analysis for recurrent events is to simply count the occurrences over a specified time frame. These counts could be distributed Poissonically, quasi-Poissonically, or negatively binomially (Wang et al., 2009). Analysis procedures for event times should be favored over straightforward counting methods whenever patients are not all fully watched but are instead subject to an underlying censoring mechanism. Although this circumstance occurs considerably more frequently in clinical applications, our focus is on event time models rather than counting models.

Because it overcomes the limitations of conventional regressions like logistic and linear regressions, we have used survival analysis in this study. One of the popular methods for analyzing data on time to an incident is the Cox proportional hazard model.

In other words, regardless of the variables, every person is equally at risk of experiencing an event like a recurrence of an illness. Nevertheless, the usual Cox model ignores all subsequent events and just takes into account the time until the first occurrence. The model by Andersen and Gill (Andersen et al., 1993), which is based on the popular Cox proportional hazards model, is the analysis technique used most frequently for recurrent time-to-event data (Cox, 1972). Whether the observed event times relate to the same patient or various patients, the Andersen-Gill model presupposes independence between all observed event times. The frailty is additionally incorporated into the model to account for variability brought on by unobserved subject-specific characteristics that are not otherwise taken into account by the other predictors. These hidden subject-specific variables might contribute to within-subject correlation. We refer to each cluster of observations (i.e., subject) as having the same level of frailty by using the term "shared frailty." By expanding the shared frailty model, Pickles et al. (1994) and Yashin et al. (1995) allowed for various but associated frailties among observations within a group. Frailty models take into account unobserved heterogeneity that develops in a data-set as a result of some observations being more prone to failure and, thus, more fragile than others(Pickles et al., 1994; Yashin et al., 1995).

The joint frailty model, the piecewise-constant hazards model, and the shared random effects model are some common joint modeling approaches for recurring and terminal occurrences. The choice of model depends on the particular research objective and the nature of the data, as these models make varying assumptions about the link between recurrent and terminal occurrences. One

feature of joint modeling is that including data from both types of occurrences, it can result in more effective utilization of data. Moreover, biases that could develop when examining recurrent and terminal events separately can be avoided via joint modeling.

1.2 **Statement of the problem**

Prostate cancer, which accounts for 7.3% of all new instances of cancer in males, is the second most often diagnosed cancer in men globally (GLOBOCAN 2020). With the highest rates in North America, Europe, and Australia, and the lowest rates in Asia and Africa, the incidence and death rates range significantly across different populations and geographical areas. The most frequent disease among men in many African nations is prostate cancer, which is a serious public health issue throughout the continent. Prostate cancer is a serious health issue with a significant incidence of morbidity and mortality in sub-Saharan Africa. Prostate cancer is a growing public health concern in Ethiopia since it is linked to significant morbidity and mortality.

Prostate cancer is predicted to cause 130,000 new cases and 65,000 deaths in Africa in 2020, with age-standardized incidence rates of 22.7 per 100,000 men and 13.5 per 100,000 men, respectively (Ferlay et al., 2019b). Prostate cancer is projected to have caused 60,000 new cases and 30,000 deaths in sub-Saharan Africa, with age-standardized incidence rates of 19.8 per 100,000 men and 12.0 per 100,000 men, respectively. Also, it is predicted to have caused 2,500 new cases and 1,600 deaths in Ethiopia, with age-standardized incidence rates of 4.1 per 100,000 males and 2.6 per 100,000 men, respectively (Ferlay et al., 2019b).

Even though there are studies conducted on Awareness of prostate cancer and its associated factors among Ethiopian men based on cross-sectional data by (Gebru et al., 2021), Prostate Cancer Screening Practice and Associated Factors Among Men in Public Health Facilities of Hossana Town, Ethiopia by (Shanko et al., 2022) using Logistic Regression model on cross-sectional data and a retrospective study conducted by (Beksisa et al., 2020) on the Survival and prognostic determinants of prostate cancer patients in Tikur Anbessa Specialized Hospital using single event survival analysis model, among the few pieces of literature available on prostate cancer studies. To the best of my knowledge, Ethiopia has hardly any studies on the joint frailty modeling of recurrent prostate cancer events and terminal events (death) among prostate cancer patients. Thus, this thesis' goal is to close that gap. Thereby, the study has tried to respond to the following fundamental research questions.

- Which variables have a substantial impact on prostate cancer recurrence and the fatal outcome (death) brought on by recurrence?
- Additionally, how does the correlation between prostate cancer recurrence and terminal events (death) hazard behave?

1.3 **Objectives of the study**

1.3.1 **General objective**

- The study's general objective was to investigate the factors that affect prostate cancer recurrence and terminal events (death) jointly in the context of Tikur Anbessa specialized hospital.

1.3.2 **Specific objectives**

- To determine the important risk variables linked to prostate cancer recurrence and death.
- To evaluate the correlation between a hazard of terminal and recurrent event of prostate cancer patients.

1.4 **Significance of the study**

By identifying, factors associated with the recurrence of the disease and improving survival from the terminal event (death), studying the recurrence of prostate cancer and death owing to the recurrence is a technique for overcoming the problem of health in society. The potential risk factors for prostate cancer recurrence and the link with death were evaluated using a model that takes into account both recurrent episodes and death from recurrence. Several prevention and treatment programs for men could be offered based on this paradigm. The findings of this study would aid in reducing the recurrence of prostate cancer by increasing societal awareness of the risk factors for the prostate cancer recurrence and by increasing the likelihood that survivors would survive after a recurrence. It would also aid policymakers in formulating policies, particularly those that are related to health.

2. **Literature review**

2.1 **Overview of prostate Cancer Recurrence**

The male reproductive system includes a diminutive endocrine organ recognized as the prostate, which is typically the site of origin for prostate cancer. The prostate gland secretes a component

of semen referred to as prostatic fluid. Prostate cancer arises from the uncontrolled proliferation of prostate gland cells, resulting in the development of a tumor. Although prostate cancer typically has an indolent course, affecting mostly males, it is the most common type of cancer. Nevertheless, in certain circumstances, it may progress aggressively and metastasize to other anatomical locations (American Cancer Society, 2022b).

Prostate cancer represents 7.3% of all male malignancies diagnosed worldwide, ranking as the second most commonly diagnosed cancer in males. It is estimated that in 2020, prostate cancer will give rise to 1.4 million new cases and 375,000 deaths across the globe. The regions with the highest prostate cancer incidence rates are Australia/New Zealand, North America, and Western Europe, while the lowest incidence rates are observed in Asia and Africa (Rawla, 2019). In the year 2020, it was estimated that 92,000 new occurrences of prostate cancer arose in African males, resulting in 52,000 fatalities from the ailment (Ferlay et al., 2019b). Despite a lower incidence of prostate cancer in Africa than in other global regions, the disease is expanding rapidly due to demographic and aging trends. Sub-Saharan Africa exhibits some of the world's most elevated mortality rates from prostate cancer, with an estimated 27,000 novel cases and 16,000 deaths (Fitzmaurice et al., 2018). One factor contributing to this is restricted access to prostate cancer screening, diagnostic, and therapeutic interventions.

2.2 **Risk factors of prostate cancer recurrence**

Recurrence of prostate cancer, which happens when the disease reappears after initial treatment, can happen following treatment. The chance of prostate cancer recurrence may be increased by several factors.

- **Gleason Score**: The Gleason score functions as a hierarchical classification scheme employed to assess the malignant potential of prostate cancer cells. Specifically, a higher Gleason score corresponds to an increased likelihood of recurrence following therapeutic intervention, as it is indicative of a more aggressive cancer phenotype. A study conducted by D'Amico and colleagues has demonstrated that patients who exhibit a Gleason score of 8 or higher after radical prostatectomy is at a substantially elevated risk of experiencing

biochemical recurrence, defined as the presence of a prostate-specific antigen (PSA) level greater than 0.2 ng/mL. This research emphasizes the clinical significance of the Gleason scoring system as a prognostic tool for disease management and highlights the importance of accurate and precise risk stratification to optimize therapeutic decision-making (D'Amico et al., 1998).

- **Tumor Stage**: The tumor stage at the time of diagnosis is an indispensable factor in predicting the probability of cancer recurrence. A higher tumor stage, indicating a greater degree of malignancy, is positively associated with an increased likelihood of cancer recurrence following treatment. The findings from the Pound et al. investigation indicate that patients diagnosed with stage T3b or T4 prostate cancer have a higher incidence of prostate-specific antigen (PSA) recurrence after undergoing radical prostatectomy when compared to patients with stage T2 disease. Specifically, individuals with T3b or T4 disease exhibit a 5-year PSA recurrence-free survival rate of 41%, whereas those with T2 disease exhibit a significantly higher rate of 70%. This research underscores the importance of accurate staging in the management of prostate cancer, as it can significantly impact treatment selection and patient outcomes(Pound et al., 1999).
- **PSA Level**: The production of prostate-specific antigen (PSA) is a characteristic of prostate cells. A heightened risk of prostate cancer recurrence after therapy is associated with elevated levels of PSA in the bloodstream at the point of diagnosis. As such, PSA levels serve as a significant marker for disease prognosis and response to treatment. Freedland and colleagues have reported that patients who present with a pretreatment prostate-specific antigen (PSA) level greater than 20 ng/mL are at a substantially increased risk of experiencing biochemical recurrence following radical prostatectomy when compared to those with a PSA level less than 10 ng/mL. Specifically, individuals with a PSA level of >20 ng/mL exhibit a 5-year biochemical recurrence-free survival rate of 34%, while those with a PSA level of 10 ng/mL have a much higher survival rate of 77%. These findings underscore the clinical significance of PSA as a prognostic marker for prostate cancer and highlight the importance of accurate risk stratification for effective therapeutic decision-making(Freedland et al., 2003)
- **Surgical margins**: The observation of cancerous cells along the periphery of resected tissue after prostate surgery suggests that residual cancer cells may be present. This

phenomenon increases the potential for the recurrence of cancer post-surgery. The term used to describe this is positive surgical margins, which refer to the presence of cancer cells at the cut edges of the removed tissue. The Epstein et al. trial has revealed that patients who exhibit positive surgical margins after undergoing radical prostatectomy are at a substantially heightened risk of experiencing biochemical recurrence when compared to those with negative margins. Specifically, individuals with positive margins exhibit a 5-year biochemical recurrence-free survival rate of 35%, while those with negative margins have a significantly higher rate of 68%. These results emphasize the importance of achieving negative surgical margins to minimize the risk of prostate cancer recurrence following surgical intervention and highlight the clinical significance of accurate surgical technique and margin assessment in disease management (Epstein et al., 1994).

- **Radiation therapy**: Radiation therapy is a commonly used therapeutic modality for prostate cancer. However, if the delivery of radiation therapy is suboptimal, it may increase the likelihood of cancer recurrence. This underscores the importance of ensuring that the radiation therapy is administered with precision and accuracy to minimize the potential for disease recurrence. The study conducted by Michalski and colleagues has demonstrated that patients with intermediate-risk prostate cancer who undergo dose-escalated radiation therapy experience a significantly decreased probability of biochemical recurrence when compared to those treated with standard-dose radiation therapy. Specifically, individuals who receive dose-escalated radiation therapy exhibit a 5-year biochemical recurrence-free survival rate of 86%, whereas those who receive standard-dose radiation therapy exhibit a markedly lower rate of 73%. These findings underscore the clinical importance of tailored treatment strategies based on disease risk assessment and highlight the potential benefits of dose-escalated radiation therapy for optimizing prostate cancer management outcomes (Michalski et al., 2018).
- **BMI**: The likelihood of prostate cancer returning after therapy has been linked to higher BMI. For instance, research indicated that following radical prostatectomy, men with a BMI of 30 or above had a greater chance of biochemical recurrence than men with a lower BMI. Men with a BMI of 30 or above had a 61% chance of surviving 5 years without experiencing a biochemical recurrence, compared to 78% of men with a lower BMI (Freedland et al., 2008).

- **Age**: Several studies have suggested a correlation between the age at diagnosis and an elevated risk of prostate cancer recurrence following therapy. For example, the research conducted by Stephenson AJ, Scardino PT, Eastham JA, et al. highlights those men aged 70 years or older are at a higher risk of experiencing biochemical recurrence following radical prostatectomy when compared to those aged 60 years or younger. Specifically, men in the older age group exhibit a 10-year biochemical recurrence-free survival rate of 39%, whereas their younger counterparts have a notably higher rate of 66%. These results underscore the clinical significance of age as a prognostic factor in the management of prostate cancer and emphasize the need for individualized treatment strategies based on patient's age and other clinical variables (Stephenson et al., 2006).
- **Diabetes**: The probability of prostate cancer recurrence after therapy has been validated to be more probable in individuals with diabetes mellitus. As an illustration, the investigation conducted by Kyle A. Richards et al. evinced that male with diabetes mellitus have a heightened hazard of biochemical recurrence after undergoing radical prostatectomy in comparison to males devoid of diabetes mellitus (HR 1.49, 95% CI 1.18-1.87)(Richards et al., 2018).

2.3 **Overview of Recurrent event Models**

Therneau's (2000) and Lawless's (2003) works delve into the topic of recurrent events, in which individuals are at risk of experiencing a sequence of events over time, such as repeated hospitalizations. This kind of data differs from other types of data that may only measure an event occurrence once, like in Poisson or logistic regression models. To appropriately model recurrent event data using proportional hazard models, one may refer to the aforementioned works by Therneau (2000) and Lawless (2003)(Lawless, 2002; Therneau & Grambsch, 2000). In the context of recurrent event data analysis, the approach utilized depends on the assumptions made about the nature of the events. When repeated events are assumed to be identical, the counting process approach is employed. However, if the events are heterogeneous and differ in terms of disease categories or the sequence in which they occur is important, the analysis of the data requires a different approach. Therefore, understanding the underlying assumptions about the nature of the events is crucial in selecting an appropriate methodology for the analysis of recurrent event data (Andersen et al., 1993).

The hazard function in Cox and parametric models may depend on latent or unmeasured variables, which can introduce bias in the estimates of regression coefficients. To address this issue and provide a more accurate approach for modeling the recurrence of diseases, frailty models have been proposed. These models offer a solution for explaining the random variation in survival function that arises from multiple environmental, genetic, and unmeasured risk factors. The use of frailty models has become increasingly popular in medical research as they allow for the incorporation of individual-level heterogeneity in the analysis of correlated event data and can help to identify sources of variation in the occurrence of events beyond observed covariates.

The Andersen-Gill model, which is also referred to as the Prentice, Williams, and Peterson models, is a variant of the Cox proportional hazards model. This statistical method is specifically designed for the analysis of recurrent event data. Unlike the Cox model, the Andersen-Gill model accounts for the presence of multiple events per subject through the incorporation of time-dependent covariates. As such, this model is highly useful for investigating complex phenomena in a variety of academic fields (Andersen et al., 1993). Anderson KM and colleagues utilized the Andersen-Gill model to develop an updated coronary risk profile. This approach, which is based on the principles of survival analysis, provides a valuable framework for analyzing recurrent event data

and incorporating time-dependent covariates. By incorporating the Andersen-Gill model into their risk profile, the authors were able to capture the complex interdependencies among various risk factors and develop a more comprehensive tool for predicting coronary disease risk. A paper describing their updated risk profile is currently in circulation and has been announced to medical professionals, who may benefit from its application in clinical practice (Anderson et al., 1991).

The Prentice-Williams-Peterson (PWP) model is a survival analysis model that is developed to analyze recurrent event data. Specifically, this statistical model is an extension of the Cox proportional hazards model, which is a widely used method for analyzing time-to-event data. The PWP model builds upon the fundamental principles of the Cox model by allowing for the analysis of multiple events per subject, making it a valuable tool for investigating complex phenomena in a range of academic disciplines (Prentice et al., 1981). Shahid Ullah and colleagues employed the Prentice-Williams-Peterson (PWP) model, among other statistical modeling techniques, in their paper titled "Statistical modeling for recurrent events: an application to sports injuries." The PWP model, which is an extension of the Cox proportional hazards model, is particularly useful for analyzing recurrent event data and incorporating time-dependent covariates. By incorporating this model into their statistical analysis, along with other relevant techniques, the authors were able to provide a comprehensive and insightful examination of the factors associated with sports injuries (Ullah et al., 2014).

In the context of survival analysis, correlated time-to-event data can often be attributed to the presence of unobserved heterogeneity or frailty. To address this issue, nested frailty models are commonly employed. These models provide a valuable framework for analyzing clustered or hierarchical data, such as data collected from research involving patients nested within hospitals or clinics. By incorporating frailty parameters that account for the shared characteristics among subjects within clusters, nested frailty models can effectively capture the complex interdependencies present in the data. Consequently, these models have become a preferred method for conducting survival analysis in a range of academic fields, including epidemiology, biostatistics, and health services research (Rabe-Hesketh & Skrondal, 2005).

If a notable proportion of patients have been effectively treated and deemed free of recurrence, the population will consist of a blend of susceptible and insusceptible individuals. In such instances, traditional survival analysis methods, such as the Cox proportional hazards model, may not be

suitable as they presuppose equal susceptibility to the ailment across all subjects and the eventual occurrence of the event over a prolonged follow-up period. Cure models, on the other hand, take into account the presence of a cured fraction within the population, and are commonly applied in scenarios where a pronounced distinction exists between susceptible and insusceptible individuals. Price and Manatunga proposed utilizing cure frailty models to examine the reappearance of leukemia in cases where there is a cured fraction. This approach takes into account the presence of individuals who have been successfully treated and are free from the ailment and employs frailty models to account for heterogeneity in susceptibility across the population (Price & Manatunga, 2001).

A joint frailty model is a statistical approach that is used to analyze recurrent event data and the occurrence of a terminal event. In this model, both types of events are modeled using a shared frailty parameter, which represents the unobserved heterogeneity or correlation between the events within a subject. The shared frailty parameter is assumed to follow a distribution, such as a gamma or log-normal distribution that reflects the variability in the risk of the events across the study population. This statistical approach is particularly useful for investigating complex phenomena in various academic fields, including epidemiology, biostatistics, and health services research (Wei et al., 1989). Through the joint modeling of recurrent and terminal events, joint models can furnish more precise and effective estimates of the association between the two event types, along with the risk factors that are connected with them, in contrast to distinct models for each type of event. “Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data”, a published study by Chiung-Yu Huang and Mei-Cheng Wang, make use of the joint frailty model(Huang & Wang, 2004). V. Rondeau applied extensions of the frailty model to scrutinize data on recurrent events, encompassing cure frailty models that accommodate a combination of susceptible and insusceptible subjects for the event of interest, nested frailty models that address data clustered at multiple hierarchical levels, and joint frailty models that jointly examine recurrent events and death, as exemplified in the context of breast cancer analysis (Rondeau, 2010a).

3. **Data and methodology**

3.1 **Description of the study area**

The present investigation has utilized a dataset comprising information on the recurrence of prostate cancer and associated death. This dataset has been procured from the records maintained by Tikur Anbessa specialized hospital (TASH). The current investigation was conducted at the Tikur Anbessa specialized hospital (TASH), situated in the capital city of Ethiopia, Addis Ababa. Tikur Anbessa specialized hospital (TASH), which commenced operation in 1972, has been established as a healthcare provider primarily for the population of Addis Ababa city and its neighboring areas, but it attracts patients from various regions across Ethiopia. The medical institution offers various outpatient and inpatient services across several specialized units, including but not limited to surgery, gynecology and obstetrics, internal medicine, pediatrics, ophthalmology, psychiatry, radiology, and pathology.

3.1.1 **Study population**

The current investigation has used a retrospective longitudinal study that endeavors to identify instances of prostate cancer recurrence and the resulting fatalities, utilizing hospital records sourced from the TASH oncology center. The study population have comprised all male patients who were diagnosed with prostate cancer and registered at TASH during the period spanning January 1, 2018, to January 30, 2021.

3.1.2 **Inclusion and exclusion criteria**

- **Inclusion criteria:** The eligible participants for this study were newly diagnosed or referred prostate cancer patients, between the period of January 1, 2018, and January 30, 2021, who were registered with detailed and comprehensive personal and medical information. The said information would be required to be recorded in either the patient's medical chart or identification card and would serve as the criteria for inclusion in the study.
- **Exclusion criteria:** Patients whose initial diagnosis predates January 2018 were excluded from participation in the study, irrespective of their registration date. Furthermore, medical records or charts that do not contain comprehensive details such as histology reports, PSA levels, or reports on the stage of cancer were deemed incomplete and therefore, excluded from the investigation.

3.1.3 **Sample size determination**

The choice of sample was crucial since using an excessively large sample size implies resource waste and using an excessively small sample size diminishes the value of the findings. There are several considerations or factors that one must make to have the ideal sample size. The research's goal, its design, budgetary restrictions, the level of accuracy needed for generalization, and others are only a few of the problems. For this research all prostate cancer patients from January 1, 2018 up to January 30, 2021 were used as a sample which is 222 prostate cancer patients following treatment in TASH.

3.1.4 **Data collection procedure**

Upon obtaining the necessary approval from the oncology division at TASH in compliance with the relevant ethical guidelines, a trained enumerator and the primary investigator has collected the hospital-based secondary data for the study.

3.1.5 **Data structure for modeling prostate cancer recurrence and death jointly**

The prostate cancer data were meticulously scrutinized to identify and measure the duration between successive instances of prostate cancer, including subsequent recurrences (also known as recurrent event data), as well as the time of death. The required data structure for modeling the recurring events concerning the duration until the occurrence of recurrent event and subsequent terminal events (death) is displayed in the table below, which includes information on death. The follow-up period commences from the outset of the study (i.e., January 1, 2018) until January 30, 2021, and includes measurements of the duration between the start date and the initial recurrence, as well as the durations of subsequent recurrences (if any).

In addition, the study also took into account the time of the final censoring for patients who did not experience prostate cancer recurrence, and the time of the terminal event (i.e., death), to model the prostate cancer recurrence and death data in conjunction with one another. When modeling the data on recurrent and terminal events, this study included data from patients with prostate cancer who experienced one or more recurrences. The duration between two successive instances of prostate cancer recurrence for each patient were calculated to represent the time interval between the repeated episodes of recurrence.

Table 1: data layout for prostate cancer recurrence and terminal event (death) due to PCA recurrence.

Subject ID	Start	End	time	Recurrence indicator	Terminal event(death) indicator	Treatment taken	Clinical stage
1	0	8	8	1	0	Chemo	II
1	10	16	6	1	0	surgery	III
1	23	32	9	0	1	nothing	IV
2	0	8	8	1	0	surgery	II
2	12	18	6	0	1	nothing	IV
3	0	7	7	1	0	surgery	II
3	14	24	10	1	0	chemo	III
3	26	32	6	0	0	radiation	III

The start and end of a pair of variables are used to specify the length of time between prostate cancer recurrences.

- **Start**: The temporal origin of an interval is denoted by its commencement time expressed in months. The initiation of treatment for an individual with a deviant prostate gland marks the point of data registration on reappearing occurrences. In the context of prostate cancer recurrence, the inception of the time frame is conventionally designated as zero.
- **End**: The temporal reference point for an event, either its incidence, censoring (excluding cases of death), or the occurrence of the final event (death) with regards to prostate cancer, is denoted by its temporal measure expressed in months.

- **Time**: The summation of the values recorded in the Start and End columns yields the duration of exposure in months, which served as the basis for constructing the frailty models.

The recurrence indicator, a dichotomous variable, signifies the occurrence of a prostate cancer recurrence, with the value of "1" indicating its presence and "0" indicating its absence, taking into account the censoring of events not attributable to prostate-related death. Additionally, the terminal event indicator characterizes the occurrence of death caused by a recurrence of prostate cancer, taking the value of either "1" or "0". For this study, the censoring mechanism for prostate cancer-related death is not considered non-informative for the process of recurring events. The study was expected to conclude with a cohort of prostate cancer patients, some of whom may have dropped out, passed away due to non-cancer-related causes, or remained recurrence-free. This implies that the survival data is subjected to random right censoring.

3.2 **Study variables**

3.2.1 **Response variable**

The outcome variables for this study is the time until the recurrence of prostate cancer and the time until the occurrence of the terminal event (death as a result of prostate cancer), evaluated in terms of the number of months.

3.2.2 **Explanatory variables**

The following predictor variables were evaluated for their potential influence on prostate cancer recurrence and the occurrence of terminal events related to prostate cancer recurrence.

1. Age at diagnosis (in years) (<60,60 and greater)
2. Serum PSA level (ng/ml) (0-4,4.1-10,10.1-20, >=20)
3. Clinical stage at diagnosis (stageI, stageII, stageIII, stageIV)
4. The treatment taken (Chemotherapy, Radiation, Surgery, hormone therapy, or a combination of the two or more)
5. Gleason score (<=6,7, 8, and above)
6. Distant metastasis (yes, no)
7. Ever smoking (yes, no)

3.2.3 **Description of variables**

- Prostate cancer recurrence: The reappearance of prostate cancer following initial treatment is defined as recurrent prostate cancer. It should be noted that recurrent prostate cancer denotes the reappearance of same cancer that was previously diagnosed, rather than the onset of a new disease.
- Terminal event (death): The terminal event, in this context, is specifically concerned with the occurrence of death caused by a recurrence of prostate cancer, which ultimately concludes the follow-up of the patient from the risk set.
- The treatment taken: Considering tumor characteristics such as size and spread, as well as the patient's preference, the typical treatment options for men with prostate cancer include surgical intervention, radiotherapy, and chemotherapy.
- Surgery: Surgery is a frequently employed treatment modality for individuals diagnosed with prostate cancer. It entails the extirpation of the prostate gland and the adjoining tissues through a surgical procedure. Surgical intervention may be recommended as a first-line therapy for prostate cancer localized within the prostate gland, or as a subsequent treatment following other therapeutic modalities such as radiotherapy or hormone therapy.
- Hormone therapy: It is a frequently employed treatment option for individuals diagnosed with prostate cancer, also recognized as androgen deprivation therapy (ADT) or androgen suppression therapy. This therapeutic approach involves the reduction of androgen levels in the body, particularly testosterone, which can promote the proliferation of prostate cancer cells. Hormone therapy may be employed as a monotherapy or in conjunction with other treatment modalities such as radiotherapy or chemotherapy.
- Radiation: Radiotherapy entails the utilization of high-energy radiation to specifically target and eliminate cancer cells in the prostate gland. Radiotherapy may be used as a primary treatment modality for localized prostate cancer or as a secondary intervention after surgery in cases where residual cancer cells remain. It may also be employed as a palliative measure for symptom management or to decelerate the progression of advanced prostate cancer.

3.3 **Methods of data analysis**

3.3.1 **Descriptive statistics**

To assess the differences in survival functions among multiple groups of prostate cancer patients, non-parametric approaches, such as Kaplan-Meier plots, were employed to characterize survival data. Moreover, the data was synthesized by creating a frequency distribution table that summarizes the study variables.

3.3.2 **Survival data analysis**

Survival analysis refers to a set of statistical methods used to analyze data in which the outcome variable of concern is the time elapsed until a specific event takes place such as the recurrence of prostate cancer or the time until death due to prostate cancer. This "time" is measured in units such as years, months, weeks, or days, from the start of an individual's follow-up period until the occurrence of the event of interest. This event may refer to death, the incidence of a disease, a relapse from remission, recovery, recurrence, or any other pre-specified experience that an individual may undergo (Aalen, 1978). In instances where study subjects become unavailable for follow-up or the observation period is constrained, resulting in certain individuals not encountering the event of interest during the study period. The application of survival analysis is essential in contrast to alternative statistical methods. This is because survival analysis can accommodate incomplete observations and provide valid estimates of the time-to-event outcome of interest by incorporating the censoring process, where the observation is terminated before the event occurs. Hence, utilizing survival analysis for prostate cancer recurrence and death significantly enhanced the accuracy and dependability of the statistical inference in this research.

The analytical challenge of censoring is a primary consideration in most survival analyses. In the context of prostate cancer recurrence and death due to PCA recurrence, censorship occurs when information regarding an individual's relapse period is incomplete, except for the terminal event of interest (namely, death resulting from prostate recurrence). This work makes no use of either left or interval censoring approaches. Specifically, when the recording encompasses the period from the individual's start time up to a specific point before the end time, the survival time is deemed to be suitably censored. Right censoring may arise for various reasons in many cases, including but not limited to:

- Death from unrelated causes
- Loss of follow-up
- Termination of study

Survival function: The survivor function represents the likelihood that the survival time of an individual, selected at random, exceeds or equals a certain predetermined time. As such, it provides the probability of an individual surviving beyond a given time. The distribution of survival time is determined by the survivorship function, probability density function, and hazard function.

Suppose we have a random variable T that is associated with survival times, and let t denote a specific value of the random variable T. Further, let f(t) be the underlying probability density function of the survival time T. The survivor function, denoted by S(t), can be defined as follows:

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

Here, F(t) represents the cumulative distribution function, which gives the probability that a randomly selected subject has a survival time that is less than or equal to a specified time t. It can be expressed mathematically as follows:

$$F(t) = P(T \leq t) = \int_0^t f(u)du\ \,, t \geq 0 \quad (3)$$

The probability density function f(t) can be defined as follows:

$$f(t) = \frac{d}{dt}F(t) = \frac{-d}{dt}S(t) \qquad (4)$$

The hazard function can be defined as the instantaneous probability of an event occurring at a specific time t (per unit time), given that an individual has survived up to that time t without experiencing the event of interest (Kleinbaum & Klein, 2012). It is given by:

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

The cumulative hazard function can be defined as follows:

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

Thus $S(t) = e^{-\Lambda(t)}$

3.3.3 **Cox-proportional hazard model**

The analytical methodology for recurring data that is most commonly employed is the model proposed by Andersen and Gill (1982), which is built upon the Cox proportional hazards model, a widely utilized statistical framework (Cox, 1972) The Andersen-Gill model assumes that the observed event times are independent of each other, regardless of whether these events correspond to the same or different patients. The proportional hazards assumption posits that the likelihood of an individual experiencing an event at a given moment is proportional to the risk of that event occurring for another group of individuals at the same time. In addition to this assumption, the Cox proportional hazards (PH) model makes several other key assumptions. These include the assumptions that: the hazard ratio for two individuals with differing covariate values is not affected by the passage of time; time is measured on a continuous scale; censoring is random; and censoring is uninformative.

Kleinbaum and Klein (2012) suggest that recurrent survival data can be modeled using a Cox proportional hazards (PH) model, wherein each subject has a line of data corresponding to each recurrent event. In such cases, the counting process approach is typically employed, which is essentially an extension of the standard Cox PH model (Kleinbaum & Klein, 2012). In recurrent survival data, a subject may remain in the risk set for multiple time intervals, until their last interval, after which they are removed from the risk set. The hazard function for the jth individual in the i^{th} group can be derived as follows:

$$\lambda_{ij}(t) = \lambda_0(t)e^{\beta^t X_{ij}} \qquad (7)$$

Where the hazard function for the j^{th} individual in the i^{th} group (i =1..., G and j=1, ..., n_i) is expressed as a function of the baseline hazard function, $\lambda_0(t)$, the vector of regression coefficients of factors such as PSA level, CLINICAL stage level, Smoking status, treatments and others, β (with p rows and 1 column), and X_{ij}, which represents the value of the covariate vector described in study variable section, for the j^{th} individual in the i^{th} group that include as PSA level, CLINICAL stage level, Smoking status, treatments and others.

3.3.4 **Frailty models**

The underlying assumption of the Cox proportional hazards model, as proposed by (Cox, 1972), is that the observations are independent of one another, with regards to the variables being analyzed. However, this assumption may not always hold, particularly in epidemiological studies where failure times may be correlated within families or geographic units. Certain unmeasured factors, such as genetic information or environmental exposures, may influence the occurrence being studied, leading to non-independence among the observations. As a result, such studies may require specialized techniques to account for the potential correlation among observations. In an alternate scenario, when events occur repeatedly within the same individual throughout the observation period, the resulting data may exhibit correlation. To address this, frailty models can be utilized, which introduce an unobserved variable treated as a random effect, to account for shared characteristics and dependencies among the recurrent events (Andersen et al., 1993; Clayton, 1978).

The notion of "frailty" was originally introduced by (Vaupel et al., 1979) in the context of univariate survival models. Subsequently, (Clayton, 1978) utilized the approach in a groundbreaking study examining the occurrence of chronic disease in families, although he did not use the term "frailty" explicitly in his work. Later, the frailty model was extended to multivariate survival data, incorporating a shared and unobserved random effect to account for dependence in the data. A frailty model is a statistical model where an unobserved random effect is introduced to account for unmeasured heterogeneity in a population that affects the hazard function for time to recurrence. The random effect is assumed to have a multiplicative effect on the baseline hazard function. This type of model is often used in the analysis of recurrent events or correlated data where there is evidence of unobserved factors that may influence the risk of experiencing an event.

3.3.5 **Shared frailty model**

The concept of multivariate frailty models is a natural progression from the univariate frailty model, allowing individuals to share a common frailty value. In the shared frailty model, a key assumption is that both members of a pair share the same frailty W, which is the underlying reason for its nomenclature. The multivariate frailty model, which enables the sharing of frailty values among individuals, was first introduced by (Clayton, 1978) and has since been subject to

comprehensive examination in works such as (Therneau & Grambsch, 2000), as well as (Duchateau et al., 2003). Frailty models can be categorized as either individual-specific or group-specific depending on the nature of the frailties. The former pertains to frailties that are unique to each individual, while the latter encompasses frailties that are shared among groups of subjects, such as those within a family unit. Shared frailty models are particularly suitable for modeling such group-specific frailties, which can provide insights into the level of correlation within groups.

3.3.6 **Shared gamma frailty model**

The gamma distribution $\Gamma(\lambda, k)$ has gained significant traction as a suitable frailty distribution in various applications(Duchateau et al., 2003). The two-parameter gamma density function is expressed as:

$$f_{\omega}(\omega) = \frac{\omega^{\lambda}\omega^{\lambda-1}e^{-k\omega}}{\Gamma(\lambda)} \qquad (8)$$

With scaling parameter k > 0 and shape parameter λ > 0, given by the equivalent Laplace transform.

$$L(s) = \int_0^{\infty} \exp(\omega s)\, f_w(\omega) d\omega \quad (9)$$

$$= \int_0^{\infty} \exp(-\omega s) \frac{\gamma^{\lambda}\omega^{\lambda-1}\exp(-\gamma\omega)}{\Gamma(\lambda)} d\omega$$

$$= \frac{\gamma^{\lambda}}{\Gamma(\lambda)} \int_0^{\infty} \exp\big(\omega(s+\gamma)\big)\, \omega^{\lambda-1}\, d\omega$$

Letting y=ω(s+γ); $\omega = \frac{y}{s+\gamma}$ and $d\omega = \frac{dy}{s+\gamma}$ We obtain by substituting ω and dω,

$$= \frac{\gamma^{\lambda}}{\Gamma(\lambda)} \int_0^{\infty} \exp -y \left\{\frac{y}{s+\gamma}\right\}^{\lambda-1} \frac{dy}{s+\gamma}$$

$$= \frac{\gamma^{\lambda}}{(s+\gamma^{\lambda})\Gamma(\lambda)} \int_0^{\infty} \exp -y\{y\}^{\lambda-1}\, dy$$

Where by $\Gamma(\lambda) = \int_0^\infty \exp -y\{y\}^{\lambda-1}\, dy$, hence

$$L(s) = \frac{\gamma^\lambda}{(s+\gamma)^\lambda \Gamma(\lambda)} * \Gamma(\lambda)$$

$$L(s) = \frac{\gamma^\lambda}{(s+\gamma)^\lambda}$$

$$L(s) = (\gamma)^\lambda (s+\gamma)^{-\lambda} \quad (10)$$

The first and second derivatives of the Laplace transform can be used to determine the mean and variance for L(s), which exist in the vicinity of zero.

$$L^{(1)}(s) = -\lambda(\gamma)^\lambda (s+\gamma)^{-\lambda-1}$$

$$L^{(2)}(s) = \lambda(\lambda+1)(\gamma)^\lambda (s+\gamma)^{-\lambda-2}$$

When these derivatives are evaluated at s=0, the expected mean and variance are;

$$E(w) = (-1)L^{(1)}(0) = \frac{\lambda}{\gamma} \quad (11)$$

$$var(w) = L^{(2)}(0) - \left(-L^{(1)}(0)\right)^2 = \frac{\lambda(\lambda+1)}{\gamma^2} - \left(\frac{\lambda}{\gamma}\right)^2 = \frac{\lambda}{\gamma}$$

When constraint k=γ is applied to gamma frailty models, the expectation is set to 1. The frailty variable's variance is then equal to $\frac{1}{\gamma}$. Assuming that the frailty term ω is a gamma with E(ω)=1 and Var(ω)=θ then λ=γ=$\frac{1}{\theta}$. Thus, the distribution function of the frailty component 'ω' is a gamma distribution with one parameter,

$$f_\omega(\omega) = \frac{\omega^{\frac{1}{\theta}} \exp(\frac{-\omega}{\theta})}{\theta^{\frac{1}{\theta}} \Gamma(\theta)} \qquad (12)$$

For values of θ > 0 and ω> 0, it implies that individuals in group i are frail, whereas ω < 1 indicates that individuals are less frail and have a lower risk. The Laplace transform associated with this is as follows:

$$L(s) = (1+\theta s)^{\frac{1}{\theta}} \qquad (13)$$

In this scenario, it is assumed that individuals j within cluster i share the same frailty value, ω_i. The survival function for the gamma frailty distribution without conditioning on any particular value of the frailty parameter is expressed as:

$$S_\theta(t) = [1 - \theta ln\{s(t)\}]^{\frac{1}{\theta}} \quad , \theta > 0 \qquad (14)$$

The hazard function for the gamma frailty distribution without conditioning on any specific value of the frailty parameter is expressed as:

$$\lambda_\theta(t) = \lambda(t)[1 - \theta ln\{s(t)\}]^{-1} \qquad (15)$$

Where S (t) and λ (t) are the survival and hazard functions in the distributions. The shared

gamma Frailty model (conditional hazard) for individual j in cluster i is;

$$\lambda_{ij}(t|X_{ij}, \omega_i) = \omega_i \lambda_0(t) e^{X'_{ij}\beta} = \omega_i \hbar(t_{ij}) \qquad (16)$$

The $\omega_i's$ are independent and identically distributed gamma distribution, like in the univariate Frailty models.

3.3.7 **Shared log-normal frailty model**

The log-normal frailty distribution holds significant importance in statistical modeling due to its widespread application in mixed models. Its popularity can be attributed to its association with the standard assumption that the random effects conform to a normal distribution. Let $\omega \sim N(1, \sigma^2)$ be a normally distributed random effect and let the frailty be given by $Z=e^{\omega}$. The corresponding frailty is characterized by a log-normal distribution, and its mathematical function takes the following form:

$$f_z(Z) = \frac{1}{Z\sqrt{2\pi\sigma^2}} \exp\left(\frac{-((\log Z)^2)}{2\sigma^2}\right) \ , \sigma > 0, z > 0 \quad (17)$$

As a result, in the gamma frailty model, the parameter θ represents the variance of the frailty Z. In contrast, in the log-normal model, the variance of the random effect $\omega = \ln(Z)$ is denoted by σ^2. It is not appropriate to make direct comparisons between both expressions. Additionally, in the log-normal model, the expected value of the frailty variable is typically not equal to one, even though the expected value of the random effect ω is zero. The parametric frailty model that incorporates shared log-normal distribution assumes the following structure:

$$\lambda_{ij}\left(t|X_{ij},Z_i\right)=\lambda_0(t)e^{\omega_i}e^{X'_{ij}\beta}=\lambda_0(t)Z_ie^{X'_{ij}\beta} \qquad (18)$$

The model denoted as (18) involves a conditional hazard function that is reliant upon the independent Z_i, where $Z_i = e^{\omega_i}$, and is commonly referred to as the "frailty" term. These Z_i values, for i = 1..., G, are assumed to possess an identical density function, $f_z(.)$, thereby characterizing the shared log-normal frailty model. The estimation of unknown parameters is achieved via the utilization of both the Expectation-Maximization (EM) algorithm and the penalized likelihood approach, as expounded upon in (Therneau & Grambsch, 2000) work.

3.3.8 **Joint frailty model for prostate cancer recurrence and death**

The duration of a person's repeated event process may be influenced by other "terminating" events such as death, which is often associated with a greater likelihood of the recurrence of major events such as cancers and opportunistic infections. This presents a departure from the conventional statistical analysis assumption that the recurring event process is non-informatively censored by death. Therefore, joint modeling of recurring events and deaths must account for this dependence. (Liu et al., 2004) have devised a joint semi-parametric model that incorporates shared gamma frailty to estimate the intensity functions of both recurring events and death. The effect of frailty on the rates of recurrent events and death varies across different models.

We denote for prostate CA patient i (i = 1, . . ., N), Y_{ij} the j^{th} recurrent times (j = 1, . . ., n_i), C_i the censoring times (not by death), and D_i the death times. First, consider Y_{ij} as a time to event. T_{ij} = min (Y_{ij}, C_i, D_i) corresponds to each follow-up time and δ_{ij} is a binary indicator for prostate CA recurrence which is 0 if the observation is censored and if the subject died, or 1 if Y_{ij} is observed ($\delta_{ij} = I(T_{ij}=Y_{ij})$ where I(.) denotes indicator function). Similarly, we note T_i^* the last follow-up time for prostate CA patient i, which is either a time of censoring or a time of death (T_i^* = min (C_i, D_i)) and $\delta_i^* = I_{(T_i^*=D_i)}$. What we observe is ($T_{ij}$, δ_{ij}, δ_i^*). Adopting the framework proposed by (Liu et al., 2004), the joint frailty model for the hazard functions of prostate CA recurrence ($r_i(.)$) and death ($\lambda_i(.)$) can be expressed as:

$$r_{ij}(t|\omega_i)=\omega_i r_0(t)\exp\left(\beta_1'Z_{ij}(t)\right)=\omega_i r_{ij}(t) \quad (23)$$

$$\lambda_i(t|\omega_i)=\omega_i^{\alpha}\lambda_0(t)\exp\left(\beta_2'Z_i(t)\right)=\omega_i^{\alpha}h_i(t) \quad (24)$$

where $r_0(t)$ and $\lambda_0(t)$ are the baseline hazard for PCA recurrence and terminl event respectivly.

It is assumed that the impact of explanatory variables on prostate CA recurrence and death times differs. The parameters β_1 and β_2 are interpretable in the context of the subject's event history and mortality status, and represent the instantaneous probability of recurrent events and the terminal event, respectively, given the subject's history and current survival status. The model and estimation procedure are capable of accommodating external time-dependent covariates as described by (Kalbf'leisch & Prentice, 2002). Moreover, the number of prior recurrent events can also be viewed as an internal time-dependent covariate, which requires the subject's survival for its manifestation and provides direct information on the time to failure.

In this research it was assumed that the random effects ω_i (frailties) are independent. With a unit mean and variance θ, the gamma frailty density was used in this instance. The unobserved ω_i has an impact on both the prostate CA recurrence times and the death times, which accounts for the dependence between T_i^* and T_{ij} conditional on explanatory variables $Z_i(t)$, such as age at diagnosis, family history, serum PSA level, clinical stage, treatment taken, and others. The heterogeneity in the data caused by unobserved factors will be considered by the common frailty parameterω_i. In the traditional model, it is assumed that $\alpha = 0$ in the above equation (Rondeau & Joly, 2003), implying that $\lambda_i(t)$ is not dependent on ω_i, indicating that death (or the terminal event process) does not provide information on the recurrent prostate CA rate $r_i(t)$. In other words, the two rates, $\lambda_i(t)$ and $r_i(t)$ are not correlated conditional on covariates. When $\alpha = 1$, the impact of frailty is the same for both prostate CA recurrence and the death. However, when $\alpha > 1$, the rate of recurrent prostate CA and the rate of death is positively associated; higher frailty will increase the risk of both prostate CA recurrence and death.

3.3.9 **Parameter estimation of the joint frailty model**

Within the context of this research endeavor, we intend to employ the penalized likelihood approach. The aforementioned technique offers several advantages, the foremost of which is its ability to generate continuous estimations of hazard and intensity functions without imposing any parametric assumptions. Additionally, this approach obviates the need for function-based log-likelihood expressions, thereby facilitating the computation of unknown parameters.

3.3.10 **Penalized likelihood**

The utilization of the time-to-event metric facilitates the assimilation of time-dependent covariates, and the corresponding likelihood function necessitates the inclusion of deferred entries. Diverging

from the gamma frailty models that are shared as noted by (Rondeau & Joly, 2003), the complete log-likelihood of the collective frailty model does not assume a straightforward configuration due to the absence of a closed form for the integrals. Consequently, employing alternative frailty distributions such as log-normal or positive stable does not incur additional complexities. Furthermore, (Pickles, 1995)argue that the selection of frailty distribution should not significantly impact the obtained outcomes.

Construction of the full log-likelihood for the joint frailty model (23,24) with calendar timescale is put as follows:

We denote T_{ij} as the jth follow-up time for prostate CA patient i and δ_{ij} the failure indicator for the prostate CA recurrence. Similarly, we define T_i^* = min (D_i, C_i) the last follow-up time for prostate CA patient i and the death indicator δ_i^* = I ($D_i < C_i$).

The marginal contribution to the likelihood $L_i(r_0(i), \lambda_0(i), \beta, \alpha, \theta) = L_i(\phi)$, for subject i and for recurrent time j = 1…. n_i is $L_i(\phi) = \int L_i(\phi|\omega) f(\omega) d\omega$

1. The joint conditional distribution of the prostate cancer recurrence and death times given ω_i is the product of individual contributions:

$$L_i(\phi|\omega_i) = \prod_{j=1}^{n_i} \left[dR_i\left(T_{ij}|\omega_i\right)^{\delta_{ij}} * exp\left(-\omega \sum_{j=1}^{n_i} \int_{T_{ij-1}}^{T_{ij}} dR_i(t) \right) \right]$$

$$* \, d\Lambda_i(T_i^*|\omega_i)^{\delta_i^*} * \exp\left(-\omega^{\alpha} \int_0^{\infty} Y_i d\Lambda_i(t) \right) \quad (25)$$

2. The probability density function for the random effects ω is $(\omega) = \frac{\omega^{\frac{1}{\theta}-1} \exp(\frac{-\omega}{\theta})}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}}$.

3. Using the previous expressions, the i^{th} marginal contribution to the likelihood is obtained by integrating the random effects:

$$L_i(\phi) = \frac{\prod_{j=1}^{n_i} dR_i\left(T_{ij}\right)^{\delta_{ij}} * d\Lambda_i(T_i^*)^{\delta_i^*}}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}} * \int_0^{\infty} \omega^{\left(N_i^R(T_i^*) + \alpha\delta_i^* + \frac{1}{\theta} - 1\right)}$$

$$* \exp\left(-\omega \sum_{j=1}^{n_i} \int_{T_{ij-1}}^{T_{ij}} dR_i(t) - \omega^{\alpha} \int_0^{\infty} Y_i d\Lambda_i(t) - \frac{\omega}{\theta} \right) d\omega \quad (26)$$

4. We then get the full log-likelihood by $l(\phi) = \log \prod_{i=1}^{N} L_i(\phi)$

$$l(\phi) = \sum_{i}^{N} \left\{ \sum_{j}^{n_N} \delta_{ij} log r_i(T_{ij}) + \delta_i^* log \lambda_i(T_i^*) - log\Gamma\left(\frac{1}{\theta}\right) - \frac{1}{\theta} log\theta \right.$$
$$+ log \int_0^{\infty} \omega^{(N_i^R(T_i^*) + \alpha\delta_i^* + \frac{1}{\theta} - 1)} exp\left(-\omega \sum_{j=1}^{n_i} \int_0^{T_i^*} dR_i(t) \right.$$
$$\left. \left. - \omega^{\alpha} \int_0^{T_i^*} \mathrm{d}\Lambda_{\mathrm{i}}(\mathrm{t}) - \frac{\omega}{\theta} \right) d\omega \right\} \quad (27)$$

with $T_{i0} = 0$ and $T_{ini} = T_i^*$ (for each prostate CA patient, assuming that the last observation time is a censoring time or a death time and not a prostate relapse time), $\Lambda_i(t) = \int_0^t \lambda_i(u)du$ the cumulative hazard function for death, with joint frailty cumulative hazard function for death $\Lambda_i(.|\omega) = \omega^{\alpha}\Lambda_i(.)$, and $\mathrm{r}_i(t) = \int_0^t r_i(u)du$ the cumulative hazard function for prostate recurrence, with joint frailty cumulative hazard function for prostate CA recurrence $\mathrm{R}_i(.|\omega) = \omega\mathrm{R}_i(.)$. In the majority of cases, it is rational to anticipate that the baseline hazard functions will exhibit a smooth behavior, thereby rendering the application of piecewise constant models unsuitable for hazard function modeling. To incorporate such a priori knowledge, a likelihood penalty term will be employed, which assigns high values to functions that are characterized by roughness (Finbarr O & sullivant, 1988; Joly et al., 1998). The roughness penalty function is represented by the sum of two squared norms of the second derivative of the hazard functions (Finbarr O & sullivant, 1988). The penalized log-likelihood is thus defined as:

$$pl(r_0(.), \lambda_0(.), \beta, \alpha, \theta) = l(\phi) - K_1 \int_0^{\infty} r_0''^2(t)dt - K_2 \int_0^{\infty} \lambda_0''^2(t)\, dt \quad (28)$$

Where $l(r_0(.), \lambda_0(.), \beta, \alpha, \theta)$the full log likelihood is defined in (27) and $K \geq 0$ is a positive smoothing parameter that controls the trade-off between the data fit and the smoothness of the functions. Maximization of (28) defines the maximum penalized likelihood estimators (MPnLE) $\widehat{r_0}(t), \widehat{\lambda_0}(t), \widehat{\boldsymbol{\beta}}, \hat{\alpha}\ and\ \hat{\theta}$.

3.3.11 **Computational procedure (algorithm)**

This study used the robust Marquardt algorithm, developed by (Marquardt, 1963), to estimate the parameters of the joint frailty model. This algorithm is a hybrid method that combines the Newton-Raphson and steepest descent algorithms. It is more stable than the Newton-Raphson algorithm while maintaining its rapid convergence property in the proximity of the maximum (Fletcher, 2000). The iterative process is terminated when the difference between the log-likelihoods of two consecutive iterations is negligible, the coefficients are deemed stable, and the gradient is sufficiently small.

3.3.12 **Assessing model adequacy**

Irrespective of the model type employed or the selection of variables for inclusion, it was crucial to assess the degree to which the model fits the observed data. The goodness of fit of a joint frailty model depends on its ability to accurately capture the prostate CA recurrence and survival patterns inherent in the data. Evaluating the fit of a statistical model using residuals was a valuable tool. Residuals are an essential component in gauging the adequacy of a model, regardless of the analytical setting. The Cox-Snell residual, introduced by Cox and Snell in 1968, is the most commonly used in the analysis of survival data and it was also used in this research. It is defined as the residual value for the i^{th} individual in the sample:

$$r_{ci} = e^{\beta' x_{ij}} \widehat{H}_0(t_i) = \widehat{H}_i(t_i) = -\log\left(\hat{S}_0(t_i)\right) \quad (29)$$

Where $\widehat{H}_0(t_i)$ an estimate of the baseline cumulative hazard function, $\widehat{H}_i(t_i)$ and $\hat{S}_0(t_i)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} prostate CA recurrence and death at t_i. The hazard function shall be observed to follow an approximately 45-degree linear relationship in the plot of $-\log\left(\hat{S}_0(t_i)\right)$ versus time, and/or in the plot of $\log(-\log\left(\hat{S}_0(t_i)\right))$. If a straight line with a slope of 1 passing through the origin is observed in the plot of $\log(time)$, it will be concluded that the model provides a better fit to the data. An approximate likelihood cross-validation criterion (LCV) is used to determine the joint and reduced models' goodness of fit as described by (Gray, 1992).

3.4 **Ethical consideration**

The study has received ethical approval from Hawassa University's research ethics review board. After receiving written consent from the TASH oncology department, the researcher collected the data following the official written letter of cooperation from the department of statistics to the hospital requesting authorization. Due to the use of a retrospective longitudinal study design, no informed consent was obtained for the investigation. The physical copy and electronic copies of all the collected data were kept in a locked cabinet with a password-protected computer to safeguard the confidentiality of any information about the patients and their clinical history. The de-identified information was only be accessible to the researcher and kept in a safe location. De-identified and coded data were the sole subject of all analyses. There was no communication between the patients and the researcher during data collection.

Statistical software

The latest version of R statistical software was used for the analysis of this research.

4. **Result and discussion**

4.1 **Descriptive analysis**

In this study, 222 men who followed prostate cancer treatment in TASH between 1st January 2018 and 30th January 2021 were considered. The response was time to recurrence of prostate cancer and death due to prostate CA recurrence. From the total of 432 recurrent observation, about 210 (61.1%) of them experienced recurrence of prostate cancer, 192(44.5%) of the experienced death (terminated) event and 222 (38.9%) were censored. From the total recurrent observation 182(42.1%) are treated with chemotherapy of which 93(51%) are events. Of the total recurrent observation 145(33.5%) are treated with hormone of which 68(47%) have shown recurrence of prostate cancer. From the total recurrent observation 105(24.3%) are treated with both combination of two or more treatments, of which 51 have shown prostate recurrence. From the total of 432 prostate cancer recurrent observation 212(49%) are smokers, of which 157(74%) have shown prostate cancer events recurrence.

From the total prostate cancer recurrence observation 274(63.4%) are above 60 years of age of which 135 (49.3%) have shown prostate cancer recurrence event. Also from the total prostate cancer recurrence events 216(50%) have distance metastasized prostate tumor, of this 125(58%)

have shown prostate cancer recurrence. From the total of prostate cancer recurrence observation 202(46.7%) have PSA level between 10 and 20 of which 100(49.5%) have shown the recurrence of prostate cancer, similarly 131(30.3%) patients have PSA level greater than 20 of this 55(42%) have shown prostate cancer recurrence and finally 99(23%) patients have PSA level less than 10, from this 57(57.6%) have shown prostate cancer recurrence. From the total of 432 prostate cancer recurrent event observations 247(57.1%) are in clinical stage 4 of which 121(48.9) have experienced prostate cancer recurrence. And also 185(42.8%) patients are in clinical stage 3, of this 121(65.4%) have shown prostate cancer recurrence.

Out of the total recurrent observations 192 observations have terminated (dead) due to recurrence, of those 80(41.16%) are treated with chemotherapy, 66(34.3%) have been treated with hormone therapy and 46(23.9%) of them are treated with combination of two or more treatments. From the total of terminated prostate cancer patients due to recurrence 129 are smokers. From the total terminated prostate CA patients 122 are above the age of 60 and, 70 are between the ages 40-60.

Of the total terminated events 84(43.75) prostate cancer patients had distant metastasized tumor and 108(56.25) of them didn't have distant metastasized tumor. In addition, from the total terminated (dead) prostate cancer patients 68 of them had PSA level greater than 20, 89 of them had PSA level between 10 and 20 and, 35 of the, had PSA level less than 10. Finally from the total terminated observation of prostate cancer patients 105(54.6%) of them were in stage 4 and 87(45.3%) of them were in stage 3. The average number of recurrent event per subject is 0.924 and the Proportion of subjects with a terminal event 0.549. The median follow-up time 500 days.

Table 2: descriptive statistics of explanatory variables of prostate cancer recurrence and death due to prostate cancer.

Variable name	Recurrent status				Terminal event		
treatment	Category	Censored	events	total	censored	Death	total
	Chemo	89(49%)	93(51%)	182(42.1%)	9(10%)	80(41.6%)	89(46%)
	hormone	77(53%)	68(47%)	145(33.5%)	11(14%)	66(34.3%)	77(40%)
	combination	54(52%)	51(48%)	105(24.3%)	8(15%)	46(23.9)	54(28%)
smoking	Yes	55(26%)	157(74%)	212(49%)		129(67.2%)	

	No			220(50.9%)		63(32.8%)	
Age	Above 60	139(51%)	135(49%)	274(63.4%)	17(12%)	122(63.5%)	139(72%)
	40-60			158(36.57)		70(36.5)	
Distance Metastasize	Yes	91(42%)	125(58%)	216(50%)	7(8%)	84(43.7%)	91(47%)
	No			216(50%)		108(56.25%)	
PSA level	Less than 10	42(43%)	57(57.6%)	99(23%)	7(17%)	35(18.2%)	42(22%)
	10-20	102(50.5%)	100(49.5%)	202(46.7)	13(13%)	89(46.3)	102(53%)
	Greater than 20	76(58%)	55(42%)	131(30.3)	8(10%)	68(35.4)	76(40%)
Clinical stage	Stage 3	64(34.5%)	121(65.4%)	185(42.3%)	0	105(54.6%)	105(55%)
	Stage 4	126(51.1%)	121(48.9)%	247(57.1)	39(31%)	87(45.3%)	126(67%)

Plot of survival baseline functions

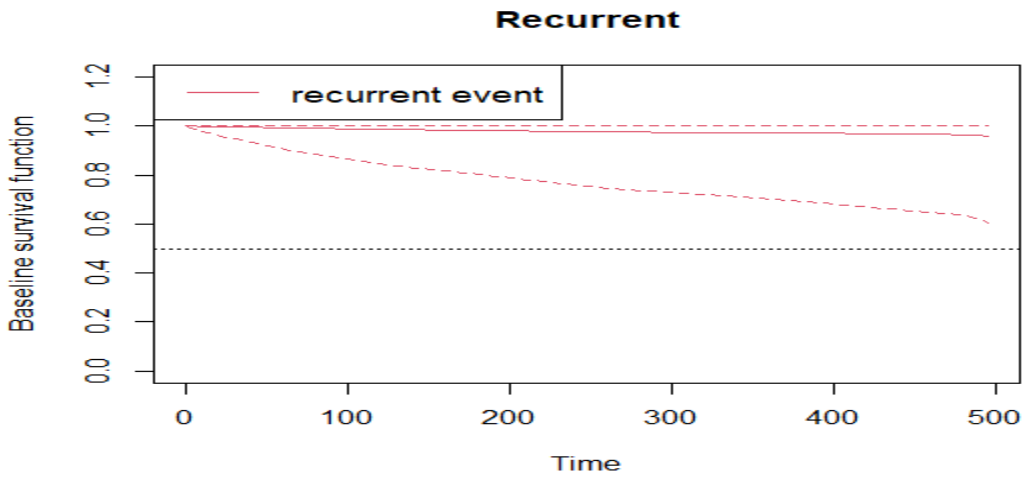

Figure 1. Survival baseline functions for recurrent event

The above figure shows how the survival of the recurrence of events changes over time. From the plot we can examine that the survival for recurrence event change over time.

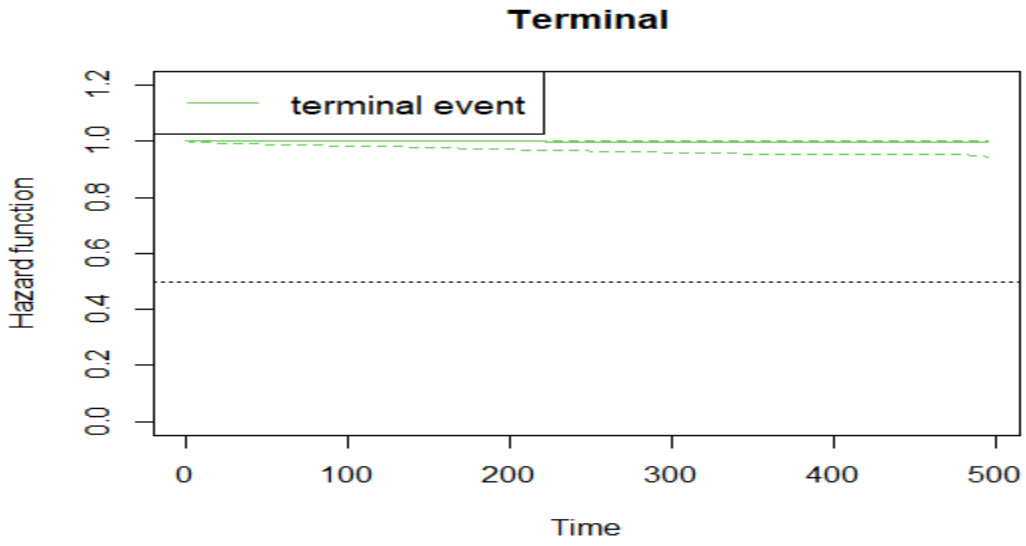

Figure 2 Survival baseline functions for death event

The above figure shows how the survival for death changes over time. From the plot we can examine that the survival for terminal event (death) did not change (constant) over time.

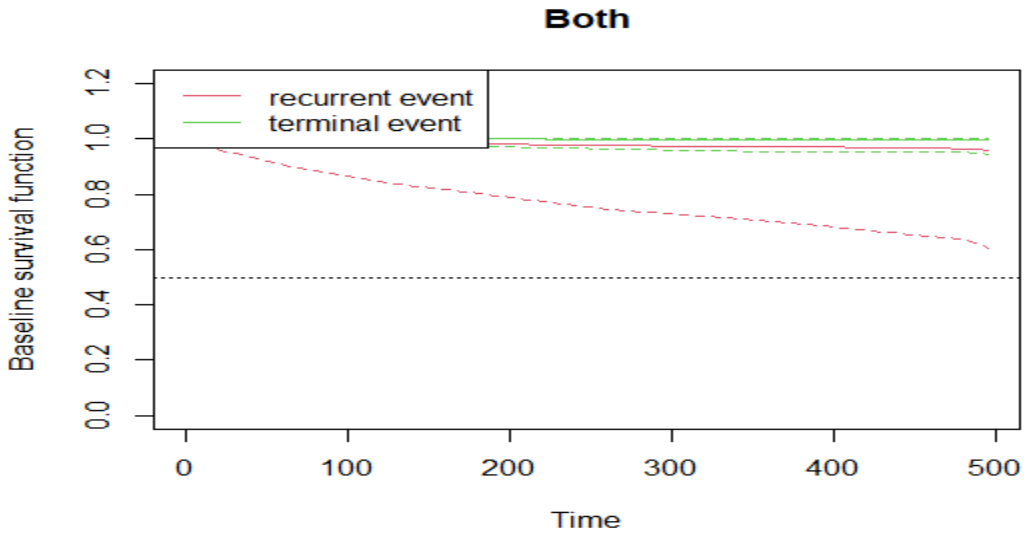

Figure 3: Survival baseline functions for death and recurrent events (joint frailty model)

The above figure representing the baseline hazard function for a specific event type (death and recurrent events). These curves show different shapes, indicating different hazard patterns over time. Terminal event (Death) has a constant baseline survival over time, while recurrent event has a changing baseline survival over time.

4.2 Cox proportional hazard model

To determine the clinical and socio demography covariates which are associated with the observed time to recurrent event of prostate cancer patients, first we fitted Cox proportional hazard model for each risk factor before proceeding to more complicated models. Variables with p-value less than or equal to 25% in the uni-variable analysis were considered for multivariable model (Hosmer et al., 2008; Bursac, et al., 2008). For multivariable analysis, variables with P-value less than or equal to 5% were selected as significant covariates.

The result from the standard Cox PH model using penalized likelihood for prostate cancer recurrence and cancer death due to recurrence are displayed in the below table is presented on table 3 below. It is observed that age, smoking, distance metastasize, PSA level, stage, and Gleason score were associated with time to recurrence of prostate cancer and terminal event (death) of patients. The Standard Cox PH model considers different line of data contributed by the same subjects as independent contributions from different subjects.

Table 3: Cox-PH uni-variable parameter estimate of PCA recurrence and death.

Variable name	Status	Category	B	Exp(β)	SE coef(H)	P-value	95% CI of coefficients
Treatment	Recurrent event	Chemo	-0.3168	0.7284	0.2335	0.1749	(0.46 - 1.15)
		Hormone	-0.2561	0.7740	0.2065	0.2149	(0.52 -1.16)
	Terminal event (death)	Chemo	5.0202	151.44	0.1822	< 0.0001	(105.95-216.49)
		Hormone	4.9429	140.18	0.1800	< 0.0001	(98.51-199.51)
	Recurrent event	Yes	0.705863	2.02559	0.22200	0.001475	(1.31 - 3.13)

Smoking	Terminal event(death)	Yes	1.61438	5.02476	0.19532	< 0.0001	(3.43 - 7.37)
Age	Recurrent Event	Above 60	0.93340	2.54316	0.20190	0.00037839	(1.71 -3.78)
	Terminal event(death)	Above 60	1.03512	2.81546	0.15918	<0.00078853	(2.06 -3.85)
Distance Metastasize	Recurrent event	Yes	0.913405	2.4928	0.20191	0.000060777	(1.68 -3.70)
	Terminal event(death)	Yes	1.04661	2.84799	0.15931	0.000050539	(2.08 - 3.89)
PSA level	Recurrent event	10-20	-0.2403	0.7863	0.2065	0.2443	(0.52 - 1.18)
		Greater than 20	-0.3397	0.7119	0.2065	0.1458	(0.45 -1.13)
	Terminal event(death)	10-20	4.9679	143.728	0.1819	< 0.0001	(100.61-205.33)
		Greater than 20	4.9918	147.210	0.1846	< 0.0001	(102.51-211.39)
Clinical stage	Recurrent event	Stage 4	2.19403	8.97133	0.27160	0.00066613	(5.27-15.28)
	Terminal event(death)	Stage 4	1.425005	4.15787	0.16142	< 0.0001	(3.03 - 5.71)

For the treatment category, although it is insignificant individuals receiving chemotherapy had a lower risk of recurrent events (hazard ratio = 0.7284, 95% CI: 0.46 to 1.15, p = 0.1749) compared to hormone therapy. However, for the terminal event of death, individuals undergoing chemotherapy faced a substantially higher risk (hazard ratio = 151.44, 95% CI: 105.95 to 216.49, p < 0.0001) compared to those on hormone therapy. Regarding smoking, individuals who smoked had a significantly higher risk of both recurrent events (hazard ratio = 2.02559, 95% CI: 1.31 to 3.13, p = 0.001475) and terminal events (hazard ratio = 5.02476, 95% CI: 3.43 to 7.37, p < 0.0001) compared to non-smokers. Age above 60 was associated with an increased risk of both recurrent events (hazard ratio = 2.54316, 95% CI: 1.71 to 3.78, p = 0.00037839) and terminal events (hazard ratio = 2.81546, 95% CI: 2.06 to 3.85, p <0.00078853).

For patients with metastasis, the presence of metastasis increased the risk of recurrent events (hazard ratio = 2.4928, 95% CI: 1.68 to 3.70, p = 0.000060777) and terminal events (hazard ratio = 2.84799, 95% CI: 2.08 to 3.89, p = 0.000050539). In terms of PSA level, individuals with PSA levels between 10-20 or greater than 20 had similar risk profiles for recurrent events. However, for terminal events, higher PSA levels were associated with significantly increased risks: hazard ratio = 143.728 (95% CI: 100.61 to 205.33, p < 0.0001) for PSA levels between 10-20 and hazard ratio = 147.210 (95% CI: 102.51 to 211.39, p < 0.0001) for PSA levels greater than 20. Finally, clinical stage 4 was strongly associated with higher risks of both recurrent events (hazard ratio = 8.97133, 95% CI: 5.27 to 15.28, p = 0.00066613) and terminal events (hazard ratio = 4.15787, 95% CI: 3.03 to 5.71, p < 0.0001) compared to other stages. In summary, the results indicate that various factors, including treatment type, smoking status, age, metastasis, PSA level, and clinical stage, have a significant impact on the risk of prostate cancer recurrence and cancer-related death. These findings provide valuable insights for understanding and managing prostate cancer in patients.

4.2.1 Survival analysis for time to death due to prostate cancer

Table 4: multi-variable Cox-PH parameter estimate of PCA death using penalized likelihood method.

Variables	coef	exp(coef)	SE coef (H)	SE coef (HIH)	Z	P
smoking	0.8397933	2.315888	0.3751064	0.3751064	2.238814	0.025168
gleasonscore	0.0192001	1.019386	0.0715773	0.0715773	0.268243	0.78851
distmetastyes	-0.5862019	0.556437	0.3635919	0.3635919	-1.612253	0.10691
stagelevelIV	1.0995177	3.002718	0.2600023	0.2600023	4.228876	0.00023486
treatchemo	-0.0465983	0.954471	0.2739299	0.2739299	-0.170110	0.86492
treathormone	0.0357514	1.036398	0.2749470	0.2749470	0.130030	0.89654

Smokers have approximately 2.32 times higher hazard of prostate cancer death compared to non-smokers. This effect is statistically significant ($p = 0.025168$). For a one-unit increase in Gleason score, there is a 1.94% increase in the hazard of prostate cancer death. However, this effect is not statistically significant ($p = 0.78851$). Distant metastasis have a 44.36% lower hazard of prostate cancer death compared to those without distant metastasis. This effect is not statistically significant but shows a trend ($p= 0.10691$). Patients with stage IV cancer have approximately 3.00 times higher hazard of prostate cancer death compared to patients with lower stages. This effect is highly significant ($p = 0.00023486$). Neither chemotherapy nor hormone therapy significantly affect the hazard of prostate cancer recurrence or death due to recurrence ($p > 0.05$).

4.2.2 Survival analysis for prostate cancer recurrence using shared Gamma frailty

In recurrent events data like prostate cancer recurrence, subjects may have more than one events of interest. Thus, patients with the same id are considered as correlated. An extension of the Cox model can be considered by taking into account the clustered structure of the data. Thus clustering can be considering as a random effect. In the shared gamma frailty model, first uni-variable analysis was conducted and significant variables at 25% level of significance were taken to the

multiple shared gamma frailty model in order for not excluding important variables (Bursac, et al., 2008). Results of uni-variable estimate is presented in table 5 of the appendix. The result from the shared Gamma frailty model is presented on the table below. It is observed that treatment, clinical stage, smoking status, age category and distance metastasis were the only significant covariates selected from the multiple shared Gamma frailty model.

Table 6: Parameter estimates of Shared Gamma Frailty model using a Penalized Likelihood

Variables	**Coef**	**exp(coef)**	**SE coef (H)**	**SE coef (HIH)**	**Z**	**P**
Smoking	0.8936875	2.444126	0.257360	0.257360	3.472523	0.00051559
gleasonscore	0.0535901	1.055052	0.050926	0.050926	1.052312	0.29266
distmetastyes	-0.4105161	0.663308	0.242895	0.242895	-1.690095	0.091010
stagelevelIV	1.8839669	6.579554	0.235932	0.235932	7.985219	0.0000014433
treatchemo	-0.1708429	0.842954	0.173636	0.173636	-0.983912	0.032516
treathormone	-0.2246434	0.798801	0.173857	0.173857	-1.292119	0.019632
serumPSA in b/n 10-20	-0.1461095	0.864063	0.2408350	0.2408350	-0.606679	0.54406000
serumPSA than 20	-0.1534174	0.857772	0.2529721	0.2529721	-0.606460	0.54421000
agecat40-60	-2.2071964	0.110009	0.9666588	0.9666588	-2.283325	0.02241100
Frailty parameter						
Theta	(SE (H)	P				
0.000015363	0.00000613069	0.5				

Smokers have approximately 2.44 times higher hazard of prostate cancer recurrence compared to non-smokers. This effect is statistically significant (p = 0.00051559). For a one-unit increase in Gleason score, there is a 5.5% increase in the hazard of prostate cancer recurrence. However, this effect is not statistically significant (p = 0.29266).Patients with distant metastasis have a 33.7% lower hazard of prostate cancer recurrence compared to those without distant metastasis. This effect is not statistically significant but shows a trend (p = 0.091010). Patients with stage IV cancer have approximately 6.58 times higher hazard of prostate cancer recurrence compared to patients with lower stages. This effect is highly significant (p = 0.0000014433). Chemotherapy and hormone therapy have a 16% and 21% lower hazard of prostate cancer recurrence compared to combined treatment with as significant effect(p=0.032 and p=0.019).Age category(40-60) have 89% lower hazard of prostate cancer recurrence compared to age greater than 60. The frailty parameter represents the variability in individual risks within the population. In this model, frailties contribute negligibly to the hazard of prostate cancer recurrence (p = 0.5).

4.2.3 Survival analysis for prostate cancer recurrence using shared Log-normal frailty

Similarly, we conducted uni-variable and multivariable analysis for the shared log-normal frailty model. The result of uni-variable analysis indicate treatment taken, age category, Gleason score, clinical stage, smoking habit and PSA level were statistically significant at 25% level of significance as presented in the appendix (Table 7). The result from the shared log-normal frailty model is presented on table 4.5 below. It is observed that age category, clinical stage, smoking status and distance metastasis were the only significant covariates selected from the saturated multiple shared log-normal frailty model.

Table 8: multivariable Parameter estimates of Shared Log-Normal Frailty model using a Penalized Likelihood.

Variables	Coef	exp(coef)	SE coef (H)	SE coef (HIH)	Z	P
smoking	1.0904331	2.975563	0.3427935	0.3427935	3.181021	0.0014676
gleasonscore	0.0903252	1.094530	0.0594713	0.0594713	1.518802	0.12881

distmetastyes	-0.6937967	0.499675	0.342097	0.342097	-2.028072	0.0425530
stagelevelIV	2.1241997	8.366199	0.2853807	0.2853807	7.443391	< 0.0001
treatchemo	-0.2274180	0.796588	0.2739299	0.2739299	-1.06440	0.28714
treathormone	-0.1988653	0.819660		0.2749470	-0.86557	0.38672
serumPSAlevel 10-20	-0.2156501	0.806017	0.2421129	0.2421129	-0.89070	0.37309
serumPSAlevel than 10-20	-0.2393736	0.787121	0.2572651	0.2572651	-0.93045	0.35214
agecat40-60	-2.3011602	0.100143	0.972365	0.972365	-2.36656	0.01795400

Frailty parameter		
Theta	(SE (H)	P
0.651628	0.201354	0.0006056

Smokers have approximately 2.97 times higher hazard of prostate cancer recurrence compared to non-smokers. This effect is statistically significant (p =0.0014676).For a one-unit increase in Gleason score, there is a 9.4% increase in the hazard of prostate cancer recurrence. However, this effect is not statistically significant (p = 0.12881).Patients with distant metastasis have a 50.032% lower hazard of prostate cancer recurrence compared to those without distant metastasis. This effect is statistically significant (p = 0.0425530). Patients with stage IV cancer have approximately 8.4 times higher hazard of prostate cancer recurrence compared to patients with lower stages, this effect is highly significant (p < 0.0001). Chemotherapy and hormone therapy do not have a significant impact on the hazard of prostate cancer recurrence (p > 0.05). The frailty parameter represents the variability in individual risks within the population. In this model, frailties contribute non-negligibly to the hazard of prostate cancer recurrence (p = 0.0006056).

4.3 Joint model for PCA recurrence and terminal event (death)

4.3.1 Joint gamma frailty model with frailty

For the joint Gamma frailty model of PCA recurrence and death, we have conducted uni-variable and multivariable analysis for the shared gamma frailty model. The result of uni-variable analysis indicate treatment taken, age category, Gleason score, clinical stage, smoking habit and PSA level were statistically significant at 25% level of significance as presented in the appendix (Table 9).Multivariable parameter estimate of the joint PCA recurrence and death using penalized likelihood estimation method are showed in the below table.

Table 10: Parameter estimates of joint Shared Gamma Frailty model using a Penalized Likelihood

Recurrences						
Variables	coef	exp(coef)	SE coef (H)	SE coef (HIH)	Z	p
smoking	1.294586	3.649485	0.3344539	0.3281765	3.87075	0.0010850
gleasonscore	0.165163	1.179585	0.0735218	0.0460968	2.24644	0.024676
distmetastyes	-0.695132	0.499009	0.3222219	0.3214946	-2.15731	0.030982
stagelevel.IV	2.203925	9.060509	0.2946161	0.2615119	7.48067	< 0.00007
treatchemo	-0.252304	0.777008	0.1973160	0.1947648	-1.27868	0.20101
treathormone	-0.221585	0.801247	0.1988154	0.1958212	-1.11453	0.26505
Terminal event						
smoking	1.774741	5.898752	0.5576986	0.5349894	3.182258	0.0014613
gleasonscore	0.414106	1.513017	0.0655705	0.0452316	6.315434	< 0.0002
distmetastyes	-1.106649	0.330665	0.5430165	0.5324937	-2.037967	0.041553
stagelevel.IV	2.135104	8.457927	0.3912891	0.3650436	5.456590	0.000485
treatchemo	-0.131032	0.877189	0.3607889	0.3600764	-0.363183	0.71647
treathormone	0.189902	1.209131	0.3524943	0.3522019	0.538738	0.59007
Frailty parameter	Coeff	SE (H)	SE (HIH)	p-value		
Theta	0.705969	0.144464	0.142975	5.1242e-07		
Alpha	1.13195	0.386713	0.348615	0.0034213		

The above summary outlines the results of a Joint Frailty model applied to the hazard function for prostate cancer recurrence and cancer-related death due to recurrence. This model considers several predictor variables and frailty parameters to analyze their impact on the recurrence and terminal event processes. For both recurrence and terminal events, smoking significantly increases the hazard, with a hazard ratio of 3.65 and 5.90 respectively with a significant effect (p=0. 0010850 and p=0.0014613), indicating that smokers have a substantially higher risk of recurrence and cancer-related death compared to non-smokers. Higher Gleason scores are associated with an increased hazard for both recurrence (hazard ratio of 1.18) and terminal events (hazard ratio of 1.51) with significant p-value (0.024676, < 0.0002). This suggests that patients with more aggressive tumor grades have a higher risk of both recurrence and cancer-related death.

The presence of distant metastases decreases the hazard for recurrence (hazard ratio of 0.50), indicating that patients with distant metastases have a lower risk of recurrence. However, it increases the hazard for terminal events (hazard ratio of 0.33), signifying that these patients have a higher risk of cancer-related death after recurrence occurs. Patients diagnosed with Stage IV cancer have significantly higher hazards for both recurrence (hazard ratio of 9.06) and terminal events (hazard ratio of 8.46). This highlights the critical impact of advanced cancer stage on increasing the risk of both recurrence and cancer-related death. Chemotherapy treatment does not significantly influence the hazards for both recurrence (hazard ratio of 0.78) and terminal events (hazard ratio of 0.88) on this particular study. This suggests that chemotherapy does not substantially alter the risk of recurrence or cancer-related death in this context.

Hormone therapy also does not have a significant impact on hazards for both recurrence (hazard ratio of 0.80) and terminal events (hazard ratio of 1.21). This implies that hormone therapy does not markedly affect the risk of recurrence or cancer-related death in this study context. Additionally, the frailty parameters provide insights into the variability of individual risks within the population. The variance of frailties (theta) is 0.705969 with a significant value (p=5.1242e-07), suggesting that there is considerable heterogeneity in cancer recurrence risks among individuals. The alpha parameter, which represents the impact of frailties on the terminal event, is 1.13195 with significant value (p=0.0034213). This indicates that frailties also play a significant role in determining the risk of cancer death due to recurrence i.e. death and recurrence are correlated for PCA patients.

4.4 Proportional assumptions

i. For time to recurrent model

Table 11: proportional assumption test using statistical tests for PCA recurrent events.

Variables	Chisq	df	P
smoking	0.0383	1	0.845
distmetast	0.8921	1	0.345
Stagelevel	2.8218	1	0.093
Treat	0.5548	3	0.758
GLOBAL	4.7892	6	0.571

The above table provides results for the proportional hazard assumption tests in a survival model for time to recurrent events. The variable smoking does not violate the proportional hazard assumption (p-value = 0.845), suggesting that the hazard of recurrent events for smokers versus non-smokers remains constant over time. The variable Distance of Metastasis does not violate the proportional hazard assumption (p-value = 0.345), suggesting that the hazard of recurrent events does not vary significantly based on the distance of metastasis.

The variable stage level shows a borderline result (p-value = 0.093), indicating that the hazard of recurrent events is proportional across different stage level. The variable Treatment does not violate the proportional hazard assumption (p-value = 0.758), suggesting that the hazard of recurrent events remains proportional across different treatment categories. The global test for the proportional hazard assumption, considering all variables, is not statistically significant (p-value = 0.571). This suggests that, overall, the proportional hazard assumption holds for the model of time to recurrent events when considering all the variables together.

ii. For time to death model

Table 12: proportional assumption test using statistical tests for PCA death events.

Variables	Chisq	df	P
smoking	0.214	1	0.643
distmetast	0.634	1	0.426
stagelevel	0.755	1	0.385
Treat	4.774	3	0.092
GLOBAL	7.907	6	0.245

The above table presents the results of the proportional hazard assumption tests for a survival mo del related to time to death. The variable smoking satisfies the proportional hazard assumption (p -value = 0.643), indicating that the hazard of death remains constant over time for smokers and n on-smokers. The variable Distance of Metastasis satisfies the proportional hazard assumption (p-value = 0.426), indicating that the hazard of death does not significantly vary based on the distan ce of metastasis.

The variable stage level meets the proportional hazard assumption (p-value = 0.385), suggesting t hat the hazard of death remains proportional across different stages. The variable Treatment show s a borderline result (p-value = 0.092), indicating that the hazard of death remains constant over t ime for Treatment. The global test for the proportional hazard assumption, considering all variabl es, is not statistically significant (p-value = 0.245). This suggests that, overall, the proportional ha zard assumption holds for the model of time to death when considering all the variables together.

4.5 Model comparison

Table 13: model comparison

Models	penalized marginal log-likelihood	LCV
Shared gamma frailty model	-1380.17	3.23651
Shared Log-Normal frailty model	-1377.15	3.2295

Shared gamma joint frailty model	-0.05	0.0818404

The Shared gamma frailty model has a penalized marginal log-likelihood of -1380.17, indicating a good fit to the data. The LCV value of 3.23651 suggests that the model has moderate predictive accuracy for recurrence events. The Shared Log-Normal frailty model has a penalized marginal log-likelihood of -1377.15, indicating a good fit to the data for death events. The LCV value of 3.2295 Suggests that the model has reasonable predictive accuracy for recurrence events. The Shared gamma joint frailty model has a very low penalized marginal log-likelihood of -0.05, indicating an excellent fit to the data. The extremely low value suggests that this model fits the data almost perfectly. The LCV value of 0.0818404 confirms the exceptional predictive accuracy of this joint model. In summary, the joint model performs exceptionally well, providing an almost perfect fit to the data and outstanding predictive accuracy. The recurrence and death models also perform well individually, with good fits and reasonable predictive accuracy, especially for Shared Log-Normal frailty model events.

4.6. Discussion

The current study aimed to determine the factors that affect prostate cancer recurrence and terminal events (death) jointly in the context of Tikur Anbessa specialized hospital. According to the current study, smokers were substantially more likely than non-smokers to have terminal events, or death (hazard ratio = 5.02476). This is not comparable to the research done in the same field by Jemal B et al., 2020, which found that the multivariable Cox regression did not statistically significantly indicate smoking status.

In this study for patients with distant metastasis, the presence of metastasis increased the risk of terminal events (hazard ratio = 2.84799). Which is not similar to the study conducted in the same area by Jemal B et.al., 2020 which revealed presence/absence of distant metastasis were not statistically significant in the multivariable Cox regression. This is not comparable to the research done in the same field by Jemal B et al., 2020, which found that the multivariable Cox regression did not statistically significantly indicate presence/absence of distant metastasis. Higher PSA levels have been related to significantly higher risks for terminal events: hazard ratio = 143.728 for PSA levels between 10 and 20 and hazard ratio = 147.210 for PSA levels above 20. Supported

by Jemal B et al.'s 2020 findings, which showed a substantial variation in survival rates depending on PSA levels upon diagnosis. Individuals who receive an early diagnosis generally have a far higher chance of surviving than those who receive an advanced diagnosis. According to the current study, higher Gleason score were substantially more likely have PCA recurrence (hazard ratio = 1.179585). This is comparable to the research done in the same field by (GROSSFELD et al., 2003) , Patients with a serum PSA at diagnosis of greater than 20 ng./ml. and a Gleason score of less than 8 had a significantly higher likelihood of remaining disease-free 5 years after surgery than similar patients with a Gleason score of 8 or greater (45% versus 0%, $p < 0.05$).

Compared to earlier stages, clinical stage 4 in the current investigation was substantially linked to increased chances of terminal events (hazard ratio = 4.15787). Numerous studies have also shown that survival rates decreased progressively with the advanced stage of the disease upon diagnosis (Beksisa et al., 2020) earlier stage of cancer at diagnosis and ADH treatment were significantly associated with better survival. Accordingly, the likelihood of death increases by 3 fold as the cancer stage is one step higher. But also there is study that contradict with the result of this context indicating Clinical tumor stage does not appear to be an important predictor in its case(GROSSFELD et al., 2003). According to a supported study carried out in the same field, patients with early diagnoses had a far higher overall survival rate than those with advanced diagnoses. In a comparable manner, Ghanaian research (Kkwesi, 2015) shows that patients with early stages of the illnesses at diagnosis (stage I & II) have a better prognosis than those with later stages (III & IV).

5. Conclusion

In a comprehensive analysis of 222 men undergoing prostate cancer treatment at TASH, spanning from January 2018 to January 2021, it was observed that treatment type, smoking habits, age, metastasis presence, PSA levels, and clinical staging significantly influenced the recurrence of prostate cancer and subsequent cancer-related mortality. Patients receiving chemotherapy had a lower risk of recurrent events but faced a substantially higher risk of death compared to those on hormone therapy. The choice of treatment should be carefully tailored to individual patient profiles. Smoking was associated with a significantly increased risk of both prostate cancer recurrence and cancer-related death. Smoking cessation interventions are crucial for improving

outcomes. Patients above the age of 60 had an elevated risk of both recurrent events and terminal events, emphasizing the importance of age-appropriate care and monitoring.

The presence of distant metastasis significantly increased the risk of both recurrent events and cancer-related death, underscoring the need for vigilant management of metastatic cases. While patients with PSA levels between10-20 or greater than 20 had similar risks of recurrent events, higher PSA levels were strongly associated with increased risks of cancer-related death, emphasizing the significance of monitoring and early intervention for patients with elevated PSA levels. Patients in clinical stage 4 exhibited markedly higher risks of both recurrent events and terminal events, highlighting the importance of timely diagnosis and management, especially for advanced-stage cases. The frailty parameters revealed substantial variability in individual risks, emphasizing the importance of considering individual differences in predictive models.

These findings provide valuable insights for healthcare practitioners in developing more effective and individualized strategies for the understanding and management of prostate cancer in patients at Tikur Anbessa Specialized Hospital.

6. Recommendation

Based on the findings from the study "Joint modelling of recurrent events in prostate cancer and time to death of patients in Tikur Anbessa specialized hospital," the following recommendations can be made:

- Clinicians should carefully consider the choice of treatment for prostate cancer patients, taking into account the risks and benefits associated with chemotherapy and hormone therapy. A personalized treatment approach may be beneficial to optimize outcomes.
- Given the significantly higher risk of both recurrent events and terminal events among smokers, healthcare providers should emphasize the importance of smoking cessation programs for prostate cancer patients, as quitting smoking may improve their overall prognosis.
- Healthcare providers should be vigilant in monitoring and providing tailored care to prostate cancer patients above the age of 60, as they are at an increased risk of recurrence and death. Early detection and management strategies specific to this age group should be implemented.

- Timely detection and management of metastasis in prostate cancer patients are crucial. Patients with metastatic disease have a substantially higher risk of both recurrent and terminal events, necessitating focused treatment and surveillance.
- Regular monitoring of PSA levels is essential, particularly for patients with PSA levels between10-20 and greater than 20, as these levels are associated with higher risks of terminal events. Close monitoring and prompt intervention may help improve patient outcomes.
- Patients in clinical stage 4 have a significantly higher risk of both recurrence and death. Healthcare providers should consider more intensive and targeted treatment strategies for these patients.
- Patients should be educated about the risks associated with their specific clinical profile, including treatment type, smoking history, age, metastasis status, PSA levels, and clinical stage. Informed patients may be more proactive in managing their health and adhering to treatment plans.

These recommendations aim to improve the management and care of prostate cancer patients, ultimately enhancing their quality of life and prognosis. It is essential for healthcare providers to consider these factors when developing treatment strategies and providing support to patients with prostate cancer.

6.1 Limitation

Some major limitations of this study are:

- It is conducted on a secondary data which is keen to bias and incomplete information.
- As the data was covered from patient's card, important risk factors or variable are not obtained as desired.
- Even though the study is conducted retrospectively based on hospital, financial and physical burdens for patients to follow up on their treatment contributes to incomplete data.
- Computational complexities attribute for managing the data during analysis since the study have used newly introduced package for modeling the data.

7. **References**

1. Aalen, O. (1978). Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models. *Https://Doi.Org/10.1214/Aos/1176344198*, *6*(3), 534–545. https://doi.org/10.1214/AOS/1176344198
2. American Cancer Society. (2022a). *Prostate Cancer | Prostate Cancer Information and Overview*. https://www.cancer.org/cancer/prostate-cancer.html
3. American Cancer Society. (2022b). *Prostate Cancer | Prostate Cancer Information and Overview*. https://www.cancer.org/cancer/prostate-cancer.html
4. Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. https://doi.org/10.1007/978-1-4612-4348-9
5. Anderson, K. M., Wilson, P. W. F., Odell, P. M., & Kannel, W. B. (1991). An updated coronary risk profile. A statement for health professionals. *Circulation*, *83*(1), 356–362. https://doi.org/10.1161/01.CIR.83.1.356
6. Beksisa, J., Getinet, T., Tanie, S., Diribi, J., & Hassen, Y. (2020). Survival and prognostic determinants of prostate cancer patients in Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: A retrospective cohort study. *PLoS ONE*, *15*(3). https://doi.org/10.1371/journal.pone.0229854
7. Bray, F., Parkin, D. M., Gnangnon, F., Tshisimogo, G., Peko, J. F., Adoubi, I., Assefa, M., Bojang, L., Awuah, B., Koulibaly, M., Buziba, N., Korir, A., Dzamalala, C., Kamate, B., Manraj, S., Ferro, J., Lorenzoni, C., Hansen, R., Nouhou, H., … Chingonzoh, T. (2022). Cancer in sub-Saharan Africa in 2020: a review of current estimates of the national burden, data gaps, and future needs. *The Lancet. Oncology*, *23*(6), 719–728. https://doi.org/10.1016/S1470-2045(22)00270-4
8. Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*(1), 141–151. https://doi.org/10.1093/BIOMET/65.1.141
9. Cochran, W. (1977). Subsampling with units of unequal sizes. *Sampling Techniques*, 303. https://archive.org/details/Cochran1977SamplingTechniques_201703
10. Cook, R., & Lawless, J. (2007). The Statistical Analysis of Recurrent Events. *The Statistical Analysis of Recurrent Events*. https://doi.org/10.1007/978-0-387-69810-6
11. Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/J.2517-6161.1972.TB00899.X
12. D'Amico, A. V., Whittington, R., Bruce Malkowicz, S., Schultz, D., Blank, K., Broderick, G. A., Tomaszewski, J. E., Renshaw, A. A., Kaplan, I., Beard, C. J., & Wein, A. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*, *280*(11), 969–974. https://doi.org/10.1001/JAMA.280.11.969
13. David G. Kleinbaum. (2011). *Survival Analysis: A Self-Learning Text*. https://www.goodreads.com/book/show/104704.Survival_Analysis

14. Duchateau, L., Janssen, P., Kezic, I., & Fortpied, C. (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society Series C*, *52*(3), 355–363. https://doi.org/10.1111/1467-9876.00409
15. Epstein, J. I., Walsh, P. C., Carmichael, M., & Brendler, C. B. (1994). Pathologic and Clinical Findings to Predict Tumor Extent of Nonpalpable (Stage T1 c) Prostate Cancer. *JAMA*, *271*(5), 368–374. https://doi.org/10.1001/JAMA.1994.03510290050036
16. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019a). 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
17. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019b). 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
18. Finbarr O, & SULLIVANt. (1988). *FAST COMPUTATION OF FULLY AUTOMATED LOG-DENSITY AND LOG-HAZARD ESTIMATORS** (Vol. 9, Issue 2). http://www.siam.org/journals/ojsa.php
19. Fitzmaurice, C., Akinyemiju, T. F., Al Lami, F. H., Alam, T., Alizadeh-Navaei, R., Allen, C., Alsharif, U., Alvis-Guzman, N., Amini, E., Anderson, B. O., Aremu, O., Artaman, A., Asgedom, S. W., Assadi, R., Atey, T. M. H., Avila-Burgos, L., Awasthi, A., Saleem, H. O., Barac, A., … Naghavi, M. (2018). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016 a systematic analysis for the global burden of disease study global burden of disease cancer collaboration. *JAMA Oncology*, *4*(11), 1553–1568. https://doi.org/10.1001/JAMAONCOL.2018.2706
20. Fletcher, R. (2000). Practical Methods of Optimization. In *Practical Methods of Optimization*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118723203
21. Freedland, S. J., Aronson, W. J., Csathy, G. S., Kane, C. J., Amling, C. L., Presti, J. C., Dorey, F., & Terris, M. K. (2003). Comparison of percentage of total prostate needle biopsy tissue with cancer to percentage of cores with cancer for predicting PSA recurrence after radical prostatectomy: Results from the SEARCH database. *Urology*, *61*(4), 742–747. https://doi.org/10.1016/S0090-4295(02)02525-6
22. Freedland, S. J., Wen, J., Wuerstle, M., Shah, A., Lai, D., Moalej, B., Atala, C., & Aronson, W. J. (2008). Obesity is a significant risk factor for prostate cancer at the time of biopsy. *Urology*, *72*(5), 1102–1105. https://doi.org/10.1016/J.UROLOGY.2008.05.044
23. Gebru, T., Mekonen, H., & Getahun, N. G. (2021). *Awareness of prostate cancer and its associated factors among Ethiopian men, 2021: A cross-sectional study*. https://doi.org/10.21203/rs.3.rs-670064/v1
24. Global Cancer Observatory. (2020). *Prostate Source: Globocan 2020 Number of new cases in 2020, both sexes, all ages*. https://gco.iarc.fr/today
25. Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, *87*(420), 942–951. https://doi.org/10.1080/01621459.1992.10476248

26. Huang, C. Y., & Wang, M. C. (2004). Joint Modeling and Estimation for Recurrent Event Processes and Failure Time Data. *Journal of the American Statistical Association*, *99*(468), 1153–1165. https://doi.org/10.1198/016214504000001033
27. Joly, P., Commenges, D., & Letenneur, L. (1998). *A Penalized Likelihood Approach for Arbitrarily Censored and Truncated Data: Application to Age-Specific Incidence of Dementia* (Vol. 54, Issue 1). http://www.jstor.orgURL:http://www.jstor.org/stable/2534006http://www.jstor.org/stable/2534006?seq=1&cid=pdf-reference#references_tab_contents
28. Kalbf'leisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data Second Edition WILEY-INTERSCI ENCE*.
29. Kleinbaum, D. G., & Klein, M. (2012). *Statistics for Biology and Health Survival Analysis A Self-Learning Text Third Edition*. http://www.springer.com/series/2848
30. Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data*. https://doi.org/10.1002/9781118033005
31. Liu, L., Wolfe, R. A., & Huang, X. (2004). Shared Frailty Models for Recurrent Events and a Terminal Event. In *Biometrics* (Vol. 60).
32. Marquardt, D. W. (1963). AN ALGORITHM FOR LEAST-SQUARES ESTIMATION OF NONLINEAR PARAMETERS*. In *J. Soc. INDUST. APPL. MATH* (Vol. 11, Issue 2).
33. Michalski, J. M., Moughan, J., Purdy, J., Bosch, W., Bruner, D. W., Bahary, J. P., Lau, H., Duclos, M., Parliament, M., Morton, G., Hamstra, D., Seider, M., Lock, M. I., Patel, M., Gay, H., Vigneault, E., Winter, K., & Sandler, H. (2018). Effect of Standard vs Dose-Escalated Radiation Therapy for Patients With Intermediate-Risk Prostate Cancer: The NRG Oncology RTOG 0126 Randomized Clinical Trial. *JAMA Oncology*, *4*(6). https://doi.org/10.1001/JAMAONCOL.2018.0039
34. National Cancer Institute. (2023). *Prostate Cancer—Patient Version - NCI*. https://www.cancer.gov/types/prostate
35. Pickles, A. (1995). A COMPARISON O F FRAILTY MODELS FOR MULTIVARIATE SURVIVAL DATA. In *STATISTICS IN MEDICINE* (Vol. 14).
36. Pickles, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., & Silberg, J. (1994). Survival models for developmental genetic data: Age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology*, *11*(2), 155–170. https://www.academia.edu/12982218/Survival_models_for_developmental_genetic_data_Age_of_onset_of_puberty_and_antisocial_behavior_in_twins
37. Pound, C. R., Partin, A. W., Eisenberger, M. A., Chan, D. W., Pearson, J. D., & Walsh, P. C. (1999). Natural history of progression after PSA elevation following radical prostatectomy. *Journal of the American Medical Association*, *281*(17), 1591–1597. https://doi.org/10.1001/JAMA.281.17.1591
38. Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, *68*(2), 373–379. https://doi.org/10.1093/BIOMET/68.2.373
39. Price, D. L., & Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*, *20*(9–10), 1515–1527. https://doi.org/10.1002/SIM.687

40. Rabe-Hesketh, S., & Skrondal, A. (2005). Multilevel and Longitudinal Modeling Using Stata. *Biometrics*, *62*(3), 951–951. https://doi.org/10.1111/J.1541-0420.2006.00588_16.X
41. Rawla, P. (2019). Epidemiology of Prostate Cancer. *World Journal of Oncology*, *10*(2), 63–89. https://doi.org/10.14740/WJON1191
42. Richards, K. A., Liou, J. ing, Cryns, V. L., Downs, T. M., Abel, E. J., & Jarrard, D. F. (2018). Metformin Use is Associated with Improved Survival for Patients with Advanced Prostate Cancer on Androgen Deprivation Therapy. *The Journal of Urology*, *200*(6), 1256–1263. https://doi.org/10.1016/J.JURO.2018.06.031
43. Rondeau, V. (2010a). Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling*, *52*(7–8), 949–955. https://doi.org/10.1016/J.MCM.2010.02.002
44. Rondeau, V. (2010b). Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling*, *52*(7–8), 949–955. https://doi.org/10.1016/j.mcm.2010.02.002
45. Rondeau, V., Filleul, L., & Joly, P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, *25*(23), 4036–4052. https://doi.org/10.1002/sim.2510
46. Rondeau, V., & Joly, P. (2003). *Maximum Penalized Likelihood Estimation in a Gamma-Frailty Model*.
47. Shanko, A., Abute, L., & Tamirat, T. (2022). Prostate Cancer Screening Practice and Associated Factors Among Men in Public Health Facilities of Hossana Town, Ethiopia. *Patient Related Outcome Measures*, *Volume 13*, 229–238. https://doi.org/10.2147/prom.s380898
48. Solomon, S., & Mulugeta, W. (2019). Diagnosis and Risk Factors of Advantage Cancers in Ethiopia. *Journal of Cancer Prevention*, *24*(3), 163–172. https://doi.org/10.15430/JCP.2019.24.3.163
49. Stephenson, A. J., Scardino, P. T., Eastham, J. A., Bianco, F. J., Dotan, Z. A., Fearn, P. A., & Kattan, M. W. (2006). Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *Journal of the National Cancer Institute*, *98*(10), 715–717. https://doi.org/10.1093/JNCI/DJJ190
50. 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. https://doi.org/10.3322/caac.21660
51. 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
52. Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. https://doi.org/10.1007/978-1-4757-3294-8
53. Ullah, S., Gabbett, T. J., & Finch, C. F. (2014). Statistical modelling for recurrent events: an application to sports injuries. *British Journal of Sports Medicine*, *48*(17), 1287–1293. https://doi.org/10.1136/BJSPORTS-2011-090803

54. Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454. https://doi.org/10.2307/2061224
55. Wang, Y. C., Meyerson, L., Tang, Y. Q., & Qian, N. (2009). Statistical methods for the analysis of relapse data in MS clinical trials. *Journal of the Neurological Sciences*, *285*(1–2), 206–211. https://doi.org/10.1016/J.JNS.2009.07.017
56. Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, *84*(408), 1065–1073. https://doi.org/10.1080/01621459.1989.10478873
57. WHO. (2021). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer
58. WHO Africa. (2020). *Cancer*. https://www.who.int/health-topics/cancer#tab=tab_1
59. Yashin, A. I., Vaupel, J. W., & Iachine, I. A. (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, *5*(2), 145–159. https://doi.org/10.1080/08898489509525394
60. Yuan, M., Zhang, X., & Diao, G. (2017). Cure Frailty Models. In *Wiley StatsRef: Statistics Reference Online* (pp. 1–7). Wiley. https://doi.org/10.1002/9781118445112.stat07953

Appendices

Appendix one

Table 5 : Shared gamma uni-variable estimate using penalized likelihood method

Variable	coef	exp(coef)	SE coef(H)	SE coef (HIH)	z	P
smoking	0.781158	2.184	0.218782	0.218782	3.57048	0.0003563 2
distmetastyes	0.438882	1.55097	0.213482	0.213482	2.05583	0.039799
gleasonscore	0.0888565	1.09292	0.061120	0.061120	1.4538	0.146
serumPSAlevelbetween 10-20	-0.229912	0.794604	0.218850	0.218850	-1.0505 5	0.29347
serumPSAlevelgreater than 20	-0.365317	0.693977	0.246774	0.246774	-1.4803 7	0.13877
stagelevelIV	2.01887	7.52978	0.258059	0.258059	7.82327	< 0.005
treatchemo	-0.312768	0.73142	0.200531	0.200531	-1.5597 0	0.11883

treathormone	-0.219027	0.80330	0.204460	0.204460	-1.0712 4	0.28406
agecat40-60	-2.08083	0.124827	0.262817	0.262817	-7.9174 2	<0.0024

Table 7: Shared Log-Normal uni-variable estimate using penalized likelihood method

Variable	coef	exp(coef)	SE coef(H)	SE coef (HIH)	z	p
smoking	0.751875	2.12097	0.219916	0.219916	3.41891	0.0006287 2
distmetastyes	0.43166	1.53981	0.214993	0.214993	2.00778	0.044667
gleasonscore						
serumPSAlevelbetwee n 10-20	-0.256334	0.773884	0.213923	0.213923	-1.19825	0.23082
serumPSAlevelgreater than 20	-0.397036	0.672310	0.243957	0.243957	-1.62748	0.10363
stagelevelIV	2.11525	8.29164	0.279355	0.279355	7.57189	< 0.003
treatchemo	-0.322600	0.724264	0.197767	0.197767	-1.63121	0.10285
treathormone	-0.218387	0.803815	0.200984	0.200984	-1.08658	0.27722
agecat40-60	-2.18841	0.112094	0.286972	0.286972	-7.62587	< 0.0024

Table 9: Uni-variable Parameter estimates of joint Shared Gamma Frailty model using a Penalized Likelihood method

Variable	category		coef	exp(coef)	SE coef(H)	SE coef (HIH)	z	p
smoking	recurrent	smokingy es	1.24726	3.48078	0.229628	0.259529	5.43163	< 0.005
	death	smokingy es	2.77833	16.0921	0.37857	0.371445	7.33901	< 0.002
distmetas tasis	recurrent	distmetas yes	0.891991	2.43998	0.21122	0.229601	4.22305	< 0.0024

	death	distmetas yes	2.37569	10.7585	0.306198	0.28664	7.75867	< 0.0085
gleasonsc ore	recurrent							
	death							
PSA level	Recurrent event	serumPS Alevelbet ween 10-20	0.1882931	1.20719	0.217983	0.218850	0.863798	0.38770
		serumPS Alevelgre ater than 20	0.0987822	1.10383	0.240720	0.254463	0.410362	0.68154
PSA level	Death	serumPS Alevelbet ween 10-20	2.57853	13.1777	0.327270	0.342186	7.87889	< 0.0003
		serumPS Alevelgre ater than 20	2.34761	10.4605	0.341392	0.377699	6.87658	< 0.0061
Stage level	Recurrent	Stage IV	2.54796	12.7810	0.2524	0.3456	10.092	< 0.0001
	Death	Stage IV	3.0379	20.8614	0.2914	0.2491	10.424	< 0.0001
Treatmen t	Recurrent	treatchem o	-0.047926 3	0.953204	0.197256	0.209585	-0.24296 5	0.80803
		treathorm one	0.0369711	1.037663	0.200984	0.204157	0.183951	0.85405
	Death	treatchem o	0.848668	2.33653	0.29485	0.326156	2.87830	0.0039982
		treathorm one	1.198340	3.31461	0.30132	0.303002	3.97697	0.0000698
Age category	Recurrent	agecat40-60	-1.96514	0.140137	0.266489	0.365275	-7.37418	< 0.00016
	Death	agecat40-60	-1.08441	0.338102	0.260115	0.255838	-4.16895	< 0.00306