SEMI-MARKOVIAN ANALYSIS OF THE PROGNOSIS OF BREAST CANCER DURING MANAGEMENT IN KENYA

PETER KAMAU WAWERU

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN STATISTICS OF THE UNIVERSITY OF EMBU

AUGUST, 2025

DECLARATION

This research project is my original work and has not been presented for the award of a degree in any other university.

Signature: ..... Date: 22/08/25

Peter Kamau Waweru
B531/EW/1660/2022
Department of Mathematics and Statistics
University of Embu

This research project has been submitted for examination with our approval as the University Supervisors.

Signature: ..... Date: 25/08/25

Dr. Edwine Benson Atitwa
Department of Mathematics and Statistics
University of Embu

Signature: ..... Date: 25/08/2025

Dr. Zakayo Ndiku Morris
Department of Mathematics and Statistics
University of Embu

Signature: ..... Date: 22/08/2025

Dr. Peter Ntoiti Kailemia
School of Nursing
Meru University of Science and Technology

DEDICATION

With my deepest gratitude and love, this work is dedicated to special people who have shaped, supported and inspired me through the journey. To my loving parents, Jane Waweru and Dedan Waweru, you have been a constant source of love, encouragement, unwavering support, and endless sacrifices, which have shaped the person I am today. You instilled in me the values of hard work, perseverance, hope and above all, faith in God. This achievement is as much yours as it is mine. To my wonderful brothers, Samuel Muhia Waweru, Antony Kang'ethe Waweru and Brian Maina Waweru. Thank you for your encouragement and genuine belief in my abilities. Each of you has contributed in a unique way, and I am forever grateful for the special bond we share. To my dear uncle Godfrey, you have been a source of guidance, advice, generosity and invaluable support. Finally, I dedicate this work to my dear self for daring to dream big, staying on track and embracing the struggles and not giving up at whatever cost, even when the road seemed very uncertain.

ACKNOWLEDGEMENT

I am deeply grateful to God for granting me the grace, strength, good health and perseverance that have brought me to the successful completion of this research project. I would like to express my sincere gratitude to my supervisors, Dr. Edwine Benson Atitwa, Dr. Zakayo Ndiku Morris and Dr. Peter Ntoiti Kailemia, for their invaluable guidance, support and mentorship throughout my research journey. Your rich insights and encouragement have been instrumental in guiding and shaping the direction of this research work. I am truly grateful for your effort and time that you invested in this work and for my academic growth. I also wish to acknowledge and thank the Department of Mathematics and Statistics at the University of Embu and the University at large for providing the necessary environment for carrying out this research. I also wish to acknowledge my classmates, Lawrence Mbuu Matuku, Washington Kimani Wainaina, Bernard Kimani Kamau and Felix Ouma for the discussions, encouragement and shared experiences. Your camaraderie made sure that the journey was lighter and more enjoyable. To my parents, you believed in me from the very beginning, and your love and support have been the bedrock of my life. I will forever be indebted to you. I also wish to thank myself for the commitment, sacrifices, late nights and resilience which have made it possible. It has not been easy, but it has been worth every moment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

CDC	Centre for Disease Control
DEB	Disorder Eating Behaviour
ED	Eating Disorder
EHR	Electronic Health Record
FHR	Fatal Heart Rate
FBC	Female Breast Cancer
GBCI	Global Breast Cancer Initiative
GLOBOCAN	Global Cancer
GoK	Government of Kenya
IARC	International Agency for Research on Cancer
LMICs	Low- and Medium-Income Countries
MCSMM	Multi-Chain Semi-Markov Model
MOH	Ministry of Health
NCCP	National Cancer Control Program
NCI	National Cancer Institute
WHO	World Health Organization

LIST OF SYMBOLS

$\boldsymbol{\pi}$	Pi
$\boldsymbol{\tau}$	Tau
$\boldsymbol{P_{ij}}$	Probability of moving from state *i* to *j*
$\boldsymbol{\pi_i}$	Steady state of breast cancer at state *i*
$\boldsymbol{S_i}$	Breast cancer state *i*

ABSTRACT

Breast cancer is a major health burden not only globally. It is the most commonly diagnosed type of cancer globally and in Kenya. In 2022, 7,243 new cases of breast cancer were reported, accounting for approximately 16.2% of all cancer cases diagnosed, with a mortality rate of 11.6% which translates to 3,398 deaths. This study aimed to determine the prevalence and analyze female breast cancer (FBC) prognosis between diagnosis and treatment, taking a case study of two counties in Kenya. Data for this study was obtained from two cancer registries in two county hospitals with a sample of 300 health records. After data cleaning, 150 records were eligible for analysis. Key variables of interest in the study were staging information of FBC at diagnosis and treatment, time taken between diagnosis and treatment, as well as the waiting time before transitioning to the subsequent stage. One of the approaches that can be used to gain insight into how breast cancer progresses over time is the application of semi-Markov analysis, which was used to analyze the prognosis of breast cancer in two counties in Kenya. This was obtained by determining the prevalence of FBC at diagnosis and at treatment and finding the transitional probabilities between different cancer states. The results of the analysis showed that FBC stage III was the most prevalent at diagnosis and treatment initiation, with a prevalence of 36% and 34.7% respectively. The chances of moving from stage II to stage IV increase significantly from 1.6% at $t = 1$ to 45.49% at $t = 72$. There is a higher probability of remaining in stage III (0.985) at $t = 1$ compared to a probability of 0.326 at $t = 72$. The results outline the necessity of timely diagnosis and initiation of interventions, which may help in clinical decision-making, resource allocation, and inform public health policies.

CHAPTER ONE
INTRODUCTION

1.0 Introduction

This chapter gives a presentation on the background in the formation of the study, the problem statement, general and specific objectives, justification, significance, and the scope of the study.

1.1 Background Information

The World Health Organization (WHO) lists cancer as one of the top causes of mortality worldwide (WHO, 2025). According to the health organization, cancer is a broad category of illnesses that can originate in any organ or area of the body in humans when cells grow uncontrolled and unnaturally and then spread to other areas of the body and eventually other organs. In the year 2020, statistics released by the WHO showed that approximately one in every six deaths was a result of cancer. The report further indicated that a third of the deaths that were cancer-related were due to high body mass index, use of tobacco, consumption of alcohol, lack of physical activity, and low vegetable and fruit intake (WHO, 2025).

Statistics released by the WHO showed that in the year 2022, 20 million new cases of cancer were reported, while roughly 10 million people died from cancer (WHO, 2024). The statistics showed that at least one in every five persons had a likelihood of being diagnosed with cancer in their lifetime. In terms of cancer mortality, it was approximated that 1 in every 12 women and 1 in every 9 men die of cancer (WHO, 2024). This underscores the growing burden of cancer around the entire globe. Breast cancer, lung cancer, cervical cancer, liver cancer, prostate cancer, skin cancer, colon, rectal, and other cancers are among the most common cancer types.

Cancer is classified into various types according to the location in the body where it first develops or according to the tissue or fluid from which it originates. Some of the most common types of cancer include Breast cancer, Lung cancer, Cervical cancer, Liver cancer, Prostate cancer, Skin cancer, and Colon and Rectum cancers, among others. According to McGarvey *et al*. (2022), cancer presents a significant source of burden not only in the United States but also in the entire globe. Cancer continues to

pose a significant source of economic burden and has projection estimates of approximately $246 billion by the year 2030 in the United States of America (McGarvey *et al*., 2022). The global cancer statistics released in 2020 by GLOBOCAN revealed that breast cancer, which is more prevalent in females, surpassed lung cancer, which is more prevalent in males gender in terms of prevalence rate (Sung *et al*., 2021).

In the year 2022, similar statistics were released by the WHO, and it revealed that breast cancer was the most prevalent type of cancer among females (WHO, 2024). When both sexes are combined, breast cancer becomes the second most diagnosed type of cancer. However, the total number of new cases increased from 2.26 million in 2020 to approximately 2.31 million in 2022 (IARC, 2024). According to the 2020 report released by the WHO, the most common cancer was breast cancer, which accounted for 2.26 million deaths out of the approximate total number of deaths of around 10 million worldwide (WHO, 2025). According to the 2020 GLOBOCAN statistics, an estimated number of deaths of 520,158 occurred out of the 801,392 new cancer cases reported in Sub-Saharan Africa in 2020 (Bray *et al*., 2022). In all these cases, Breast Cancer accounted for the largest percentage of cases in both men and women.

In Kenya, cancer ranks third in terms of cause of death, behind cardiovascular and infectious diseases (Jani *et al*., 2021). Statistics show that between 2012 and 2018, the cancer-related mortality rate rose by approximately 16%, and the number of new cancer cases is expected to be on a rising trajectory for the next two decades by over 120% (Jani *et al*., 2021). According to the International Agency for Research on Cancer (IARC) and the GLOBOCAN report, breast cancer was ranked as the most commonly diagnosed cancer by type. It constituted approximately 16% of the total number of new cases in Kenya in the year 2020 (Ferlay *et al*., 2021). WHO adds to this and notes that breast cancer is the most prevalent cancer type not only in Kenya and Africa at large but also in the entire globe (WHO, 2024).

According to Bray *et al*. (2022), the burden of cancer in Africa is rising, especially in Sub-Saharan Africa. States in this region need to have enhanced systems of surveillance that can measure the magnitude of the cancer problem so that monitoring and planning for the disease can be enhanced. Olaleye & Ekrikpo (2017) noted that Sub-Saharan Africa has had relatively higher mortality rates from cancer, which has

been attributed to the changes in diet, lifestyle, and population dynamics across Africa. The two researchers concluded that there is an urgent need for cancer diagnosis and treatment across the region.

In a bid to control cancer rates by reducing the incidence and mortality as well as improving the life of cancer patients, the Government of Kenya (GoK) came up with the National Cancer Control Program (NCCP). The first draft of the program was implemented between 2017-2022, and the second program is to be implemented between 2023-2027. The program aims to make sure that evidence-based strategies are implemented to aid in prevention, early detection, diagnosis, treatment, and palliation, as well as optimally using the available resources and using them in the best way possible. According to NCCP (2023), projections show that by the year 2028, Kenya will have an estimated 58,000 new cases of cancer (MOH, 2023). In all these cases, it is estimated that the leading cancer type is breast cancer. This, therefore, calls for timely planning to aid in cancer control interventions, which include prevention, early detection, diagnosis, and treatment, as well as palliation.

In 2021, WHO launched the Global Breast Cancer Initiative (GBCI). The initiative aimed at reducing the increasing breast cancer burden and preventing 2.5 million deaths by the year 2040 (WHO, 2023). During the same year, Kenya launched the Breast Cancer Screening and Early Diagnosis Action Plan 2021-2025, which was in line with the GBCI. The Kenya Cancer Policy 2019-2030 has eight key themes that aid in the control of cancer, and among the themes are access to quality, affordable, and sustainable cancer care and improved survivorship care coordination (MOH, 2020). The policy also aims to ensure the promotion of cancer research and surveillance and the sustainable financing of cancer care.

Pillar five of the Kenya Cancer Control Strategy (2023-2027) highlights the importance of strengthening cancer research in Kenya and incorporating the findings in cancer protocols and policies. The National Cancer Taskforce report, released in 2022, identified that there is low prioritization of cancer research and called for cancer research to address the most prevalent cancers in the country for control. The available breast cancer research has focused on genomic and molecular aspects of the disease. The continued breast cancer burden in the country indicates that it is an area that requires prioritized cancer control strategies informed by relevant research.

1.2 Problem Statement

The increased cancer mortality rate has become not only a national but also a global threat. Different policies and programs, like the NCCP, have been formulated to ensure proper planning and care for cancer patients. However, these policies have been formulated without taking into account mathematical models, which are critical in understanding the future progression of breast cancer. When coming up with the policies and legislations, the policymakers have focused only on epidemiological aspects of breast cancer with limited attention to the dynamics of the disease. This study, therefore, sought to bridge the gap by analyzing breast cancer progression and informing accurate policy-making and legislation towards halting the mortality rate and progression of the disease by having proper resource planning by the governments; national and county governments. To do this, there was a need to mathematically understand the transition of cancer patients from one stage of breast cancer to the other. In semi-Markov modeling, transitional probabilities are critical in understanding the progression of different chronic diseases, breast cancer being one of them. This study, therefore, sought to determine the prevalence of breast cancer, obtain the probabilities between different breast cancer stages and construct a probability matrix. The probabilities were then used to find the steady states of breast cancer and estimate the management cost of offering the necessary healthcare to patients in different breast cancer states for proper resource planning and allocation.

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study was to analyze the prevalence and prognosis of breast cancer during management in Kenya.

1.3.2 Specific Objectives

1. To determine the prevalence of breast cancer at each stage in Kenya
2. To construct the transition probability matrix of breast cancer in Kenya
3. To compute the steady states of the different breast cancer stages in Kenya.
4. To estimate the cost of managing breast cancer at each stage in Kenya.

1.4 Research Questions

1. What is the prevalence of breast cancer at each stage in Kenya?
2. What are the transition probabilities of moving from one stage of breast cancer to the other in Kenya?
3. What are the steady states of breast cancer at each stage in Kenya?
4. What is the estimated cost of managing breast cancer at each stage in Kenya?

1.5 Justification

Breast cancer has been noted to be the second most diagnosed type of cancer, accounting for more than 2.31 million cases in 2022 (IARC, 2024). The majority of breast cancer cases and deaths reported occur in countries categorized as low and medium-income. In Kenya, breast cancer is the leading type of cancer in terms of incidence rate, with an approximate number of 6,799 new cases in the year 2020 (MOH, 2021). In 2022, breast cancer had the highest incidence rate among all cancer types diagnosed in both males and females. A total of 7,243 new cases of breast cancer were reported in the year. Out of the 29,317 cancer deaths that were reported, 18,003 deaths were of female gender. To gain a comprehensive understanding of the progression of breast cancer and develop effective mitigation and control strategies, it is crucial to analyze the transition from one stage to the other. Transition probabilities of breast cancer best give the progression dynamics of the disease so that necessary measures can be put in place. Finding the steady states of breast cancer distributions is critical in gaining insight into the proportions of breast cancer patients in each particular state. To understand the states that are predominant and how the progression of breast cancer stabilizes over time, obtaining the steady states is important.

To identify critical states that have a central role in the prognosis of breast cancer and its dynamics, the steady states were of great importance. By using steady states, it is possible to assess the relative level of care that is required for cancer patients in each state. Calculation of the total financial impact after assessing the estimated cost per patient in every state is critical in the estimation of the financial impact of offering care to breast cancer patients. This is of critical essence not only to national and sub-national governments but also to households, so that there is proper planning and preparedness to offer relevant cancer care to patients.

1.6 Significance of the Study

The study findings will help researchers and healthcare professionals understand how breast cancer evolves from one state to another. To control breast cancer, it is important to understand the progression and the different dynamics of the disease. Transitional probabilities best give valuable insights into the dynamics of breast cancer. The findings of the study will help promote preparedness in the health sector and come up with early detection strategies by the MOH in both the national and county governments. The steady states obtained will also give a long-term distribution of the cancer patients across the different states when the system is at equilibrium. It will also enhance improved breast cancer treatment strategies and patient care from the MOH. The findings will enable proper and accurate resource allocation from both levels of Government, as well as the affected households. The study findings will also quantify the financial impact breast cancer has on patients and health systems. This will play a significant role in the allocation of resources from the National and County governments, as well as the family and relatives of a patient. The study findings will also inform policy development in health care as far as breast cancer is concerned in terms of resource allocation and planning.

1.7 Scope of the Study

The area of study in this research was two Kenyan counties, Meru and Nyeri, which are in the Mt. Kenya region. According to a report released by the Ministry of Health, cases of breast cancer in these two counties have been on a rising trajectory, which calls for concerted efforts in combating it and taking the necessary care for those who are diagnosed with the disease. Additionally, these two counties are unique as they give an average representation of cancer care since both counties have a cancer center with the necessary equipment and resources to offer at least the minimum care required for cancer patients. This study incorporated the use of secondary data obtained from breast cancer health records kept and maintained in cancer center hospitals from the two counties of focus for a period between 2013 and 2023. The variables this study focused on included staging information of breast cancer patients at diagnosis, cancer stage pathways of different patients, prescribed average treatment modalities per stage, and the outcomes.

1.8 Key Assumptions of the Study

1 Data completeness and accuracy. In this study, it is assumed that the data was collected accurately and recorded in the health records that were obtained from the two cancer center hospitals.

2. This study also assumed that all the breast cancer patients sought medical attention and their transitions were recorded.
3. This study also assumed that the data was accurately recorded in the cancer registry and that no external interventions affected the progression of the disease.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter presents a theoretical and empirical review of semi-Markov modeling and its application in the determination of transition probabilities and steady states, as well as its use in cost estimation analysis across various fields.

2.1 Theoretical Review

A Markov Chain is a powerful model that describes the progression of something through different states. The point of focus in a Markov Chain is the Markov Property, which requires that, given the history of a subject, the present state only depends on the recent past. This, therefore, means that the process is memoryless since the chances of an event occurring depend not on the entire history of occurrence but only on the recent history. In this study, a semi-Markov process was considered. In a semi-Markov process, the amount of time spent in each state before a transition to the next state occurs is an arbitrary random variable that depends on the next state the process will enter (Asanjarani *et al*., 2022). This is one of the major Markov assumptions that is taken into account in a semi-Markov model and was also employed in this study. Consider any random process $X = \{X_1, X_2 \ldots\}$. The process can be referred to as a Markov Chain if the Markov property is satisfied. The property can be expressed as

$$P\{X_{t+1}/X_t = X_{t,}X_{t-1,} \ldots, X_0\}$$

Therefore, a semi-Markov process is a stochastic process where, after entering a given state i, the process randomly determines the time taken to transition from the state. The waiting time before the transition can be represented as τ, and therefore, the

$$P_{(i,\tau)j} = \Pr\{X_{n+1} = j | X_n \}$$

2.2 Prevalence of breast cancer

The Center for Disease Control (CDC) defines prevalence as the number of cases of a disease, the number of people infected, or the number of patients with certain attributes during a particular period (CDC, 2023). It is usually expressed in the form of a rate. In the health sector and medical epidemiology, prevalence is defined as the proportion of persons with a given condition at a specified point in time (Tenny &

Hoffman, 2017). Prevalence over a period gives a more accurate picture of the overall prevalence since it considers all persons with the condition between certain time instants.

Based on the need to plan for cancer in the future, it is important to have a long-term view of breast cancer prevalence. Mitchell *et al.* (2021) conducted a study to establish cancer prevalence before industrialization. In particular, they sought to determine whether the prevalence of cancer before the Industrial Revolution and tobacco was different from the current prevalence. After the analysis of data obtained from different cemeteries in the Cambridge area, they found that 3.5% had evidence of metastases. After comparing the findings to the proportion of people who died of bone metastasis as of then, it was evident that cancer was less prevalent before the Industrial Revolution, with between 9% and 14% in adults (Mitchell *et al*., 2021).

Hong *et al*. (2020) conducted a study to report cancer statistics in Korea in terms of incidence, survival, mortality, and prevalence in 2017. Data from the Korea National Cancer Incidence Database covering all the years between 1999 to 2017 was analyzed and was used to evaluate the incidence, prevalence and survival rates. A trend analysis was performed, and the results obtained revealed that cancer incidence rates increased annually by 3.5% between the years 1999 and 2011 and then dropped by 2.7% in the same period. Patients who were diagnosed with cancer between 2013 and 2017 had a survival rate of 70.4% and this showed a prevalence of approximately 1.87 million cases.

Given the growing burden of colorectal cancer and its rapid change within populations, Wong *et al*. (2019) conducted a review to describe the prevalence of this type of cancer in Asia. To achieve this, they used data from the International Agency for Research on Cancer (IARC) database obtained from the WHO. In addition to prevalence, they summarized the risk factors that were associated with colorectal cancer in Asia. Data on the prevalence of the disease for five years was analyzed, and it showed that Asia had the highest prevalence of the disease, with a total of 4,789,635 new cases for the five years. From the statistics, it was evident that men and women were affected by the disease almost equally, with 2,595326 and 2,194,309, respectively. The five-year prevalence was represented as 62.8/100,000.

2.3 Transition Probability Matrix and Steady States

Various studies have been carried out by different researchers who employed semi-Markov modeling to achieve the expected results. A study conducted by Vargas-Calixto *et al*. (2022) employed the use of Multi-Chain Semi-Markov Models (MCSMMs) in determining the evolution and progression of fatal heart rate (FHR) and uterine pressure (UP) patterns. According to the researchers, the models helped in estimating the probability of transitioning between the different patterns of FHR and UP. The researchers based their analysis on the methodological approach proposed by Onu *et al*. (2017), who used multi-chain Semi-Markov models in modeling varying-time behaviour of physiological processes that follow particular sequences. These models take into consideration the modeling of Semi-Markov chains as a function of time taken in the particular transition state.

The researchers had three possible states of the fatal heart rate (FHR), namely BAS, DEC, and ACC, while uterine pressure had only two states: RIN and CON. The following formula was used to estimate the transition probabilities between different states.

$$P(i \rightarrow j) = \frac{n(i \rightarrow \mathrm{j})}{\sum_{j}^{3} n(i \rightarrow \mathrm{j})}$$

where $P(i \rightarrow j)$ is the proportion of patients that moved from the state i to j, given that $n(i \rightarrow \mathrm{j})$ is the number of transitions that occur from state i to j. The researchers did not take into account self-transition probabilities but rather took into account the dwell time in each state. The transition probabilities were estimated between different states and the dwell times at each time epoch. The results obtained by the researchers showed that using Multi-Chain Semi-Markov Models is a reliable approach to analyzing certain risks out of the transition probabilities.

Mirzapour *et al*. (2019) conducted a study to develop a Markov and a semi-Markov model for estimating and predicting respiration-induced intra-fraction motion by using a surrogate motion signal. The researchers aimed to model transition probabilities between different motion phases contained in two-cycle types using a discrete semi-Markov model. This model was used to model the time spent at each transition state before transitioning to the next state. In the study, MRI data for nine patients was used to evaluate the performance of the Markov and semi-Markov

models that had been developed. The researchers used the semi-Markov modeling since it provided flexibility and was more accurate for obtaining short-term trajectory prediction of the transitions. The researchers were able to calculate probabilistic transitions between various motion cycle states in their study.

Ahmad *et al*. (2019) conducted a study to estimate the annual probabilities of changing disability levels among patients in Australia who had Relapsing-Remitting Multiple Sclerosis (RRMS) and identified that there was a slow progression of the disease. Multiple sclerosis is a health condition that affects the central nervous system and, over time, leads to disability in the patient. The diseases had different episodes of acute neurological deterioration, which are later followed by either partial or full recovery, which is commonly referred to as relapsing-remitting. To achieve the objective of the study, the researchers applied a three-state continuous Markov model using longitudinal data of the patients diagnosed with the disease.

Ahmad *et al*. (2019) studied the transition probabilities between three disability states of the disease, namely, No/Mild, Moderate, and Severe, to explain the progression of the disease. They employed the use of the 'msm' package in R software, which enables fitting a multistate model, and defined a 3 by 3 matrix with different transition intensities based on the disability state. The researchers found that a significant proportion of individuals who had the disease (Multiple Sclerosis) tended to remain in the same disability state over the period under study, which revealed that there was a slow progression of the disease.

A study was conducted by Rodriguez *et al*. (2021) whose main objective was to use Electronic Health Record (EHR) data to estimate the transition probabilities between different states of paediatric eating disorders (ED). In their study, they note that EHR data is useful in estimating the transition probabilities between different states in a model, but there is no specific guideline for the appropriate method to estimate the probabilities. The researchers applied three potential methods in estimating the transition probabilities after having mapped three health states of ED, namely, bulimia nervosa, anorexia nervosa, and any other eating disorder.

The methods applied in estimating the transition probabilities between the health states were the multistate Markov model, simple last-first proportions, and the independent survival model. EHR data was obtained from outpatient, inpatient, and

different emergency department visits by patients who were diagnosed with eating disorders. The findings of the study revealed that the transition probabilities differed significantly across the three estimating methods (Rodriguez *et al*., 2021). The researchers noted a challenge in using the EHR data because some of it is incomplete, but they also noted that the data is important in bringing out real-world insights regarding healthcare patterns and outcomes.

Another study was conducted by Baechle *et al.* (2019) to test the prevalence and the development of the disorder eating behaviour (DEB) by the use of establishing transition probabilities of the disorder in the adolescents. The data used by the researchers in the study was an example of a longitudinal type of data with 1,318 observations which were used to calculate the probability of transition within the age categories of the age-specific transition probability. A first-order Markov model was incorporated in the study to estimate the probabilities and were fitted using the regressive logistic models. The "current" state, which generally represented the present DEB state of the respondents and those who did not have any disorder, was used as the dependent variable, while the "previous" state was used as the independent variable. The transition probabilities obtained from the model enabled the researchers to have deeper insights into the course of DEB from one state to the other over time.

2.4 Cost Assessment and Estimation

McGarvey *et al.* (2022) conducted a retrospective study that aimed at getting a good understanding of the possible effects of early diagnosis by enforcing healthcare costs on various cancer patients depending on the type and the stage of being diagnosed. The analysis was done using Optum's de-identified Integrated Claims-Clinical Data. The data included data from commercially insured members and data from Medicare Advantage. The researchers extracted data from adults who were newly diagnosed with solid cancer tumours, the cancer stage at diagnosis between 2016 and 2020, as well as continuous enrollment for not less than one month after the diagnosis. The mean standardized costs for patients with breast, lung, cervical, ovarian, and colorectal cancers were calculated each month using an annual and cumulative basis for the four years of the study. For every month, costs for those who underwent continuous enrollment into the three types of cancer were taken into account to come up with a standard cost for each cancer care per month. The researchers found out that the mean costs of cancer care generally increased annually by stage of diagnosis.

Another research was conducted by Šlegerová & Kopečková (2023) and aimed at estimating the cost-effectiveness in the treatment of breast cancer after adding pertuzumab in the treatment of breast cancer metastatic. To achieve the objectives of the study, the researchers applied semi-Markov modeling with different transition states. The states were derived from the four different treatment cost states given as first-line medication, no medication, next-line medication, and finally, death. The model was used to estimate the costs from the perspective of the healthcare payers. The researchers used secondary data in their study, which was obtained from the Czech clinical practice recorded in their health records register. The researchers indicated that in health economics analysis of chronic diseases, Markov processes are critical in estimating the duration of time taken in different states and more so in cost-effectiveness studies. The cost states used in the study were directly connected to the different treatment stages of breast cancer patients.

2.5 Research Gap

Despite the advancement and application of semi-Markov models in different healthcare fields, the models have not been used to understand the progression dynamics of breast cancer. The existing literature has put more focus on both Markov and semi-Markov models being used in different health applications. Dalabanjan *et al.* (2021) did a study on the prognosis of cancer and applied a Semi-Markov Process. However, the gap in their study was that they did not have patient data to study the prognosis of cancer. In Kenya, mathematical models like regression modeling on tumour growth have been applied to investigate cancer development and progression (Yin *et al.*, 2019). Given that the transitions of breast cancer are non-homogeneous, semi-Markov models are appropriate for understanding the transition and progression dynamics of the disease. This study, therefore, seeks to use breast cancer patient records data to gain insights into the transition dynamics and estimate the management cost of offering necessary care to breast cancer patients.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter outlines the sources of data that will be used in this study and how it will be analyzed to meet the objectives of the study.

3.1 Data Source and Description

The data for this study was obtained from population-based cancer registries and health records of breast cancer patients from two public hospital registries in Kenya, with a sample of 300 breast cancer health records. The data obtained included information on patients with breast cancer diagnosis, staging information, and time taken before treatment initiation.

3.2 Data Analysis

After obtaining the required data, it was entered into Microsoft Excel, where it was cleaned to determine the number of valid cases that would be taken into consideration. After data cleaning, 150 records were eligible for analysis, which was implemented in R version 4.4.2. This study classified breast cancer into 4 possible states. The 4 states can be denoted as $S_1, S_2, S_3,$ and S_4.

Let

$S_1 = Breast\ cancer\ Stage\ I$

$S_2 = Breast\ cancer\ Stage\ II$

$S_3 = Breast\ cancer\ Stage\ III$

$S_4 = Breast\ cancer\ Stage\ IV$

To study the progression of cancer as a Markov chain using the states listed above, we had the following assumptions to consider.

i) States S_1 to S_4 are the four possible states of cancer that patients can be diagnosed in.

ii) The time taken in each of the states is independent. This means the distribution of time taken in each state does not depend on the time taken in the other state.

If a patient changed from one state to another, it was considered that our event of interest had occurred. Therefore, the state space *S* was represented as

$$S = \{S_1, S_2, S_3, S_4\}$$

FBC staging can be at any of the four possible states from S_1 to S_4, which are the possible states where treatment can be initiated to offer care to them. It is worth noting that the data used in this study was rightly censored since, at the time of data collection, some of the patients had not transitioned through all the four possible states.

In this study, $D_{i,j}(t)$ was taken to represent the distribution function of the length of stay of a patient in a particular state, say S_i, before treatment initiation, where they were staged to know the appropriate care to offer to the patients. The distribution $D_{i,j}(t)$ was continuous at arbitrary points, which meant it was a semi-Markov process.

3.3 Prevalence of Breast Cancer

Exploratory data analysis was conducted to gain more information from the data and, specifically, the distribution of the number of breast cancer cases at each stage. The number of cases per stage is divided by the total number of breast cancer cases and then multiplied by 100 to express it as a percentage to determine the prevalence of breast cancer per stage.

$$4 \quad = \frac{ni}{N} \times 100 \qquad (1)$$

where, n_i – number of breast cancer patients at stage *i,* given *i=1,2,3,4*

N – Total number of breast cancer patients

3.4 Transition Probability Matrix

In this study, semi-Markov models were employed to capture the transitions of breast cancer between different states in diagnosed patients. Semi-Markov models allow flexible waiting times before patient states change and transition to another. This flexibility of the sojourn times allows the use of semi-Markov models to estimate breast cancer progression among patients. This is because the waiting times or the time spent in each of the states vary significantly from patient to patient. This study focuses on the transition of breast cancer between the time they are screened and staged at diagnosis and the time the patients seek treatment in public hospitals.

One of the key components of the semi-Markov models is the intensity matrix (Q matrix). The matrix captures the instants of transition rates between different states in a system, in this case, cancer stages. The rates help understand the intensity at which patients transition from one stage to another. In semi-Markov modeling, the Q matrix is the foundation for determining and deriving the transition probability matrix *P(t).* The intensity matrix is defined as follows:

$$\mathrm{Q} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{pmatrix} \qquad (2)$$

where q_{ij}, shown in the matrix, represents the rate at which patients move from one state, *i* to *j.* One of the key principles of intensity matrices is that the leading diagonal entries, q_{ii}, are given by taking the negative sum of the other entries in a particular row. This can be represented as shown below;

$$q_{ii} = -\sum_{i \neq j}^{n} q_{ij} \qquad \forall\, i, j = 1,2,3,4 \qquad (3)$$

n is equal to the four possible states of breast cancer considered in this study.

Estimates were obtained by using the observed transition counts and the total time spent in a particular stage of cancer to determine the entries of the intensity matrix. From the data obtained, all the transition counts from one stage to another were summarized in a transition count matrix N. The matrix gives the record of transition counts observed between a pair of states, say, i and j. The matrix N can be generalized as shown below.

$$\mathrm{N} = \begin{pmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \\ n_{41} & n_{42} & n_{43} & n_{44} \end{pmatrix} \qquad (4)$$

where transition counts from stage i to j, provided that $i \neq j$ is given by N_{ij}. Counts of patients who were diagnosed in stage i and remained in the same stage are given by N_{ii}, which represent self-transitions in a particular stage i.

The estimation of the Q matrix also requires that we have the transition time matrix, which is used together with the transition count matrix to obtain the intensity matrix.

The transition time matrix accounts for the time taken by patients in a given state i before they transit to state j. This includes the self-transitions for the patients who remain at the same stage over time. This study considers the waiting time in months. The generalization of the transition time matrix is given as follows;

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ t_{41} & t_{42} & t_{43} & t_{44} \end{pmatrix} \quad (5)$$

where T_{ij} is a representation of the total time spent in stage i before transitioning to stage j. T_{ii} represents the total time taken in stage i without moving to another stage. Entries in the transition time matrix are calculated as follows

$$T_{ij} = \sum_{k \epsilon S_{ij}} t_k \quad \text{for } i \neq j \quad (6)$$

$$T_{ii} = \sum_{k \epsilon S_{ii}} t_k \quad \text{for } i = j \quad (7)$$

$$\forall\ i, j = 1,2,3,4$$

where;

t_k represents the time a patient k spent in a given stage i.

S_{ii} shows the set of patients who did not move from stage i and, therefore, remained in the given state.

S_{ij} represents the set of patients who moved from stage i to stage j, as observed in the data.

When determining the intensity rates in the Q matrix, the off-diagonal elements q_{ij} were given by the following formula;

$$q_{ij} = \frac{N_{ij}}{T_{ij}}, \forall\ i, j = 1,2,3,4 \quad (8)$$

$$q_{ii} = \frac{N_{ii}}{T_{ii}}, \forall\ i = 1,2,3,4 \quad (9)$$

Once the operations in the above equations are performed, the resultant matrix must be modified to suit the principles of the Q matrix. The principles require that the sum of all the entries in a given row in the Q matrix should add up to 0. Equation (3) is taken into account to satisfy the principles. The diagonal elements represent the rate

at which patients exit a given stage. The other elements off the diagonal of the matrix are non-negative, as shown below;

$$q_{ij} \geq 0 \quad \text{for } i \neq j \tag{10}$$

All the rows in the intensity matrix should suit the following principle;

$$\sum_j q_{ij} = 0 \quad \forall\, i = 1,2,3,4 \tag{11}$$

The modified matrix will be given as follows

$$\mathrm{Q} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{pmatrix} \tag{12}$$

It is worth noting that the matrix above in equation (12) is just an empirical Q matrix calculated from the observed waiting times before transition and the transition counts. Therefore, to make sure that the estimates of the Q matrix account for potential biases and filter any statistical noise, it is important to carry out model fitting of the Q matrix. Therefore, the empirical Q matrix is fitted in a semi-Markov model, which will ensure that the model assumptions, like that of continuous time distribution, are satisfied.

3.5 Model Fitting

Given that the observed waiting time were continuous, as seen in the data, a semi-Markov model was fitted in order to estimate the Q matrix. The likelihood function for the model fitted is given as follows;

$$L = \prod_{i=1}^{n} \prod_{i=1}^{m} \left[q_{ij} f_{ij}(t_{ij}) \right]^{N_{ij}} \tag{13}$$

where;

q_{ij} is the calculated transition intensity from state i to state j

t_{ij} is the total time taken before transitioning from state i to state j

N_{ij} represents the count of the transitions from state i to state j

$f_{ij}(t)$ shows the density function of the distribution of the time taken in state i before moving to the next state j.

m represents the number of possible states in the system

n shows the number of patients considered in the study.

The likelihood function finds sets of parameters in the intensity matrix that help in maximizing the probability of observing the data. It is a product of the density function of the distribution of sojourn times and the observed data, the empirical intensities.

The log-likelihood of equation (14) gives simplified derivatives, which are critical in the optimization of the estimates of the Q matrix. Estimates of complex models like the semi-Markov are refined by the log-likelihood function, which facilitates a more robust estimation of the Q matrix. The function is given as follows;

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{m} N_{ij}\left(\log q_{ij} + \log f_{ij}\left(t_{ij}\right)\right) \qquad (14)$$

Therefore, the resultant entries in the estimated Q matrix of the maximum likelihood estimate are given as

$$\hat{q}_{ij} = \frac{N_{ij}}{\sum_j \int_0^\infty t \cdot f_{ij}(t)dt} \qquad (15)$$

As such, the intensity matrix obtained after the Maximum Likelihood Estimation (MLE) is represented as follows

$$\hat{Q} = \begin{pmatrix} \hat{q}_{11} & \hat{q}_{12} & \hat{q}_{13} & \hat{q}_{14} \\ \hat{q}_{21} & \hat{q}_{22} & \hat{q}_{23} & \hat{q}_{24} \\ \hat{q}_{31} & \hat{q}_{32} & \hat{q}_{33} & \hat{q}_{34} \\ \hat{q}_{41} & \hat{q}_{42} & \hat{q}_{43} & \hat{q}_{44} \end{pmatrix} \qquad (16)$$

3.6 Transition Probabilities

Since the process is continuous-time semi-Markov, the determination of the transition probability from one state to another at time t will be expressed by;

$$P(t) = \{P_{ij}(t)\} \qquad (17)$$

given that $P_{ij}(t)$ is the probability of being in state j after time t, given that the patient started at state i, which can be given as

$$P_{ij}(t) = \Pr(\text{being in state } j \text{ at time } t \mid \text{started at } j) \qquad (18)$$

The transition probabilities above $(P_{ij}(t))$ satisfy the Kolmogorov forward differential equations of the following form;

$$\frac{dP(t)}{dt} = \hat{Q}P(t) \qquad (19)$$

where;

t is time in months.

$\hat{Q}$ is the transition intensity matrix.

$P(t)$ is the transition probability matrix.

To find the solution to the differential equation (18), the matrix exponential based on the Taylor series expansion gives the solution.

$$P(t) = e^{\hat{Q}t} \tag{20}$$

The matrix exponential is based on the following series

$$e^{\hat{Q}t} = I + \hat{Q}t + \frac{(\hat{Q}t)^2}{2!} + \frac{(\hat{Q}t)^3}{3!} + \frac{(\hat{Q}t)^4}{4!} + \cdots + \frac{(\hat{Q}t)^n}{n!} \tag{21}$$

where I is the identity matrix, t is time in months, and $\hat{Q}$ is the transition intensity matrix. Therefore, the specific transition probabilities from state i to state j are given as

$$P_{ij}(t) = (e^{\hat{Q}t})_{ij} \tag{22}$$

with a probability condition given as

$$\sum_j P_{ij}(t) = 1 \ \ \forall \ i\text{'s} \tag{23}$$

The transition probabilities obtained will be represented in a probability matrix P, which can be generalized as shown below.

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} \tag{24}$$

where P_{ij} is the probability of transitioning from breast cancer stage i to stage j. According to Igwenagu & Egemba (2021), each row must sum up to 1 to satisfy the probability condition outlined in equation (25), as shown below;

$$P_{11} + P_{12} + P_{13} + P_{14} = 1 \tag{25}$$

$$P_{21} + P_{22} + P_{23} + P_{24} = 1 \tag{25}$$

$$P_{31} + P_{32} + P_{33} + P_{34} = 1 \tag{25}$$

$$P_{41} + P_{42} + P_{43} + P_{44} = 1 \tag{25}$$

3.7 Determination of Steady States

To achieve the second objective of calculating the steady states of the different cancer stages, we took into consideration the steady-state probabilities. The steady-state probabilities give the distributions of patients in each of the four states in a long-term view when the system is at equilibrium, where the probabilities of being at a particular state remain constant over time. For Markov and Semi-Markov chains, steady states are arrived at by solving a system of linear equations involving the transition probabilities obtained in the first objective and the probability of a patient being at a

particular stage and a given time. Using the transition probabilities obtained in this study, equations describing each of the breast cancer states will be written, describing how populations in each breast cancer state will have changed over time in the long run.

To solve the equations and find the steady-state values of every breast cancer state, we will set each of the equations equal to zero and solve iteratively for the steady-state values. It is worth noting that at a steady state, the rate of change of breast cancer patients in terms of progression to other states will be equal to zero.

If P is the transition probability matrix and π is the steady state row vector, then

P

The system of linear equations is often solved using matrix methods as summarized π given that $\sum_{j=1}^{n} \pi = 1$ (26)
in the equation below;

$$(P^T - I)\pi^T = 0 \text{ given that } \sum_{j=1}^{n} \pi = 1 \qquad (27)$$

where P is the transition matrix.

P^T is the transpose of the transition matrix.

I is the identity matrix.

π is the steady state vector. Its transpose π^T is used for solving the system as a column vector

Steady states tell us where patients are most likely to end up in the long run as the disease progresses, especially when we account for how long they stay in a particular stage.

The above equation (27) expands to a system of equations as shown below

$$\begin{bmatrix} p_{11}-1 & p_{21} & p_{31} & p_{41} \\ p_{12} & p_{22}-1 & p_{32} & p_{42} \\ p_{13} & p_{23} & p_{33}-1 & p_{43} \\ p_{14} & p_{24} & p_{34} & p_{44}-1 \end{bmatrix} \times \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (28)$$

In order to solve for π, there is a need to add a normalization condition which replaces one of the four equations so as to get a solvable system of three linear equations with four unknowns and the normalization. The normalization equation is given below

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \quad (29)$$

The system of linear equations will be solved to find the values of π_1, π_2, π_3, and π_4, which will represent the proportions of the population of patients in each cancer state.

3.8 Breast Cancer Management Cost Estimation

To achieve the third objective, a more structured approach was taken into consideration. Data on direct medical costs and indirect costs by breast cancer patients were quantified in every stage of the disease. The two cancer centers considered in this study keep extensive treatment and billing records of the breast cancer patients, especially those patients who have been diagnosed and staged and treatment was initiated. To achieve the fourth objective, records of treatment cost information was obtained from the cancer centers for a period of five years. Patient-level cost information of staging and diagnostic procedures, chemotherapeutic and hormonal therapy, surgical and radiotherapies, laboratory and imaging tests, and follow-up information was collected. In this analysis, the average costs of two patients in each stage (Stage II, Stage III and Stage IV) was taken in order to obtain the full costs of their treatment so that the average cost to treat a patient in this specific stage can be calculated.

One of the major constraints faced in the process of data collection is that no adequate cost records were available and applicable to patients in Stage I. This gap was occasioned by a number of factors. To begin with, a low level of breast cancer diagnosis at Stage I in the patient records was observed in the two centers. This is consistent with the national wide patterns wherein breast cancer is usually diagnosed at later stages as there is a delay in the screening process, a lack of awareness and also an unavailability of early diagnostics services (Kailemia *et al.*, 2023). The cost of stage I could not be found and therefore a proportional estimation model was used in estimating the cost where it was considered that stage I cost is 32 percent of the average cost of treating stage II as detailed by Sun *et al.* (2018).

The costs incurred in the treatment of each patient were disaggregated into standard items that are normally used when handling breast cancer. These parts entailed diagnostic imaging, biopsies, work in the laboratory, staging assessment, chemotherapy, surgery, radiotherapy, hormonal therapy, and clinical follow-ups. The

arithmetic mean of the sum treatment costs of two representative patients in each stage was used to calculate the average cost per stage by use equation 30.

$$Average\ Cost = (Patient\ A\ cost + Patient\ B\ Cost)/2 \quad (30)$$

$$Stage\ I\ estimated\ cost = 0.32 * Average\ stage\ II\ cost \quad (31)$$

3.9 Diagnostics Checks

After the data was analyzed, diagnostic checks were performed to identify and make corrections to anything that might compromise the quality of the data and findings. Specifically, the steady states were verified and matrix validation was done. This was to ensure that the findings of the study were accurate and will provide reliable results that are informative.

3.10 Ethical Consideration

This study underwent review and approval by the relevant Institutional Ethics Review Committee (IERC), reference number CUIERC/NACOSTI/651. The permit covers the use of anonymized secondary data without any additional consent. Fully anonymized secondary data from the registries was used, and therefore, no additional informed consents were required. A research license was also sought from the National Commission for Science, Technology and Innovation and was granted. The license number of this study was 224145. The Ethics approval and the research license are contained in the appendices section as Appendices II and III, respectively.

CHAPTER FOUR

RESULTS AND INTERPRETATION

4.1 Prevalence of breast cancer

Descriptive statistics for the prevalence of FBC at diagnosis are displayed in **Figure 4.1** below. From the analysis, breast cancer stage III is the most prevalent stage at diagnosis, which accounted for 36% of the total number of cases that were diagnosed over the period. Breast cancer stage II had a prevalence of 34.7%, while stages IV and I had a prevalence of 24% and 5.33%, respectively. This indicates that in Kenya, a majority of the breast cancer cases that are diagnosed are at stage III, which is an advanced stage of the disease.

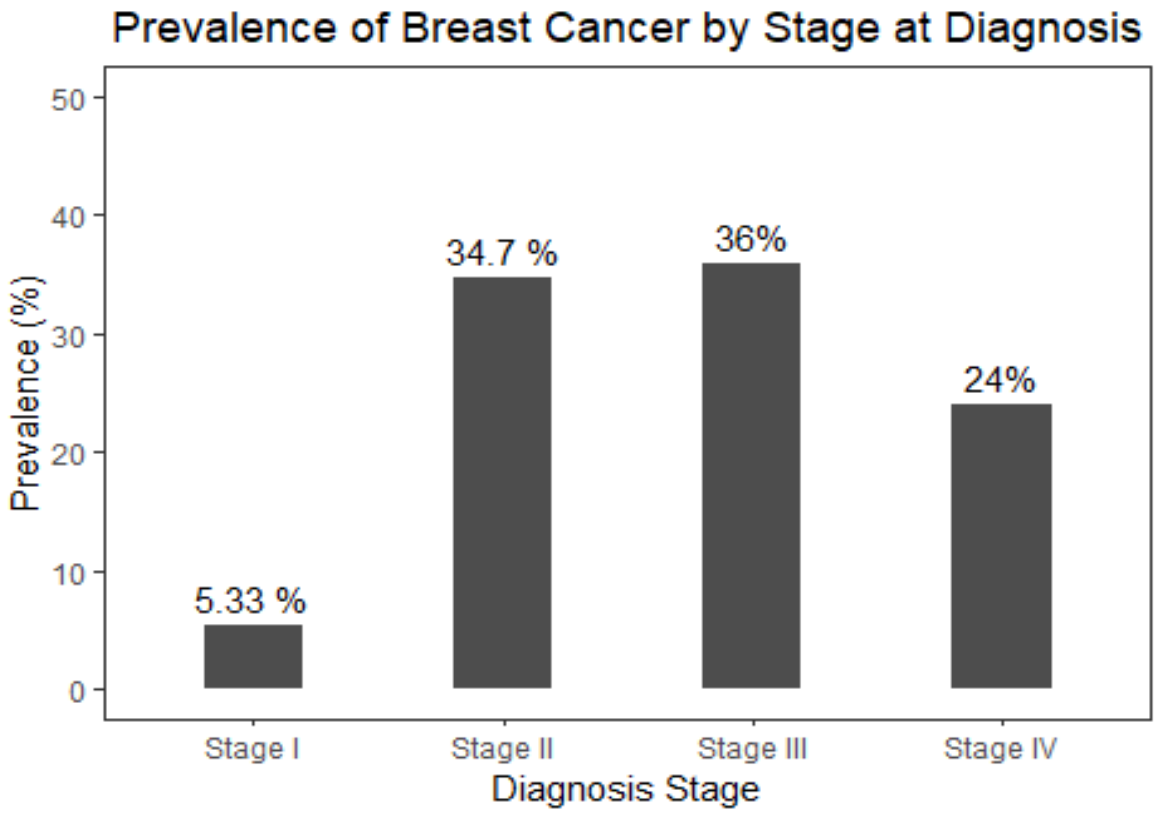


Figure 4.1: Stage Prevalence at Diagnosis

The prevalence of FBC per stage at the time of treatment initiation differs from that of the diagnosis stage. From the analysis, 34.7% of the patients were at stage III of breast cancer, 32.7% at stage II, and 26% and 6.67% of the patients were at stage IV

and I, respectively. It, therefore, indicates that between the time when the patients were diagnosed with breast cancer and the time they sought treatment, there were transitions that occurred. **Figure 4.2** below shows the prevalence of FBC per stage at the treatment initiation stage.

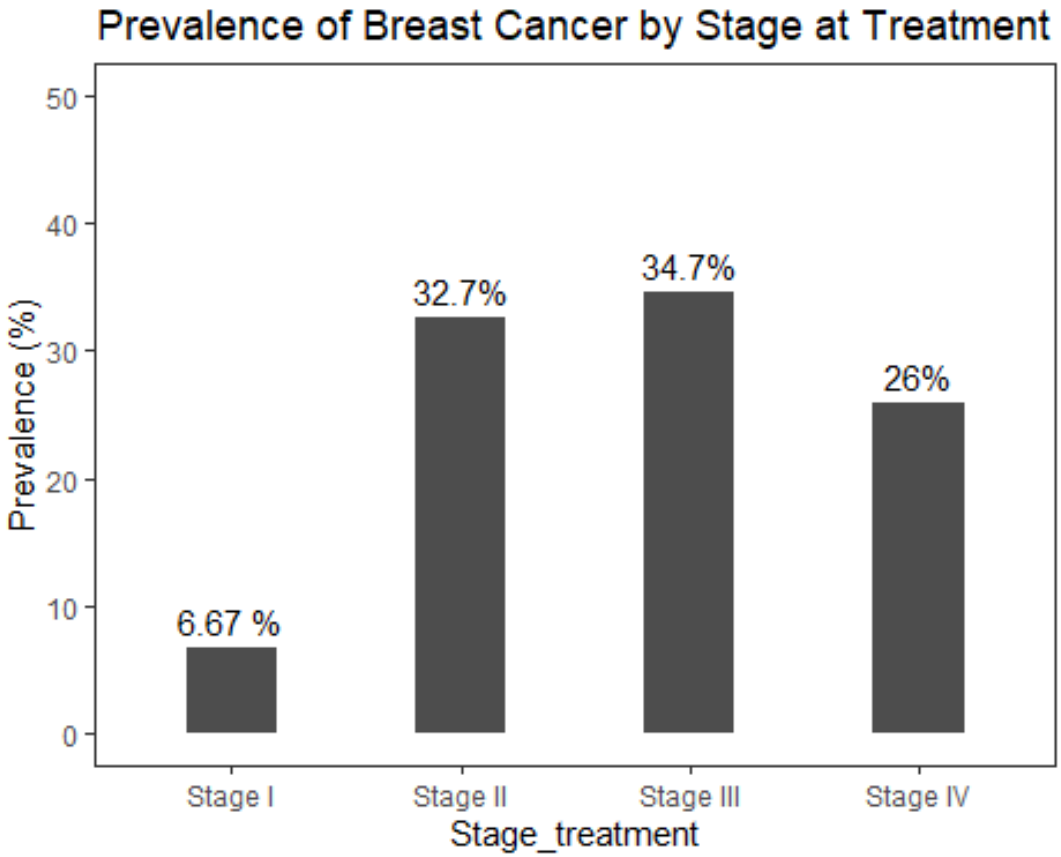


Figure 4.2: Stage Prevalence at Treatment

4.2 Transition Probabilities

To estimate the transition probabilities from one stage of breast cancer to another between diagnosis and treatment initiation, this study defined four stages of breast cancer. When defining the semi-Markov model, the four stages were taken as the possible states where transitions can occur. The basis of the transition probabilities of a continuous process is the Q matrix given by equation (2). From the analysis of the data, Equation (4) yields;

$$N = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 2 & 48 & 0 & 2 \\ 0 & 1 & 52 & 1 \\ 0 & 0 & 0 & 36 \end{pmatrix} \tag{32}$$

The above matrix represents the transition counts of patients from one stage to another between diagnosis and treatment. As seen in equation 26, 8 breast cancer patients diagnosed at stage I remained in the same stage and took a total of 12 months. Out of the 52 patients diagnosed with Stage II, 48 remained in the same stage and took a total of 118 months; 2 transitioned back to Stage I and took 4.5 months, while two patients had progressed to Stage IV and took 4 months before they sought treatment. 52 out of the 54 patients diagnosed at stage III remained at the same stage, taking a total of 126 months; 1 patient who took 2 months transited backwards to stage II, while 1 patient had transited to Stage IV after taking 4 months. All 36 patients diagnosed with breast cancer stage IV remained at the same stage as at the time they sought treatment and took 66.5 months.

The transition time matrix, as outlined in equation (5), is given as follows;

$$\mathrm{T} = \begin{pmatrix} 12 & 0 & 0 & 0 \\ 4.5 & 118 & 0 & 4 \\ 0 & 2 & 126 & 4 \\ 0 & 0 & 0 & 66.5 \end{pmatrix} \tag{33}$$

Time in the above matrix is given in months, showing the total time taken by patients who moved from one state to another. Implementing Equations (8) and (9) yields a calculated transition intensity matrix given by Equation (12), which shows the rate at which patients move from one state to another. The calculated transit intensity is given as;

$$Q = \begin{pmatrix} 0.6667 & 0 & 0 & 0 \\ 0.4444 & 0.4068 & 0 & 0.5 \\ 0 & 0.5 & 0.4127 & 0.25 \\ 0 & 0 & 0 & 0.5414 \end{pmatrix} \tag{34}$$

To satisfy the requirements of the Q matrix as outlined by equations (3) and (11), the above Q matrix needs to be modified. The modified Q matrix is given as follows.

$$Q = \begin{pmatrix} -0.6667 & -0.6667 & 0 & 0 \\ 0.4444 & -0.9444 & 0 & 0.5 \\ 0 & 0.5 & -0.75 & 0.25 \\ 0 & 0 & 0.5414 & -0.5414 \end{pmatrix} \tag{35}$$

The results of the log-likelihood, as outlined by Equation (15), are given as follows

Table 4.1: Maximum Likelihood Estimates

Maximum likelihood estimates		
Transition intensities		
Transition	**Baseline**	
State 2 - State 1	0.016036	(0.0040097, 0.064135)
State 2 - State 2	-0.032334	(-0.0861635, -0.012134)
State 2 - State 4	0.016298	(0.0040755, 0.065172)
State 3 - State 2	0.008038	(0.0011312, 0.057114)
State 3 - State 3	-0.015569	(-0.0622587, -0.003893)
State 3 - State 4	0.007531	(0.0009917, 0.057188)
-2 * log-likelihood: 56.68035		

The estimated intensity matrix obtained from the maximum likelihood estimate is given as follows:

$$\hat{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.0160 & -0.0323 & 0 & 0.0163 \\ 0 & 0.0080 & -0.0156 & 0.0075 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (36)$$

The above equation satisfies the Q matrix requirements as given by Equations (3) and (11). From the above matrix, $\hat{q}_{11}$ is given as 0 since there is no record of breast cancer patients who were diagnosed and transitioned to another stage before seeking treatment. $\hat{q}_{44}$ is also given as 0 since, from the data obtained, it is an absorbing state. Therefore, there are no expected transitions between stage IV and any other stage. As shown in Equation 35, the rate at which patients move from stage II to I of breast cancer is 1.6% per month. The rate of moving out of stage II to any other stage is 3.23%, while patients moving from stage II to IV have a rate of 1.63% per month. The rate at which breast cancer patients move from stage III to stage I and IV is 0.8% and 0.75% respectively, while the rate of moving out of Stage III to any other stage is 1.56%. The estimated matrix is then used to find the matrix exponential, which yields

the transition probability matrices at different times. The time t is given in months, and a transition probability matrix is estimated at different numbers of months.

Equation (22) is used to estimate the transition probability matrices at various times. The transition probability matrix estimates the probability of moving from one stage to another. The matrix exponential is ideal for calculating transition probabilities continuously, assuming any particular distribution for the sojourn times. The method gives a more realistic and flexible representation of the progression of a disease, unlike other methods that assume fixed step intervals. The transition probability matrices below give the transition probability matrix of breast cancer patients in Kenya after the time (*t*) of diagnosis.

After one month ($t = 1$)

$$P(1) = e^{\hat{Q}(1)} \tag{37}$$

The resultant matrix from Equation (31) shows that based on the data obtained from the hospital, patients diagnosed at stage II of breast cancer had a probability of 0.96818 of remaining at the same stage, 0.01578 of moving back to stage I, and 0.01604 of transiting to stage IV after one month of waiting before treatment initiation. Patients diagnosed at stage III of breast cancer had a probability of 0.9846 of remaining at stage III, 0.00006 of moving to stage I, 0.0079 of moving back to stage II, and 0.0075 of transiting to stage IV. The results for different time intervals (*t*) in months are as shown below

$$P(1) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.01578 & 0.96818 & 0 & 0.01604 \\ \textbf{III} & 0.00006 & 0.00785 & 0.98455 & 0.00754 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{38}$$

After two months ($t = 2$)

$$P(2) = e^{\hat{Q}(2)} \tag{39}$$

$$P(2) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03106 & 0.93738 & 0 & 0.03156 \\ \textbf{III} & 0.00025 & 0.01532 & 0.96934 & 0.01508 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (40)$$

Average time spent by patients in months ($t = 2.4667$)

$$P(2.4667) = e^{\hat{Q}(2.4667)} \quad (41)$$

$$P(2.4667) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03475 & 0.92993 & 0 & 0.03156 \\ \textbf{III} & 0.00031 & 0.01711 & 0.96563 & 0.01695 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (42)$$

After three months ($t = 3$)

$$P(3) = e^{\hat{Q}(3)} \quad (43)$$

$$P(3) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.04585 & 0.90755 & 0 & 0.04660 \\ \textbf{III} & 0.00055 & 0.02244 & 0.95437 & 0.02264 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (44)$$

After four months ($t = 4$)

$$P(4) = e^{\hat{Q}(4)} \quad (45)$$

$$P(4) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.06017 & 0.87868 & 0 & 0.06115 \\ \textbf{III} & 0.00097 & 0.02922 & 0.93962 & 0.03019 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (46)$$

After five months ($t = 5$)

$$P(5) = e^{\hat{Q}(5)} \quad (47)$$

$$P(5) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.07403 & 0.85072 & 0 & 0.07524 \\ \textbf{III} & 0.00149 & 0.03566 & 0.92511 & 0.03774 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (48)$$

After six months ($t = 6$)

$$P(6) = e^{\hat{Q}(6)} \tag{49}$$

$$P(6) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.08746 & 0.82365 & 0 & 0.08889 \\ \textbf{III} & 0.00211 & 0.04179 & 0.91082 & 0.04528 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{50}$$

After twelve months ($t = 12$)

$$P(12) = e^{\hat{Q}(12)} \tag{51}$$

$$P(12) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.15949 & 0.67841 & 0 & 0.16210 \\ \textbf{III} & 0.00769 & 0.07248 & 0.82959 & 0.09024 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{52}$$

After thirty-six months ($t = 36$)

$$P(36) = e^{\hat{Q}(36)} \tag{53}$$

$$P(36) = \begin{bmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.34110 & 0.31223 & 0 & 0.34667 \\ \textbf{III} & 0.04835 & 0.12404 & 0.57093 & 0.25668 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{bmatrix} \tag{54}$$

After seventy-two months ($t = 72$)

$$P(72) = e^{\hat{Q}(72)} \tag{55}$$

$$P(72) = \begin{bmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.44760 & 0.09749 & 0 & 0.45491 \\ \textbf{III} & 0.11826 & 0.10954 & 0.32596 & 0.44623 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{bmatrix} \tag{56}$$

As outlined, P_{11} and P_{44} have a probability of 1 in all the matrices since they are absorbing states. Patients who were diagnosed at these stages never transitioned to any other stage before treatment. Patients diagnosed at stage I remained in the particular stage indefinitely since there were no observed transitions to other stages.

This could indicate cases of early detection of breast cancer in patients. The fact that stage IV is an absorbing state suggests that once patients reach stage IV, they do not transit back to earlier stages. Stage IV of breast cancer represents an advanced stage of cancer where metastasis has taken place and, therefore, becomes unlikely to transition to earlier stages.

The specific entries with non-zero probabilities in the transition probability matrix can be represented as shown in **Table 4.** 2. The probability of moving from stage II to stage I increases from 0.0158 at t=1 to 0.4476 at t=72, where t is the waiting time between diagnosis and treatment initiation. This means that after one month (t=1), there is a 1.5% chance of moving from stage II to stage I, and at t=72, the chances rise to 44.76%. The chances of remaining in stage II without seeking treatment decrease from 96.82% to 9.75%, with an increase in time in months from $t = 1$ to $t = 72$, respectively. The chances of moving from stage II to stage IV increase significantly from 1.6% at $t = 1$ to 45.49% at $t = 72$. There is a higher probability of remaining in stage III (0.985) at $t = 1$ compared to a probability of 0.326 at $t = 72$. However, as time (t) increases, the probability of progressing to stage IV increases significantly from 0.0075 at $t = 1$ to 0.4462 at $t = 72$. It is worth noting that a patient could have been diagnosed at a particular stage of breast cancer and still be in the same stage of cancer at treatment, but with more grown tumours than during diagnosis. This study considers distinct stages and not tumour-specific stages of breast cancer, and this is the reason a majority of the patients were observed to have remained in the same stage between diagnosis and treatment. However, this does not rule out the likelihood of patients diagnosed with breast cancer stage, say III, progressing from IIIa to IIIb.

Table 4. 2: Transition Probability Table

Time (Months)	P11	P21	P22	P24	P31	P32	P33	P34	P44
1	1	0.01577952	0.96818315	0.01603733	6.342914e-05	0.007847857	0.9845516	0.007537143	1
2	1	0.03105698	0.93737862	0.03156440	2.497138e-04	0.015324783	0.9693418	0.015083709	1
3	1	0.04584836	0.90755419	0.04659745	5.530160e-04	0.022444453	0.9543670	0.022635545	1
4	1	0.06016913	0.87867868	0.06115219	9.677133e-04	0.029220077	0.9396235	0.030188695	1
5	1	0.07403426	0.85072190	0.07524385	1.488392e-03	0.035664417	0.9251078	0.037739384	1
6	1	0.08745824	0.82365461	0.08888715	2.109838e-03	0.041789802	0.9108163	0.045284017	1
12	1	0.15949362	0.67840692	0.16209946	7.686375e-03	0.072483197	0.8295864	0.090244015	1
36	1	0.34109989	0.31222725	0.34667286	4.834668e-02	0.124036633	0.5709327	0.256684021	1
72	1	0.44760057	0.09748585	0.45491358	1.182583e-01	0.109544182	0.3259641	0.446233447	1

The transition probabilities given in the table are represented graphically, as shown in

Figure 4. 3. Each line represents the probability of transitioning between a given set of breast cancer stages as a function of time. The absorbing states (Stages I and IV) have their lines converge at a probability of 1. This shows that patients remained in the states for an indefinite time. The transient states (Stages II and III) have their probabilities fluctuating before they stabilize. From the graph, taking a case of forward progressions, the line showing the transition from stage II to IV is the steepest between 0 and 36 months of waiting for treatment.

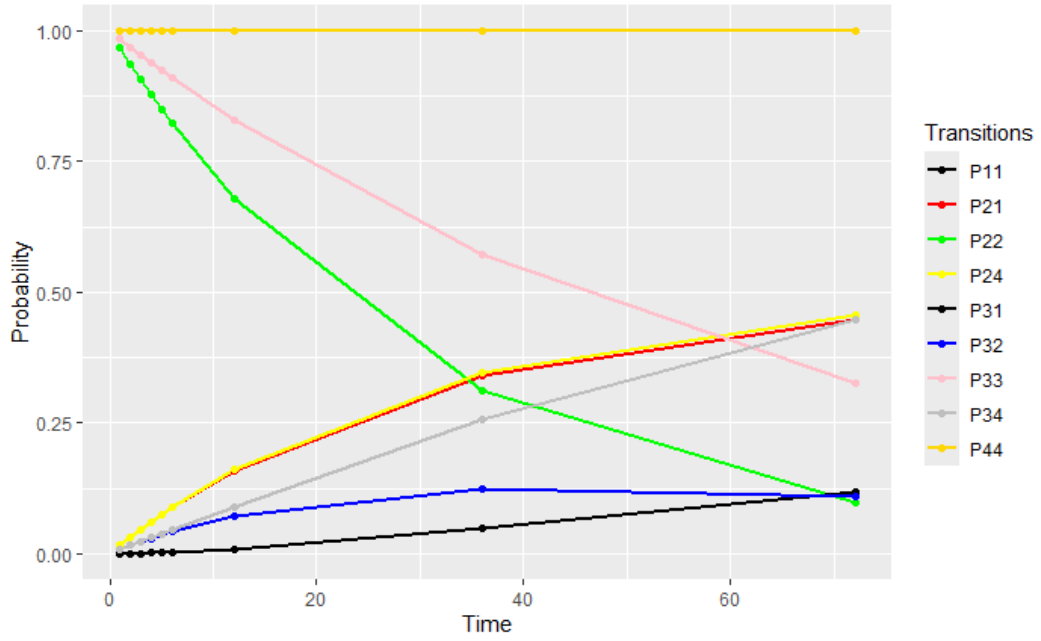


Figure 4. 3: Transition probability graph

If the time the patients take before they seek treatment after diagnosis is extended to 240 months to observe the long-term likely behaviour of the transition probabilities, we obtain **Figure 4**. Some of the key highlights from the graph are that the chances of remaining at stages of diagnosis (specifically stages II and III) decrease significantly to 0. This means the probability of remaining at stage II and III approaches 0 with time, as derived from the analysis of the observed data of breast cancer patients. With time, the probability of transitioning backwards from stage II to me and from stage III to stage II also approaches zero based on the extrapolation results, as shown in **Figure 4**. As the time taken increases before seeking treatment, the chance of progressing from

stage III to IV increases significantly. This is by the line showing the trend of the probability of moving from stage III to IV in **Figure 4**.

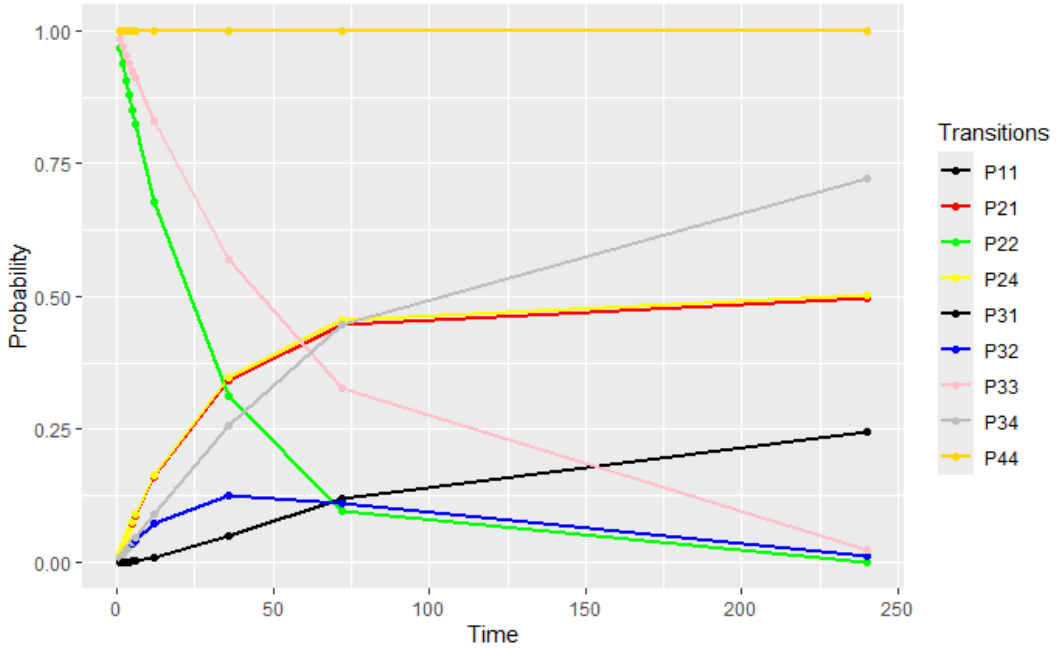


Figure 4.4: Extrapolated Transition Probability Graph

4.2.1 Transition Probability Diagram

If we take the transition probability matrix to be given by the average number of months the patients took before they sought treatment (t=2.47), then our transition probability matrix is given as

$$P = \begin{bmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03475 & 0.92993 & 0 & 0.03156 \\ \textbf{III} & 0.00031 & 0.01711 & 0.96563 & 0.01695 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{bmatrix} \quad (57)$$

A transition probability diagram can be drawn to visually represent how breast cancer patients move from one stage of the disease to the other. The probability diagram is a critical tool that helps understand disease progression and how two or more different states communicate. The respective probability diagram is given below based on the transition probability matrix cap P. Stages I and IV are given as absorbing states from the probability diagram since there is no record of transitions to other states. Stages II and III are transient states since the probability of returning to the states is non-zero. As seen in **Figure 4.5**, the data obtained from the registry had a record of many of the breast cancer patients diagnosed at stages II and III experiencing the transitions. Based on the average waiting time between diagnosis and treatment (2.47 months), there are higher chances, 93%, of remaining at stage II as compared to 3.5% and 3.3 % of transitioning to stages I and IV, respectively. The chances of being diagnosed and remaining at stage III are higher than stage II, which is 96.6%. The chances of moving from stage III to stages II and IV are equal and stand at 1.7% while moving to stage I is very unlikely and stands at 0.03%.

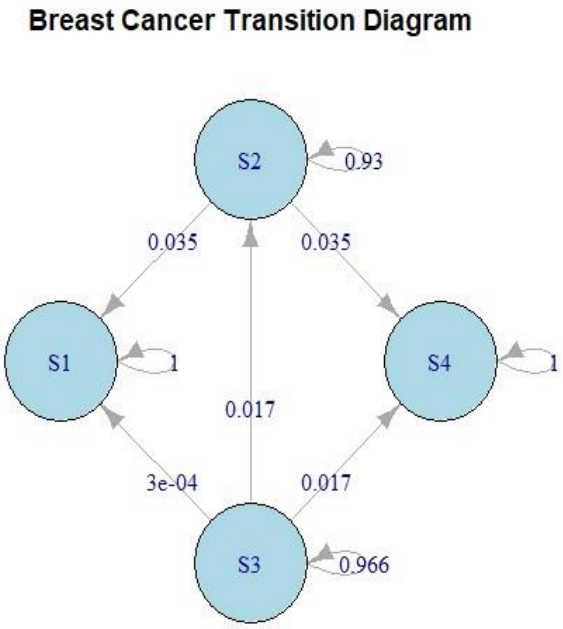


Figure 4.5: Breast Cancer Transition Probability Diagram

4.3 Steady State Distribution

Using the Markov chain package in R, the steady state probabilities of the Semi-Markov model were computed by solving Equation 28. The results of the analysis are given as

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad (58)$$

The output above reflects two distinct distributions of steady states, which, from the data obtained from the two cancer centers, are dependent on the patient's initial stage at diagnosis. The results of the analysis have two steady-state vectors since, from earlier analysis, stage I and stage IV were identified to be absorbing states. Based on the data obtained from the cancer centers, for patients diagnosed at stage II and stage III, the steady-state vector is given as

$$[0 \quad 0 \quad 0 \quad 1] \qquad (59)$$

The above steady state vector suggests that, from the data obtained, patients diagnosed at stage II and III eventually transition to stage IV with a probability of 1 before treatment initiation. This finding suggests that the long-run distribution of cancer patients who enter the cancer pathway at stage II or III will eventually be in stage IV without any effective treatment initiation. Stage four for this study is an absorbing state, and patients will transition along the pathway from stage II to stage IV. This result indicates that breast cancer progression tends to worsen as more time is taken without early intervention.

The steady state distribution of the patients diagnosed at stage I is given by the second steady state vector in the results of the analysis. For patients diagnosed at Stage I, the steady-state vector is given as

$$[1 \quad 0 \quad 0 \quad 0] \qquad (60)$$

This steady state vector suggests that, based on the data that was obtained for this study, patients who were diagnosed at stage I remained in the same stage indefinitely. These patients did experience any transitions prior to treatment, which indicated that before treatment initiation, stage I can act as a stable terminal state for some patients diagnosed at this stage. It is important to emphasize that this result does not suggest/imply any cure or biological remission, but an observed stagnation in the stage

prior to treatment. It is likely that the patients were diagnosed early enough in such a way that there were no observed transitions occurred within the time period studied.

4.4 Average Cost of Management of Breast Cancer

Error! Reference source not found. below presents the cost components which are a ssociated with the treatment and management of breast cancer at different stages, specifically stage II, III and IV, based on equation 30 and the data obtained. The data contains the costs of carrying out diagnostic procedures, imaging, treatment, supportive therapies, and follow-up, among other management costs. For each item, the mean cost was obtained and finally, the total cost per stage. The summary of the average costs is as shown in Error! Reference source not found..

Table 4.3: Average Costs of Breast Cancer Management per Stage

Cost Item	**Stage II – Average (Ksh.)**	**Stage III – Average (Ksh.)**	**Stage IV – Average (Ksh.)**
Mammogram	5500	0	0
Biopsy	6000	6000	6000
Staging	14000	14000	26500
ECHO	5000	5000	0
Chemotherapy	240000	105000	90000
Restaging	14000	14000	7000
Surgery	120000	110000	0
Radiotherapy	105000	135000	50000
Oral chemotherapy	45000	31500	0
CT scan brain	0	3500	0
Radiotherapy (brain)	0	30000	0
2nd Line Chemotherapy	0	45000	0
PET scan	0	30000	60000
Targeted therapy	0	0	597000
2nd line targeted therapy	0	0	300000
Hormonal therapy	240000	240000	240000
Laboratory workup	100000	100000	100000
Clinical follow-ups	50000	50000	50000
Total	**944,500**	**919,000**	**1,526,500**

4.4.1Treatment Type Distribution after Treatment Initiation

Figure 4.6 shows the proportion of patients who received treatment type 1, which, for purposes of this study, was taken to be surgery. 66.7% of the respondents indicated that when treatment was initiated, surgical treatment was one of the treatment interventions that they underwent. A third of the patients did not undergo this treatment type due to various reasons. The patients may have been inoperable dues to, maybe, comorbid reasons or may have declined this treatment pathway due to personal reasons. As seen in Error! Reference source not found., surgery contributed s ignificantly to the average treatment costs per stage and therefore influenced the clinical trajectory in disease management.

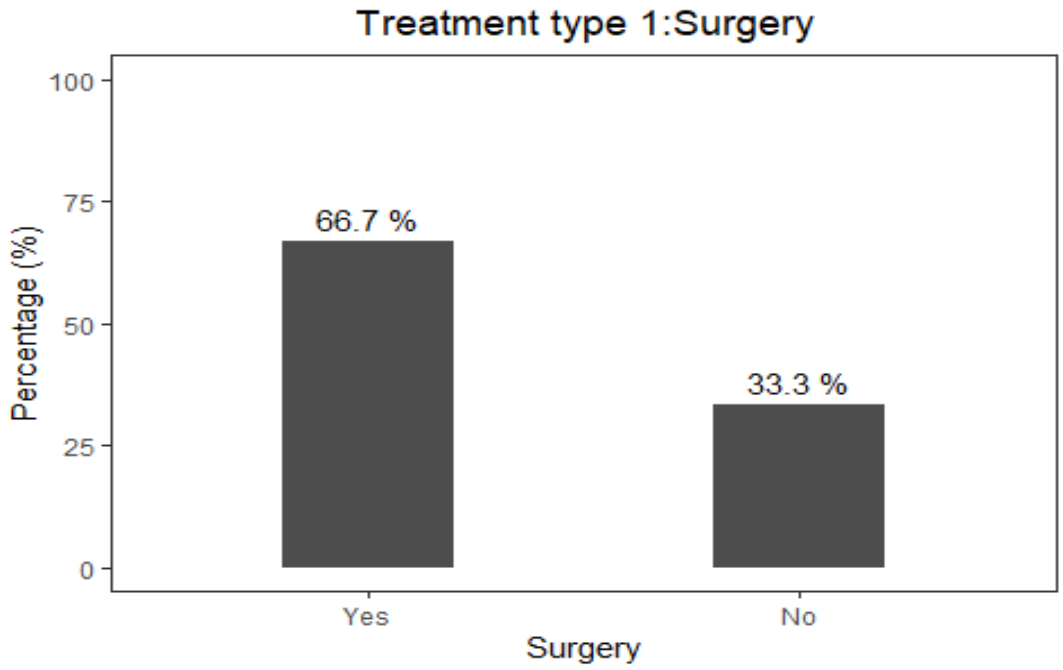


Figure 4.6: Treatment Type I (Surgery)

Chemotherapy was the second treatment type that was considered in this study, and a significant proportion of the patients (87.3%) of the patients whose records were used in this study underwent this treatment method. Only 12.7% of the patients did not undergo any form of chemotherapy, which suggests that it's a critical treatment modality in breast cancer treatment and management. **Figure 4.7** shows a visual

representation of the distribution of patients who underwent chemotherapy and those who did not.

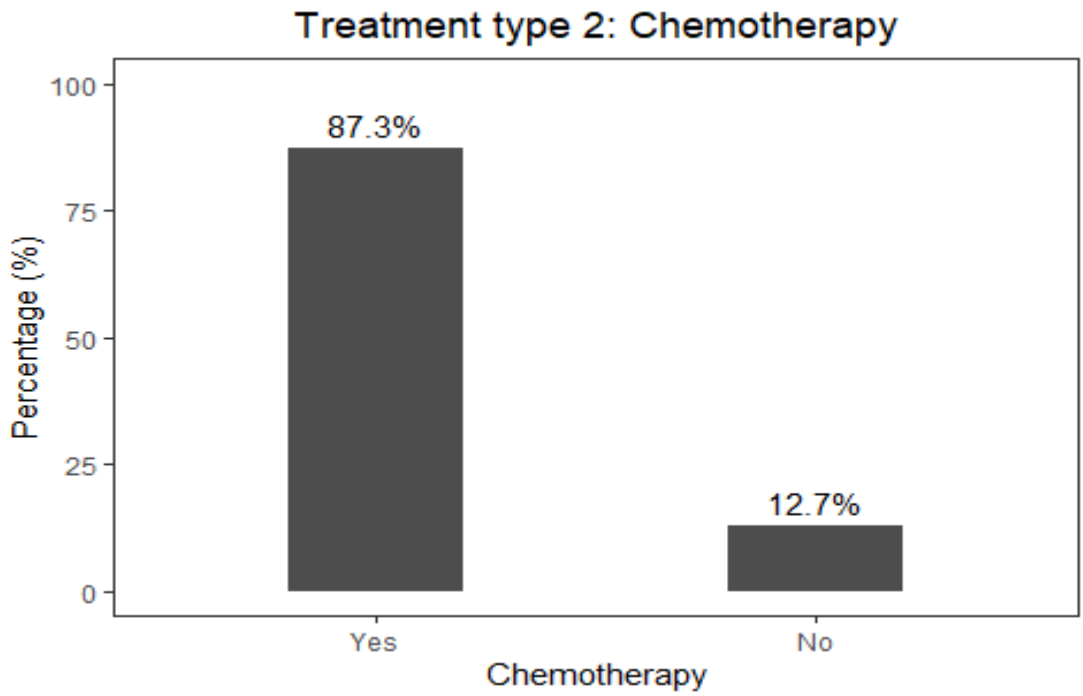


Figure 4.7: Treatment Type II (Chemotherapy)

The third treatment type considered was hormonal therapy, and the distribution of those patients who underwent this treatment type and those who did not is shown graphically in **Figure 4.8**. Only 58.7% of the total patients underwent hormonal therapy as a breast cancer treatment and management modality, while only 41.3% of the patients did not.

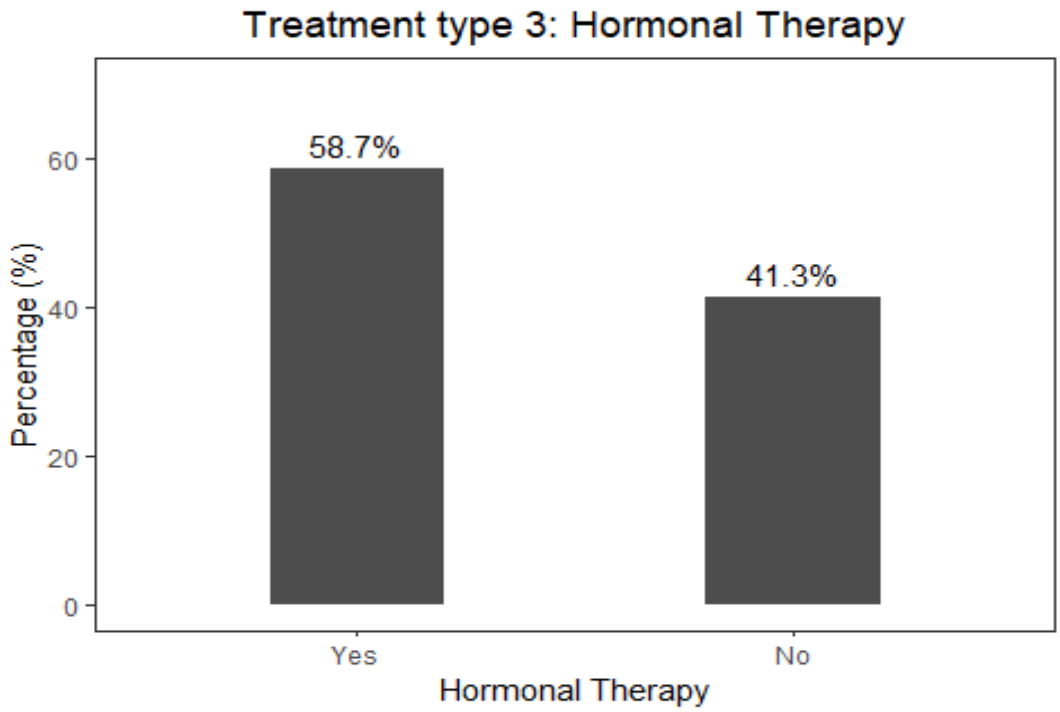


Figure 4.8:Treatment Type III (Hormonal Therapy)

68.7% of the patients indicated that they did not undergo any other treatment type other than a surgical procedure, chemotherapy or hormonal therapy, as shown in **Figure 4.9**. However, 31.3% of the patients indicated that they underwent other treatment modalities, which include radiotherapy among others.

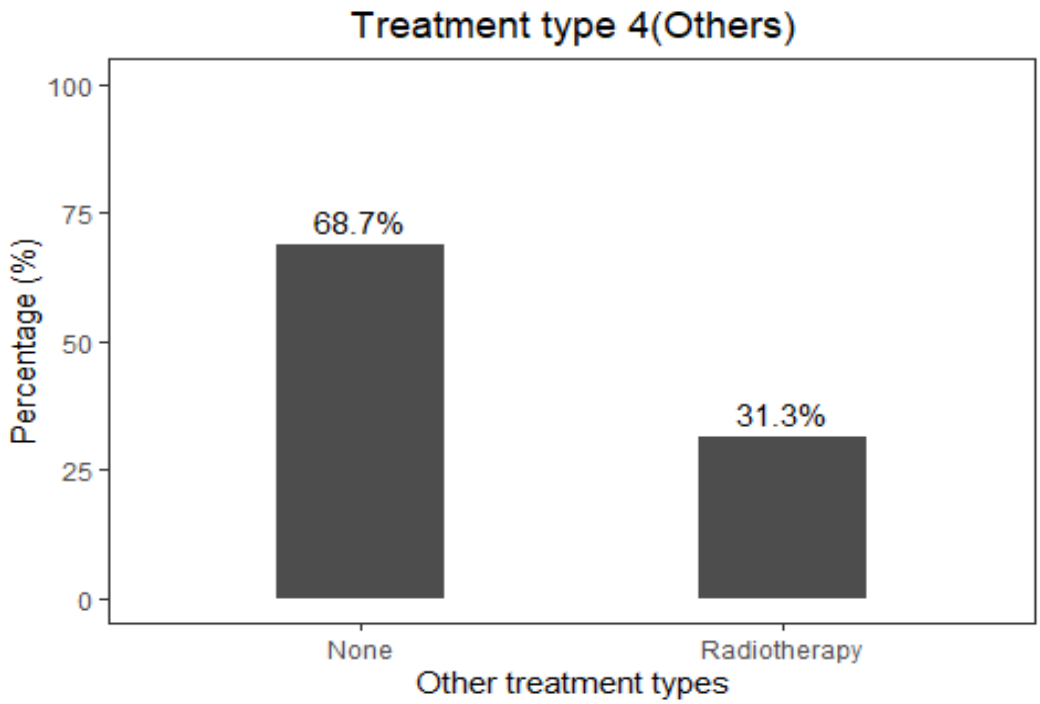


Figure 4.9: Other Treatment Types (Radiotherapy)

At the time of the data extraction from the patient files, 93.3% of the patients were still alive, while 6.67% were reported to have died. **Figure 4.10** graphically shows the proportions of the patient status at the time of data extraction.

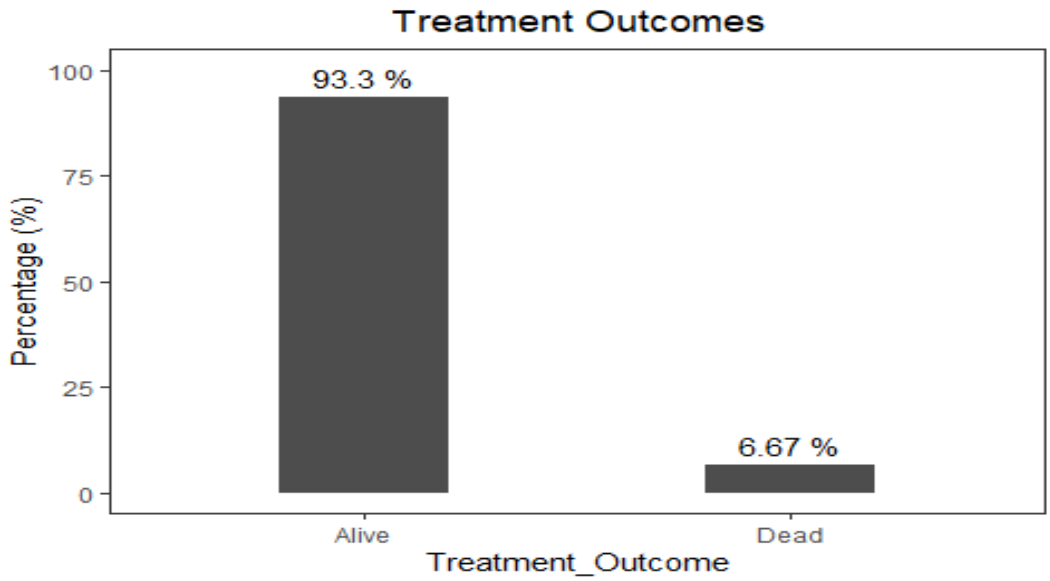


Figure 4.10: Patient Status

4.5 Application of Steady-State Probabilities in Estimating Costs

In order to calculate the long-term expected costs per patient, the semi-Markov model's steady-state probabilities were used. However, based on the steady state probability distributions obtained, patients diagnosed at stage II and III eventually will be at stage IV. This, therefore, means the long economic burden will be at stage IV. **Table 4.4** shows the estimated total cost of managing breast cancer per stage in 5 years.

Table 4.4: Estimated Long-Term Breast Cancer Management Cost

Breast Cancer Stage	Average Cost (KES)
Stage I	302,240
Stage II	944,500
Stage III	919,000
Stage IV	1,526,500

As would be expected, the average cost of the treatments rose with the degree of cancer. The cost of managing breast cancer Stage IV was found to be over four times more than that of managing Stage I in five years. **Table 4.4** above shows a summary of the average cost of managing breast cancer per stage for a period of five years. It is good to note that the consultation charges were not included in the treatment costs.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Discussion

The first objective of this study was to determine the distribution of cancer patients per stage at the time of diagnosis and treatment initiation in Kenya. The findings indicated that the majority of patients sought a diagnosis at stages II and III, whereas less of them sought the diagnosis at stages I and IV. Such staging tendency is an indication of a general trend in low-and medium-income countries (LMICs) where early-detection promotion activities are not yet developed (Chantada *et al*., 2025). Recent studies underscore similar findings across sub-Saharan Africa. A study by Ngwa *et al.* (2022) revealed that most of the sub Saharan countries reported that over 70 percent of breast cancer cases were identified at either stage II or stage III with stage III being the most common. They researchers had similar results to the ones found in this study and also help in understanding staging of late-stage diagnosis, which continues to thrive in the area. According to the study conducted by Ngwa *et al*. (2022), the late-stage diagnosis is commonly known to be caused by low awareness, cultural barriers, and delayed referrals, as well as a lack of resources to make the diagnosis.

Further support of the findings of this study is found in a study conducted by Daniel *et al.* (2023) in Kenya, that revealed that delayed diagnosis and the start of treatment contributed to progression of breast cancer. To their finding, there was time elapsed between the development of the symptoms and hospital presentation that the patients sought to know their stage. In contrast, weaker countries record higher rates of early-stage detection whereby they have developed better screening infrastructure. (Ginsburg *et al.*, 2020). The American Cancer Society (2023) notes that in the United States of America, nearly 64% of breast cancer cases are diagnosed in either stage I or stage II due to organized referral systems and routine mammography (Islami *et al*., 2024). The finding that fewer patients were diagnosed early (stage I) is consistent with the challenges in the Kenyan health context.

The second objective of this study was to construct a transition probability matrix to capture the likelihood of moving between breast cancer stages before treatment initiation. The findings showed that transitions predominantly occurred between stages II and III, with relatively rare backward transitions. This finding is consistent with a

study by Velloso *et al*. (2017), which evaluated breast cancer progression and found that forward transitions were dominant, particularly between stages II, III, and IV, with minimal backward transitions. In this study, stages I and IV were identified as absorbing states, which indicate that diagnosed patients at stages I and IV remained at the same stage before treatment was initiated.

The results from the observed data show that the chances of moving from stage III to stage II or stage I are so small, as seen in **Table 4.** 2. These results are similar to the results by Shockney (2025), who found that stage III breast cancer is locally advanced and has tumours that are larger than those of stages I and II. The transition probabilities in this study were consistent with findings from other studies, like that of Huang *et al.* (2020), who estimated transition probabilities from breast cancer stage 0 to IV. The researchers used a full Markov model to track preclinical breast cancer development across the stages. However, the transition probabilities in their findings were age-specific and not time-specific, as in this study.

In their study, Li *et al*. (2022) gave the survival contradiction of the different cancer stages, where stages II and III are further divided into more tumour-specific categories. Another study found that luminal B breast cancers (locally advanced-stage cancers) had greater growth rates of tumours than luminal A breast cancers (Alaidy *et al*., 2021). This could be the reason why, in our study, the probability of stage III breast cancer patients transitioning backwards is significantly low. Even at $t = 6$, the probability of remaining at stage III (0.9108) is higher than that of stage II (88.89).

The steady-state distributions obtained in this study are crucial since they reflect the terminal behaviour of breast cancer in the population under study. The finding that, in the long run, patients diagnosed in stages II and III will be in stage IV is consistent with the nature of breast cancer as depicted by different studies. For instance, Li *et al*. (2019) in their study found out that delays in the treatment initiation of breast cancer are a major medical problem in LMICs. The researchers further noted that delayed treatment initiation was associated with worse survival rates since patients had already progressed to advanced stages of breast cancer. The finding is also consistent with that of a study by Flores-Balcázar *et al.* (2020), which found that delays in treatment initiation in the management of cancer reduced cancer-specific survival outcomes in women who were diagnosed with breast cancer. The direct indicator of the possibility of the delay in treatment initiation burdening healthcare systems was the evidence that

those patients with stage II and stage III without any treatment inevitably transferred to stage IV. Similar results were reported by Mutebi *et al*. (2020) in their study.

The results of the stage IV were that there was a significant rise in the costs of treatment with the stage of breast cancer. The mean total cost of management of breast cancer was the greatest in stage III & IV. The findings are congruent with results of other researchers in the study of the treatment and management of breast cancer costs. Sun *et al.* (2018) studied the cost of breast cancer treatment and discovered that cost of treatment was higher when cancer was diagnosed in its later stages than when it was identified in its early stage. A study was conducted by Thomas *et al.* (2021) to investigate the various cost models within breast cancer episodes by stages. The oncological costs were mostly spent on the difficulties of the treatment of breast cancer cases at stage IV because of the complexity and the needs of the cures. In particular, the significance of this study was that the per-unit cost of stage IV breast cancer was found by the researchers at 35 percent over the average target cost of managing the disease (Thomas *et al.*, 2021).

The cost estimation of the study fits in the theoretical theories and models of health economics predicting non-linear costs rise whereby the costs rise with increase in the disease progression. This observation concurs with the result of a different study whereby, it was observed that the longer the treatment initiation delay time, the more the cost of treatment because of the increased tumour burden as well as other complications (Hanna *et al.*, 2020). This owes to the finding of this study where the costs associated with delays in diagnosis are higher and associated with poor treatment outcomes (Sun *et al.*, 2018). The lack of cost data at stage I indicates that there has been a difficulty in overall cost monitoring of outpatients or less-intensive care.

5.2 Conclusions

This study employed semi-Markov models to analyze breast cancer progression across its four possible stages. The model incorporated the use of matrices based on transition probabilities and time of waiting between diagnosis of disease and the treatment with the aim of appreciating the dynamics of the progression. Among the key conclusions of this research, it can be stated that stage I and stage IV turned out to be absorbing states according to the semi-Markovian representation of the data of breast cancer patients provided by the registry. Stage I, being an absorbing state, suggests that the

cases may have been detected early and took time before transitioning to another stage. It could have been led by the timely intervention and the efforts that have been put forward to encourage people to go for screening for breast cancer.

In this study, stage IV was found to be an absorbing state, and this phenomenon could suggest that the patients had reached an advanced stage where it is unlikely to transition backwards to earlier stages. It is worth noting that this study only considered the four possible stages (I, II, III, and IV) and therefore the transition probabilities were between the four stages and not any other stage. There could be patients who might have died after stage IV or any of the other three stages, but this study does not consider the death stage or the probability of getting to the death stage.

This aligns well with the existing literature that suggests that in advanced stages like stage IV, metastasis has already taken place. It is not possible to reverse the progression. The findings of this study confirm that stage III is highly progressive based on the results of the extrapolation of the time taken before getting treatment. Stage III patients were seen to have lower chances of regressing backwards from stage III. The probability of remaining in stage II or III after diagnosis decreases with increased patient time before treatment initiation. This suggests that time is a crucial factor in the dynamics of breast cancer progression (Kisiangani *et al.*, 2018). As the time taken increases, the probability of remaining at the stage at diagnosis reduces significantly, especially in stages II and three, which could be attributed to various factors.

The steady-state findings have significant implications for clinical practice and public health policy. Stage II and III patients should be prioritized for timely treatment, given their inevitable progression to stage IV without any intervention (Rivera-Franco & Leon-Rodriguez, 2018). These stages represent critical chances to take necessary action to prevent worsening outcomes by progressing to advanced stages like stage IV. Although the data used in this study showed records of patients before treatment, stage I provides a strategic opportunity for early breast cancer intervention. The fact that no transition occurred suggests suggest early-stage diagnosis of the disease may be stable over short windows, which aligns with the need to have early diagnosis and regular screening. For this study, stage IV remains an absorbing state and indicating that once breast cancer reaches this advanced stage, the natural trajectory is terminal, meaning there is no possibility of regression, specifically in the absence of treatment.

The use of a semi-Markov model has been proven, in this study, to be effective in capturing the progression of breast cancer in a continuous time process, which offers a more flexible approach than other models. The transition probabilities are time-specific, which makes it more possible to understand the effect of delayed treatment initiation on breast cancer patients. The application of matrix exponentiation in this study allowed the model to have time-dependent transition probabilities and patterns. The semi-Markov framework offers a powerful approach to understanding the long-term consequences of early diagnosis and treatment delays in female breast cancer management.

The steady state analysis of the pre-treatment of the progression of breast cancer reveals that when the necessary interventions are not taken after diagnosis, the disease evolves towards two terminal outcomes. One of the outcomes is the stability evident in stage I for patients who are diagnosed early, who are more likely to remain in that stage indefinitely. The second outcome is the inevitable progression towards stage IV for the patients diagnosed at the intermediate stages. This finding emphasizes the importance of early detection and the need to have regular screening so that female breast cancer can be diagnosed early enough to enable fast-tracked treatment initiation. This applies particularly to patients who are diagnosed beyond stage I, who have a chance of progressing to terminal stage faster if treatment is not initiated in time.

The last objective of this study was to estimate the cost of the management of breast cancer in each of four clinical stages and apply steady-state analysis of the semi-Markov model to evaluate the long-term budgetary consequences of the untreated disease. Economic burden of the management of breast cancer at different stages in Kenya was well understood in practicality and the context of the study, with 100 percent cost data obtained based on real patient records at two public cancer centers in Kenya. Taken together with the properties of transition dynamics expressed in steady-state probabilities, the study provides a strong framework of how breast cancer changes over the course of time without treatment, and how such evolution impacts the economic well-being of the patients and their households.

The results highlight the difference in costs of breast cancer management at various levels of the disease. In particular, the case of Stage IV treatment costs over four times more than that of managing breast cancer at Stage I. The cause of such an increase could be attributed to resource-intensive and complex interventions that are necessary

in advanced stages of the disease, including target-based treatment, prolonged chemotherapy cycles, palliative radiotherapy, and hospital-based follow-ups, as explained by Kailemia *et al.* (2023).

The study findings are in line with international and local literature, which demonstrates that the care of late-stage cancer is tremendously expensive in comparison with the early-stage treatment (Sun *et al*., 2018). Further, after the steady-state analysis, it was observed that before effective treatment, there is at all likelihood that without due treatment in good time, patients who are diagnosed with Stage II or Stage III will definitely turn to Stage IV. Such a development not only aggravates clinical outcomes but also creates a much heavier economic load not only on the health system but also on individual households. The model therefore has strengthened the fact that early-stage intervention is not only a clinical necessity it is also a major economic mechanism through sustainable management of breast cancer.

The findings of this study will only be actioned upon with a larger investment in early detection and the timely treatment thereof when it comes to an understanding of it as far as public health is concerned. The study findings suggest that the earlier the diagnosis, the simpler the treatment and management of the disease can be. This way, it is easier to manage and contain breast cancer and avoid its progression to advanced stages that are more expensive to manage (WHO, 2023). Moreover, savings due to a lack of late-stage care should be reinvested into reinforcing screening systems, training personnel, and supporting regional cancer centers in their efforts to establish the beneficial cycle of early diagnosis and resource utilization.

The analysis reveals that the stage when breast cancer is diagnosed is key in influencing the total cost of care that a patient may require. Aggregation of cost estimates and the dynamics of transitions used in the study shows that breast cancer is not only too costly but can be prevented at an early stage with huge economic gains. These observations advocate a change in policies to proactive screening and quick treatment onset of cancer as a medically and economically relevant need in Kenya's healthcare.

5.3 Recommendations

Based on the findings of this study, a number of the most important recommendations are drawn for consideration in the field of healthcare policy, clinical practice, and future research. The first recommendation is that there is a need to ensure early detection and diagnosis of breast cancer in Kenya. This may be done through the enhancement of community-based awareness and screening activities such as clinical breast examination, health education, and mobile diagnostics. Timely diagnosis is not only more successful but also makes the treatment process much easier and cheaper. Early case identification would be further enhanced by integrating the breast cancer risk assessment into regular practice in the primary health care facilities.

Early detection of breast cancer ought to be advanced with some more investment towards easing its treatment by making it affordable. Since management at the initial stage was seen to be less expensive in this study, the Government should place more emphasis on Stage I and II of the treatment procedure through reimbursement of costs incurred in surgery, hormonal therapy, and other forms of diagnostic imaging procedures. Also, the regular and steady supply of key oncology medications and treatment interventions in the county-level institutions may minimize the number of referrals and terminal cases.

Second, one should deal with the delays that tend to exist between the initial identification and the beginning of the course of action. These delays lead to the development of illnesses, and patients end up taking up more time-consuming treatment. Improving referral systems between the lower health facilities and regions cancer centers, as well as the turnaround time of diagnosis of biopsy, and scans would help minimize these delays. It is also important to ensure that after diagnosis, patients are quickly connected with oncology services.

The stakeholders in FBC control need to ensure they come up with ways of reducing the delays in treatment initiation based on the observation that the probability of progressing to stage IV increases with an increase in waiting time. It may involve reducing logistical and access barriers, improving referral pathways, and improving access to treatment.

5.4 Recommendation for Further Studies

Additionally, this study recommends further research to unearth factors that influence delayed treatment initiation by patients who have received a definitive diagnosis of breast cancer. Further research can be done to determine factors that lead to the downstaging of FBC before treatment is initiated. Finally, although the study was conducted using the data of two cancer centers, additional analysis can be performed on more significant data (multi-centers). Subsequent studies ought to look into the difference in price of treatment in various regions and facilities, evaluate expenses that households spend out-of-pocket, and test the long-run financial performance in various groups of patients. These studies would be beneficial in making results more generalizable and more robust.

REFERENCES

Ahmad, H., Van Der Mei, I., Taylor, B. V, Lucas, R. M., Ponsonby, A.-L., Lechner-Scott, J., Dear, K., Valery, P., Clarke, P. M., & Simpson, S. (2019). Estimation of annual probabilities of changing disability levels in Australians with relapsing-remitting multiple sclerosis. *Multiple Sclerosis Journal*, *25*(13), 1800–1808.

Alaidy, Z., Mohamed, A., & Euhus, D. (2021). Breast cancer progression when definitive surgery is delayed. *The Breast Journal*, *27*(4), 307–313. https://doi.org/10.1111/tbj.14177

Asanjarani, A., Liquet, B., & Nazarathy, Y. (2022). Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches. *The International Journal of Biostatistics*, *18*(1), 243–262.

Baechle, C., Hoyer, A., Stahl-Pehe, A., Castillo, K., Toennies, T., Lindner, L. M. E., Reinauer, C., Holl, R. W., Kuss, O., & Rosenbauer, J. (2019). Course of disordered eating behavior in young people with early-onset type I diabetes: prevalence, symptoms, and transition probabilities. *Journal of Adolescent Health*, *65*(5), 681–689.

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

CDC. (2023). No Title. *Center for Disease Control and Prevention*. https://www.cdc.gov/nchs/hus/sources-definitions/prevalence.htm#print

Chantada, G. L., Villanueva, G., & Abramson, D. H. (2025). Challenges for Early Diagnosis in Retinoblastoma in Low-and Middle-Income Countries. *Pediatric Blood & Cancer*, e31859.

Dalabanjan, M. S., Agrawal, P., T, D., & Suranagi, M. D. (2021). Prognosis of Cancer - A Semi Markov Process. *International Journal of Engineering and Advanced Technology*, *10*(5), 146–150. https://doi.org/10.35940/ijeat.E2695.0610521

Daniel, O., Ashrafi, A., Muthoni, M. A., Njoki, N., Eric, H., Marilynn, O., Faith, A. B., Beth, W. G., Nyakio, M., & Odero-Marah, V. (2023). Delayed breast cancer presentation, diagnosis, and treatment in Kenya. *Breast Cancer Research and Treatment*, *202*(3), 515–527.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, *149*(4), 778–789. https://doi.org/10.1002/ijc.33588

Flores-Balcázar, C. H., Flores-Luna, M. L., Villarreal-Garza, C. M., & Bargalló-Rocha, J. E. (2020). Provider delay in treatment initiation and its influence on survival outcomes in women with operable breast cancer. *Reports of Practical*

Oncology and Radiotherapy, *25*(2), 271–275.

Ginsburg, O., Yip, C., Brooks, A., Cabanes, A., Caleffi, M., Dunstan Yataco, J. A., Gyawali, B., McCormack, V., McLaughlin de Anderson, M., & Mehrotra, R. (2020). Breast cancer early detection: A phased approach to implementation. *Cancer*, *126*, 2379–2393.

Hanna, T. P., King, W. D., Thibodeau, S., Jalink, M., Paulin, G. A., Harvey-Jones, E., O'Sullivan, D. E., Booth, C. M., Sullivan, R., & Aggarwal, A. (2020). Mortality due to cancer treatment delay: systematic review and meta-analysis. *Bmj*, *371*.

Hong, S., Won, Y.-J., Park, Y. R., Jung, K.-W., Kong, H.-J., & Lee, E. S. (2020). Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2017. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *52*(2), 335–350.

Huang, Y., Li, Q., Torres-Rueda, S., & Li, J. (2020). The Structure and Parameterization of the Breast Cancer Transition Model Among Chinese Women. *Value in Health Regional Issues*, *21*, 29–38. https://doi.org/10.1016/j.vhri.2019.05.003

IARC, I. A. for R. on C. (2024). *No Title*. https://www.iarc.who.int/cancer-type/breast-cancer/

Igwenagu, C. M., & Egemba, N. C. (2021). A Multi-State Markov Modelling of Breast Cancer Progression. *IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)*, *Volume 8*(Issue 3 (September, 2021)), 172–187. http://www.mdcjournals.org/pdf/163368710320210810102358.pdf

Islami, F., Baeker Bispo, J., Lee, H., Wiese, D., Yabroff, K. R., Bandi, P., Sloan, K., Patel, A. V, Daniels, E. C., & Kamal, A. H. (2024). American Cancer Society's report on the status of cancer disparities in the United States, 2023. *CA: A Cancer Journal for Clinicians*, *74*(2), 136–166.

Jani, P., Craig, H., Are, C., & Rooprai, G. (2021). Cancer on the Global Stage: Incidence and Cancer-Related Mortality in Kenya WORLD HEALTH ORGANIZATION REGION: AFRICA. *The ASCO Post Journal*. https://ascopost.com/issues/february-25-2021/cancer-on-the-global-stage-incidence-and-cancer-related-mortality-in-kenya/

Kailemia, P. N., Lee, E. C., & Renfrew, M. J. (2023). Intersection of social determinants of symptomatic breast cancer presentation in a rural setting: A critical ethnographic study. *Journal of Advanced Nursing*, *79*(5), 1882–1897.

Kisiangani, J., Baliddawa, J., Marinda, P., Mabeya, H., Choge, J. K., Adino, E. O., & Khayeka-Wandabwa, C. (2018). Determinants of breast cancer early detection for cues to expanded control and care: the lived experiences among women from Western Kenya. *BMC Women's Health*, *18*, 1–9.

Li, Y., Hua, R., He, J., & Zhang, H. (2022). Survival Contradiction in Stage II, IIIA, And IIIB Colon Cancer: A Surveillance, Epidemiology, and End Result-Based Analysis. *Evidence-Based Complementary and Alternative Medicine*, *2022*(1), 1–

8. https://doi.org/10.1155/2022/4088117

Li, Y., Zhou, Y., Mao, F., Guan, J., Lin, Y., Wang, X., Zhang, Y., Zhang, X., Shen, S., & Sun, Q. (2019). The influence on survival of delay in the treatment initiation of screening detected non-symptomatic breast cancer. *Scientific Reports*, *9*(1), 10158.

McGarvey, N., Gitlin, M., Fadli, E., & Chung, K. C. (2022). Increased healthcare costs by later stage cancer diagnosis. *BMC Health Services Research*, *22*(1), 1155. https://doi.org/10.1186/s12913-022-08457-6

Mirzapour, S. A., Mazur, T., Sharp, G., & Salari, E. (2019). Intra-fraction motion prediction in MRI-guided radiation therapy using Markov processes. *Physics in Medicine & Biology*, *64*(19), 195006.

Mitchell, P. D., Dittmar, J. M., Mulder, B., Inskip, S., Littlewood, A., Cessford, C., & Robb, J. E. (2021). The prevalence of cancer in Britain before industrialization. *Cancer*, *127*(17), 3054–3059.

MOH. (2020). Kenya Cancer Policy 2019-2030. In *Ministry of Health* (Vol. 2020). http://guidelines.health.go.ke/#/category/7/448/meta

MOH. (2021). Breast-Cancer-Screening-and-Early-Diagnosis-Action-Plan-2021-2025. In *Ministry of Health*. http://guidelines.health.go.ke:8000/media/Breast-Cancer-Screening-and-Early-Diagnosis-Action-Plan-2021-2025.pdf

MOH. (2023). National Cancer Control Strategy 2023-2027. In *Ministry of Health*. http://guidelines.health.go.ke:8000/media/NATIONAL_CANCER_CONTROL_STRATEGY_2023-2027_7uTQQP4.pdf

Mutebi, M., Anderson, B. O., Duggan, C., Adebamowo, C., Agarwal, G., Ali, Z., Bird, P., Bourque, J., DeBoer, R., & Gebrim, L. H. (2020). Breast cancer treatment: A phased approach to implementation. *Cancer*, *126*, 2365–2378.

NCCP. (2023). National Cancer Control Program. *Ministry of Health*. https://nccp.or.ke/policies-and-guidelines/

Ngwa, W., Addai, B. W., Adewole, I., Ainsworth, V., Alaro, J., Alatise, O. I., Ali, Z., Anderson, B. O., Anorlu, R., & Avery, S. (2022). Cancer in sub-Saharan Africa: a lancet oncology commission. *The Lancet Oncology*, *23*(6), e251–e312.

Olaleye, O., & Ekrikpo, U. (2017). Epidemiology of Cancers in Sub-Saharan Africa. In *Cancer in Sub-Saharan Africa* (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-52554-9_1

Onu, C. C., Kanbar, L. J., Shalish, W., Brown, K. A., Sant'Anna, G. M., Kearney, R. E., & Precup, D. (2017). A semi-Markov chain approach to modeling respiratory patterns prior to extubation in preterm infants. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2022–2026.

Rivera-Franco, M. M., & Leon-Rodriguez, E. (2018). Delays in breast cancer detection

and treatment in developing countries. *Breast Cancer: Basic and Clinical Research*, *12*, 1178223417752677.

Rodriguez, P. J., Ward, Z. J., Long, M. W., Austin, S. B., & Wright, D. R. (2021). Applied methods for estimating transition probabilities from electronic health record data. *Medical Decision Making*, *41*(2), 143–152.

Shockney, L. D. (2025). Stage 3 Breast Cancer Overview. *National Breast Cancer Foundation*. https://www.nationalbreastcancer.org/breast-cancer-stage-3/

Šlegerová, L., & Kopečková, K. (2023). The cost-effectiveness of Pertuzumab for the treatment of metastatic HER2+ breast cancer in Czechia: A semi-Markov Model using cost states. *Value in Health Regional Issues*, *38*, 118–125.

Sun, L., Legood, R., dos-Santos-Silva, I., Gaiha, S. M., & Sadique, Z. (2018). Global treatment costs of breast cancer by stage: a systematic review. *PloS One*, *13*(11), e0207993.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). 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

Tenny, S., & Hoffman, M. R. (2017). *Prevalence*.

Thomas, R. B., Maio, V., Chen, A., Park, S., Waters, D., Keith, S. W., Walsh, K., Handley, N., & Csik, V. P. (2021). Breast cancer stage is associated with exceeding target price in the oncology care model. *JCO Oncology Practice*, *17*(11), e1660–e1667.

Vargas-Calixto, J., Wu, Y., Kuzniewicz, M., Cornet, M.-C., Forquer, H., Gerstley, L., Hamilton, E., Warrick, P. A., & Kearney, R. E. (2022). Multi-chain semi-markov analysis of intrapartum cardiotocography. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1948–1952.

Velloso, F. J., Bianco, A. F. R., Farias, J. O., Torres, N. E. C., Ferruzo, P. Y. M., Anschau, V., Jesus-Ferreira, H. C., Chang, T. H.-T., Sogayar, M. C., & Zerbini, L. F. (2017). The crossroads of breast cancer progression: insights into the modulation of major signaling pathways. *OncoTargets and Therapy*, 5491–5524.

WHO. (2023). *The Global Breast Cancer Initiative (GBCI)*. World Health Organization. https://www.who.int/initiatives/global-breast-cancer-initiative

WHO. (2024). BREAST CANCER. *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

WHO. (2025). CANCER. *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/cancer

Wong, M. C. S., Ding, H., Wang, J., Chan, P. S. F., & Huang, J. (2019). Prevalence

and risk factors of colorectal cancer in Asia. *Intestinal Research*, *17*(3), 317–329.

Yin, A., Moes, D. J. A. R., van Hasselt, J. G. C., Swen, J. J., & Guchelaar, H. (2019). A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT: Pharmacometrics & Systems Pharmacology*, *8*(10), 720–737.

APPENDICES

Appendix I: Data Collection Checklist

RESEARCH INSTRUMENT FOR CANCER REGISTRIES

I am Peter Kamau Waweru, a Master's student at the University of Embu, School of Pure and Applied Sciences, Department of Mathematics and Statistics. I am conducting a study titled "***SEMI-MARKOVIAN ANALYSIS OF THE PROGNOSIS OF BREAST CANCER DURING MANAGEMENT IN KENYA***". The purpose of the study is to determine the prevalence and analyze the prognosis of breast cancer during management using semi-Markov modeling in Kenya, taking a case study of Mt. Kenya region. This will involve constructing a transition probability matrix and determining the prevalence of breast cancer at different stages.

To support this research, I kindly request access to secondary data on breast cancer cases from your registry. Below is a checklist outlining the variables of interest which will help have the relevant information for analysis. The support provided will be invaluable to the success of this research. The data provided will be used solely for this study and will be treated with utmost confidentiality.

In case of any clarification, you can contact me or my supervisors through the contacts given below:

1. Peter Kamau Waweru
Tel.: 0704182967
Email: 1660@student.embuni.ac.ke

2. Dr. Peter Kailemia
Tel.: 0720033064
Email: ntoiti.peter@embuni.ac.ke

3. Dr. Edwine Benson Atitwa
Tel.: 0724556119
Email: atitwa.benson@embuni.ac.ke

4. Dr. Zakayo Ndiku Morris
Tel.: 0726670862
Email: ndiku.zakayo@embuni.ac.ke

A. Patient Demographics

a) Age at diagnosis .....
b) Sex….....
c) County of residence .....
d) Date of cancer diagnosis .....
e) Cancer type diagnosed……...
f) Stage at diagnosis (Tick appropriately ✓)

Stage 0 Stage I Stage II Stage III Stage IV

g) Comorbidity: Accompanying condition
   a. Medical……
   b. physical …….

B. Recurrence Details

Recurrence information

Yes No

If Yes, number and date of recurrence (1^{st}, 2^{nd} , 3^{rd} recurrence)

.....

C. Treatment Information

Treatment Type	**Cancer Stage**	**Cost in Ksh.**	**Date of initiation**	**Treatment Outcome/Response (Stage transition)**
1. Surgery				
2. Chemotherapy				
3. Radiation				
4.Hormonal therapy				
5. Others				

Appendix II: Institutional Ethical Review Approval

CHUKA UNIVERSITY

Knowledge is Wealth *(Sapientia divitia est)* Akili ni Mali

CHUKA UNIVERSITY INSTITUTIONAL ETHICS REVIEW COMMITTEE

Telephones: 020-2310512/18 **P. O. Box 109-60400, Chuka**
Direct Line: 0772894438 **Email: info@chuka.ac.ke,** **Website: www.chuka.ac.ke**

19th November, 2024

REF: CUIERC/ NACOSTI/651
TO: **Peter Kamau Waweru**

RE: Semi – Markovian Analysis of the Prognosis of Breast Cancer During Management in Kenya

This is to inform you that *Chuka University IERC* has reviewed and approved your above research proposal. Your application approval number is *NACOSTI/NBC/AC-0812.* The approval period is 19th November, 2024 – 19th November, 2025.

This approval is subject to compliance with the following requirements;

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by *Chuka University IERC.*
iii. Death and life threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to *Chuka University IERC* within 72 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to *Chuka University IERC* within 72 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to *Chuka University IERC.*

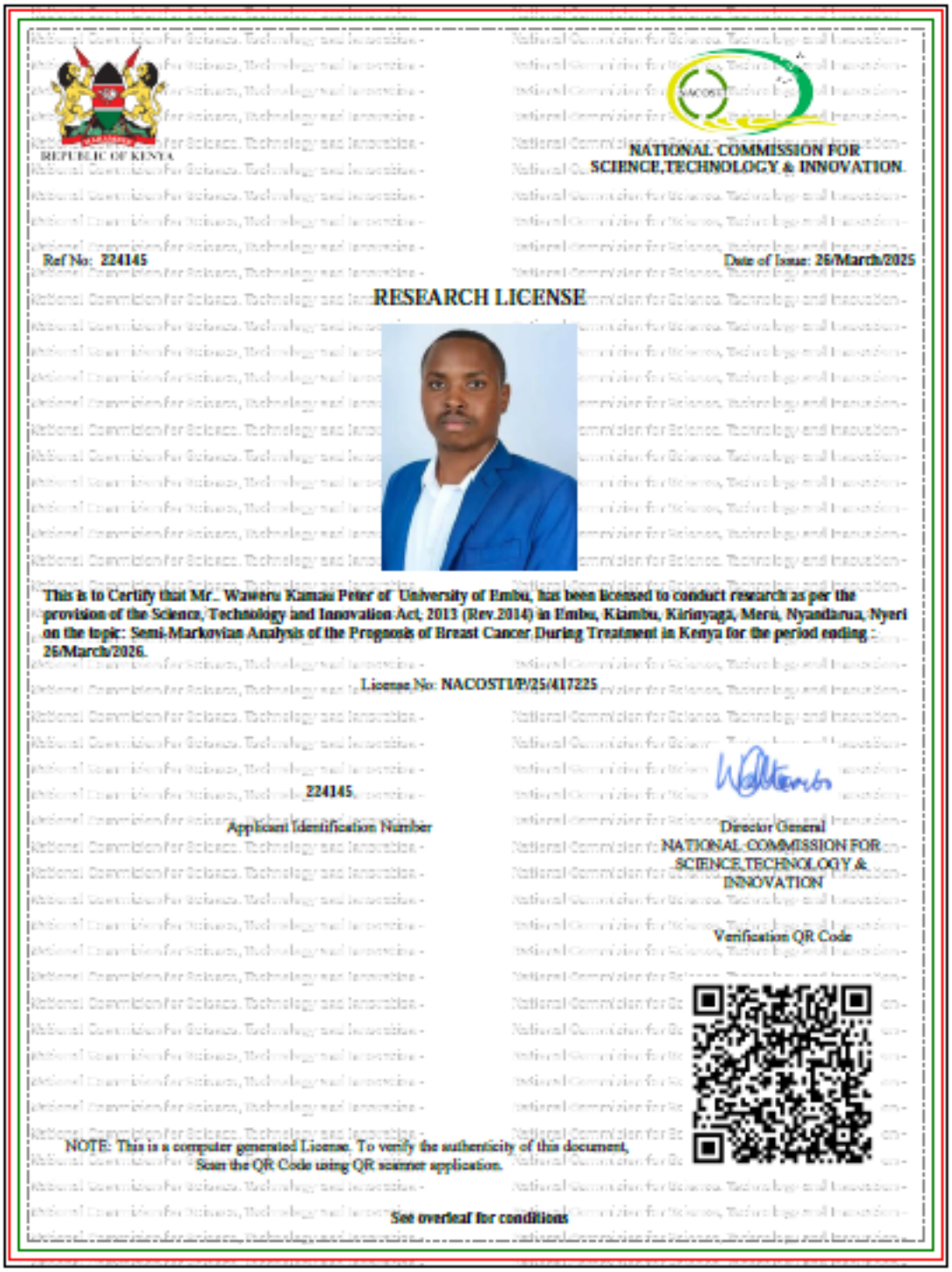
Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology and Innovation (NACOSTI) https://oris.nacosti.go.ke and also obtain other clearances needed.

Yours sincerely

Dr. Benjamin Kanga
SECRETARY

INSTITUTIONAL ETHICS REVIEW COMMITTEE
19 NOV 2024
P.O. Box 109-60400, CHUKA
CHUKA UNIVERSITY

Appendix III: National Commission for Science, Technology and Innovation (NACOSTI) Research License.



REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 224145

Date of Issue: 26/March/2025

RESEARCH LICENSE

This is to Certify that Mr.. Waweru Kamau Peter of University of Embu, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Embu, Kiambu, Kirinyaga, Meru, Nyandarua, Nyeri on the topic: Semi-Markovian Analysis of the Prognosis of Breast Cancer During Treatment in Kenya for the period ending : 26/March/2026.

License No: NACOSTI/P/25/417225

224145

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions