Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2025, 2025:85

https://doi.org/10.28919/cmbn/9271

ISSN: 2052-2541

# SEMI-MARKOVIAN ANALYSIS OF THE PROGNOSIS OF BREAST CANCER BETWEEN DIAGNOSIS AND TREATMENT INITIATION IN KENYA: A CASE STUDY OF TWO COUNTIES

PETER KAMAU WAWERU[1,*], EDWINE BENSON ATITWA[1], ZAKAYO NDIKU MORRIS[1], PETER NTOITI KAILEMIA[2]

[1]Department of Mathematics and Statistics, University of Embu, Embu, Kenya

[2]School of Nursing, Meru University of Science and Technology, Meru, Kenya



**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 transiting 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 probability of remaining at stage II or stage III after diagnosis

*Corresponding author

E-mail address: kamaupeter.waweru1@gmail.com

Received April 2, 2025

was found to decrease with the increase in the waiting time before treatment initiation. 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.

**Keywords:** semi-Markov analysis; transition probability; mortality rate; female breast cancer; public health policies.
**2020 AMS Subject Classification:** 62P10.

## 1. INTRODUCTION

According to World Health Organization (WHO), cancer is a broad category of illnesses that can originate in any organ or area of the body in humans when cells grow uncontrollably and unnaturally and then spread to other areas of the body and eventually other organs. 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. WHO identifies cancer as a major public health burden worldwide [1]. In the year 2020, WHO showed that approximately one in every six deaths globally was as a result of cancer [1]. The WHO reported that in 2022, 20 million new cases of cancer were reported, while roughly 10 million people died from cancer [2]. The report also 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 is approximated that 1 in every 12 women and 1 in every 9 men die of cancer [2]. 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 [3]. This underscores the growing burden of cancer in the entire globe.

WHO cancer reports indicate that FBC is the most prevalent type of cancer among females [2]. FBC is more prevalent in females than any other type of cancer in terms of prevalence rate [4]. According to the 2020 report released by the WHO, the most common cancer was FBC, accounting for 2.26 million new cases and a total of 685,000 deaths out of an approximate total of around 10 million deaths worldwide [1]. The 2020 GLOBOCAN report estimated that 520,158 deaths occurred among the 801,392 new cancer cases reported in Sub-Saharan Africa, where FBC represented the largest percentage of the mortality rate [5]. States in Sub-Saharan Africa need to develop enhanced surveillance systems to assess the magnitude of the cancer problem, allowing for improved monitoring and planning for the disease. Olaleye and Ekrikpo noted that Sub-Saharan Africa has experienced relatively higher cancer mortality rates, attributed to changes in diet, lifestyle, and population dynamics across the continent [6]. The two researchers concluded that

there is an urgent need for improved cancer diagnosis and treatment throughout the region. is and treatment across the region.

In Kenya, according to GLOBOCAN, FBC constituted approximately 16% of the total number of new cases in Kenya in 2020 [7]. Between 2012 and 2018, 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% [8].

In a bid to control cancer rates by reducing incidence and mortality while improving the lives of cancer patients, the Government of Kenya (GoK) established the National Cancer Control Program 2023-2027 [9]. The program aims to ensure that evidence-based strategies are implemented to aid in prevention, early detection, diagnosis, treatment, and palliative care, while optimally utilizing available resources. Projections indicate that by 2028, Kenya is expected to see approximately 58,000 new cancer cases, with FBC likely representing the largest number of cases [9]. This underscores the need for timely planning to facilitate cancer control interventions, including prevention, early detection, diagnosis, treatment, and palliative care.

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 [10]. During the same year, Kenya launched Breast Cancer Screening and Early Diagnosis Action Plan 2021-2025, which was in line with the GBCI [11]. 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 [12]. 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 [13]. The National Cancer Taskforce report released in 2022 identified that there is low prioritization of cancer research and called out for cancer research to address the most prevalent cancers in the country for control [14]. The continued breast cancer burden in the country indicates that it is an area that requires prioritized cancer control strategies informed by relevant research. It is, therefore, important to understand the transition dynamics of FBC in order to control it.

In order to contribute to the realization of the of FBC control efforts in Kenya as well as the larger GBCI, it is necessary to gain a comprehensive understanding of the progression dynamics of FBC

by analyzing the transitions from one stage to the other. Transitional probabilities give valuable insights into the probability of moving from one state to another within a given time which can be estimated using Markov models. Given that the transitions of FBC are non-homogenous, semi-Marko analysis is appropriate for understanding the transition and progression dynamics of the disease. 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. A study was done on the prognosis of cancer and applied a Semi Markov Process [15]. However, the gap in their study was that they did not have patient data to study the prognosis of cancer. This study, therefore, sought to use FBC patient records data to gain insights into the transition dynamics of cancer patients between diagnosis and treatment initiation.

The study findings may help researchers and healthcare professionals understand how breast cancer evolves from one state to another. The findings of the study can help in promoting preparedness in the health sector and come up with FBC early detection strategies by the MOH in both national and county governments. It may also contribute to improved FBC diagnosis and treatment strategies that will reduce the delays in treatment initiation. The findings may enable proper and accurate resource allocation from both levels of government and the affected households to eradicate treatment initiation barriers.

## 2. Material and Methods

### 2.1 Data

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 on breast cancer diagnosis, staging information, and time taken before treatment initiation.

### 2.2 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 two 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 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 all through 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.2.3 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 in percentage to determine the prevalence of breast cancer per stage.

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

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

$N$ – Total number of breast cancer patients

**2.4 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 using semi-Markov models to estimate breast cancer progression among patients. This is because the waiting times or the time spent in each of the state varies 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:

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

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 other entries in a particular row. This can be represented as shown below;

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

*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.

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

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 $q_{ij}$ will be given by the following formula;

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

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

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;

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

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 assure that the model assumptions, like that of continuous time distribution, are satisfied.

**2.5 Model Fitting**

Given that the observed waiting time is continuous, as seen in the data, a semi-Markov model is 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 stage $j$

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

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

$f_{ij}(t)$ 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;

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

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

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

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

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

**2.6 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;

$$P(t) = \left\{P_{ij}(t)\right\} \tag{17}$$

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

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

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

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

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 stage $j$ are given as

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

with a probability condition given as

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

The transition probabilities obtained will 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 [16], 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. MAIN RESULTS

### 3.1 Prevalence of breast cancer

Descriptive statistics for the prevalence of FBC at diagnosis are displayed in **Error! Reference source not found.** 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 stage 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.

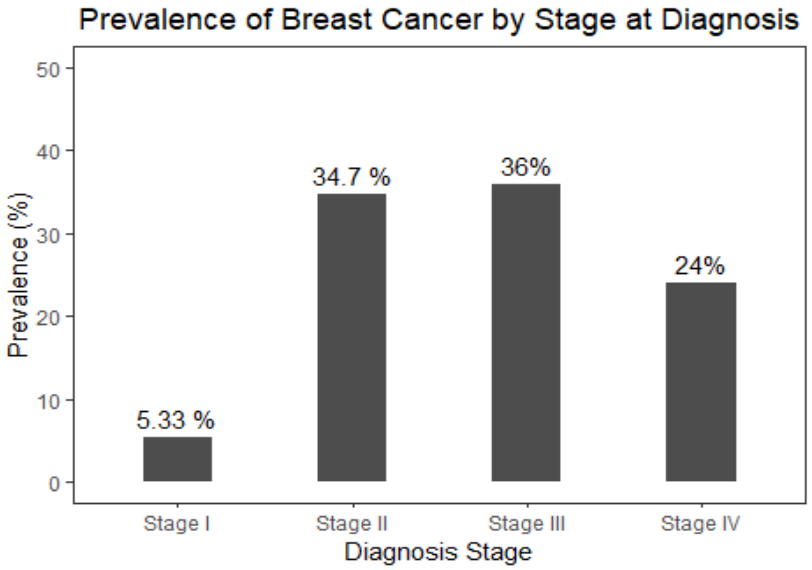


**Figure 1: Stage prevalence at diagnosis**

The prevalence of FBC per stage at the time of treatment of 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. **Error! Reference source not found.** below show the prevalence of FBC per stage at the treatment initiation stage.

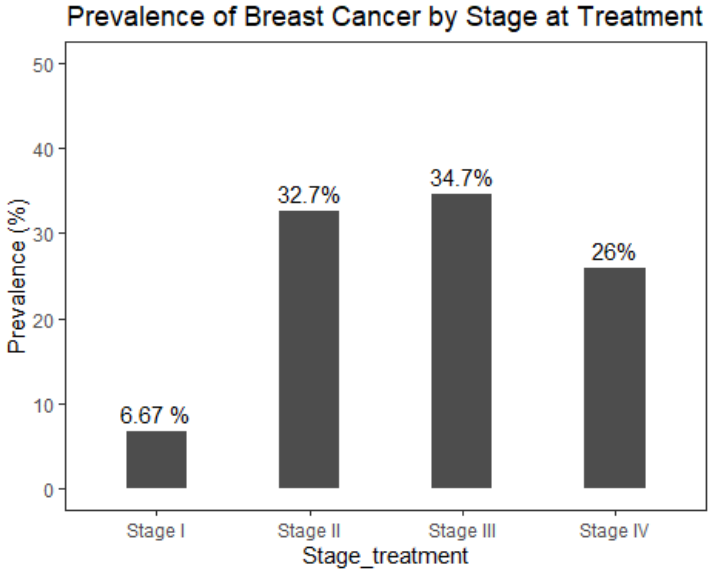


**Figure 2: Stage Prevalence at Treatment**

### 3.2 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;

$$\mathrm{N} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 2 & 48 & 0 & 2 \\ 0 & 1 & 52 & 1 \\ 0 & 0 & 0 & 36 \end{pmatrix} \tag{26}$$

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 2 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 backward 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;

$$\mathrm{T} = \begin{pmatrix} 12 & 0 & 0 & 0 \\ 4.5 & 118 & 0 & 4 \\ 0 & 2 & 126 & 4 \\ 0 & 0 & 0 & 66.5 \end{pmatrix} \tag{27}$$

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;

$$Q = \begin{pmatrix} 0.6667 & 0 & 0 & 0 \\ 0.4444 & 0.4068 & 0 & 0.5 \\ 0 & 0.5 & 0.4127 & 0.25 \\ 0 & 0 & 0 & 0.5414 \end{pmatrix} \tag{28}$$

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.

$$Q = \begin{pmatrix} -0.6667 & -0.6667 & 0 & 0 \\ 0.4444 & -0.9444 & 0 & 0.5 \\ 0 & 0.5 & -0.75 & 0.25 \\ 0 & 0 & 0.5414 & -0.5414 \end{pmatrix} \quad (29)$$

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

**Table 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:

$$\hat{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.0160 & -0.0323 & 0 & 0.0163 \\ 0 & 0.0080 & -0.0156 & 0.0075 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (30)$$

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 ($t = 1$)

$$P(1) = e^{\hat{Q}(1)} \tag{31}$$

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 II, and 0.0075 of transiting to stage IV. The results for different time intervals (*t*) in months are as shown below

$$P(1) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.01578 & 0.96818 & 0 & 0.01604 \\ \textbf{III} & 0.00006 & 0.00785 & 0.98455 & 0.00754 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{32}$$

After two months ($t = 2$)

$$P(2) = e^{\hat{Q}(2)} \tag{33}$$

$$P(2) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03106 & 0.93738 & 0 & 0.03156 \\ \textbf{III} & 0.00025 & 0.01532 & 0.96934 & 0.01508 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{34}$$

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

$$P(2.4667) = e^{\hat{Q}(2.4667)} \tag{35}$$

$$P(2.4667) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03475 & 0.92993 & 0 & 0.03156 \\ \textbf{III} & 0.00031 & 0.01711 & 0.96563 & 0.01695 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{36}$$

After three months ($t = 3$)

$$P(3) = e^{\hat{Q}(3)} \tag{37}$$

$$P(3) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.04585 & 0.90755 & 0 & 0.04660 \\ \textbf{III} & 0.00055 & 0.02244 & 0.95437 & 0.02264 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{38}$$

After four months ($t = 4$)

$$P(4) = e^{\hat{Q}(4)} \tag{39}$$

$$P(4) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.06017 & 0.87868 & 0 & 0.06115 \\ \textbf{III} & 0.00097 & 0.02922 & 0.93962 & 0.03019 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{40}$$

After five months ($t = 5$)

$$P(5) = e^{\hat{Q}(5)} \tag{41}$$

$$P(5) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.07403 & 0.85072 & 0 & 0.07524 \\ \textbf{III} & 0.00149 & 0.03566 & 0.92511 & 0.03774 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{42}$$

After six months ($t = 6$)

$$P(6) = e^{\hat{Q}(6)} \tag{43}$$

$$P(6) = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.08746 & 0.82365 & 0 & 0.08889 \\ \textbf{III} & 0.00211 & 0.04179 & 0.91082 & 0.04528 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{44}$$

After twelve months ($t = 12$)

$$P(12) = e^{\hat{Q}(12)} \tag{45}$$

$$P(12) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.15949 & 0.67841 & 0 & 0.16210 \\ \textbf{III} & 0.00769 & 0.07248 & 0.82959 & 0.09024 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{46}$$

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

$$P(36) = e^{\hat{Q}(36)} \tag{47}$$

$$P(36) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.34110 & 0.31223 & 0 & 0.34667 \\ \textbf{III} & 0.04835 & 0.12404 & 0.57093 & 0.25668 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{48}$$

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

$$P(72) = e^{\hat{Q}(72)} \tag{49}$$

$$P(72) = \begin{pmatrix} \textbf{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.44760 & 0.09749 & 0 & 0.45491 \\ \textbf{III} & 0.11826 & 0.10954 & 0.32596 & 0.44623 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \tag{50}$$

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 **Error! Reference source not found.**. 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 $t = 72$. 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 **Error! Reference source not found.**. These results are similar to the results by [17], who found out that stage III breast cancer is locally advanced and has tumors that are larger than those of stages I and II. 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 $t = 72$. The transition probabilities in this study were consistent with findings from other studies like that of Huang et al. [18] , 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. 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 tumors than during diagnosis. Li et al. [19] in their study give the survival contradiction of the different cancer stages in their study where stages II and III are further divided into more tumor-specific categories of stages [19]. This study considers distinct stages and not tumor-specific stages of breast cancer, and it 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. Another study found that luminal B breast cancers (locally advanced-stage cancers) had greater growth rates of tumors compared to luminal A breast cancers [20]. 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).

**Table 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 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.

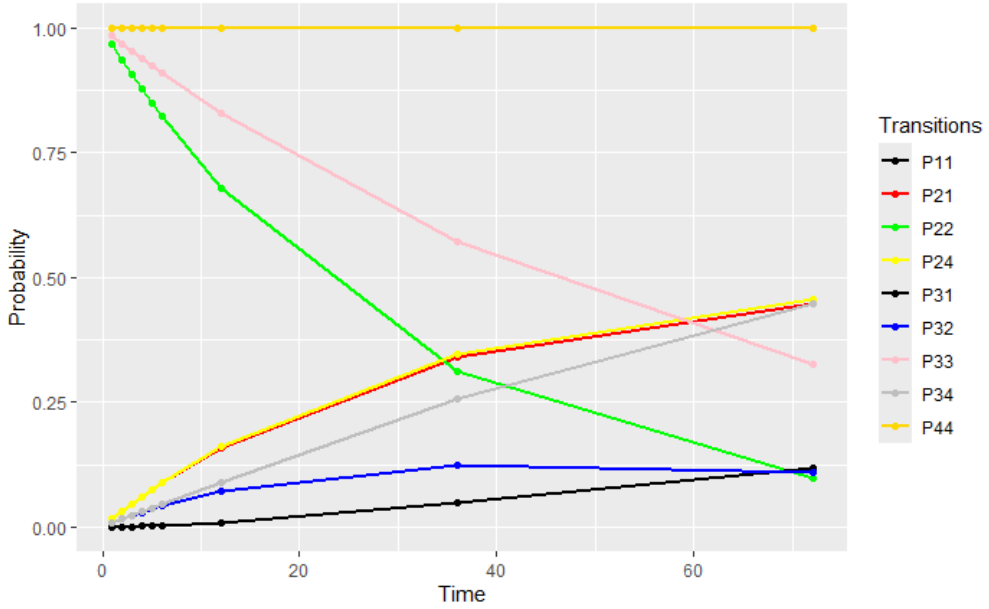


**Figure 3: Transition probability graph**

If we extend the time the patients take before they seek treatment after diagnosis to 240 months to observe the long-term likely behavior of the transition probabilities, we obtain Figure 4. Some of the key highlights from the graph are that 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 backward from stage II to me and from stage III to 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 given by the line showing the trend of the probability of moving from stage III to IV in Figure 4.

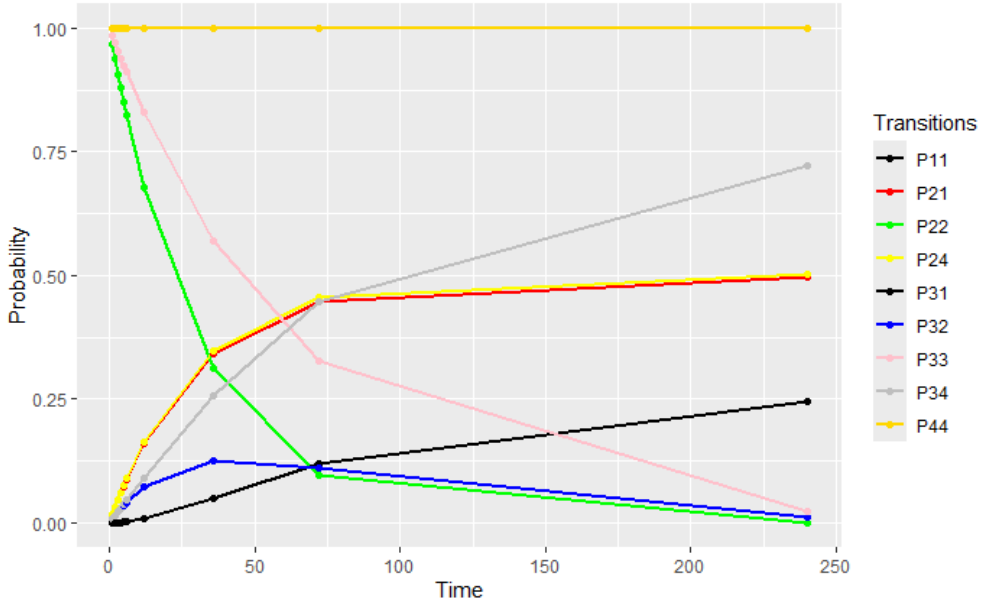


**Figure 4: Extrapolated Transition Probability Graph**

### 3.3 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

$$P = \begin{pmatrix} \boldsymbol{Stage} & \textbf{I} & \textbf{II} & \textbf{III} & \textbf{IV} \\ \textbf{I} & 1 & 0 & 0 & 0 \\ \textbf{II} & 0.03475 & 0.92993 & 0 & 0.03156 \\ \textbf{III} & 0.00031 & 0.01711 & 0.96563 & 0.01695 \\ \textbf{IV} & 0 & 0 & 0 & 1 \end{pmatrix} \quad (51)$$

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 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 the stage I is very unlikely and stands at 0.03%.

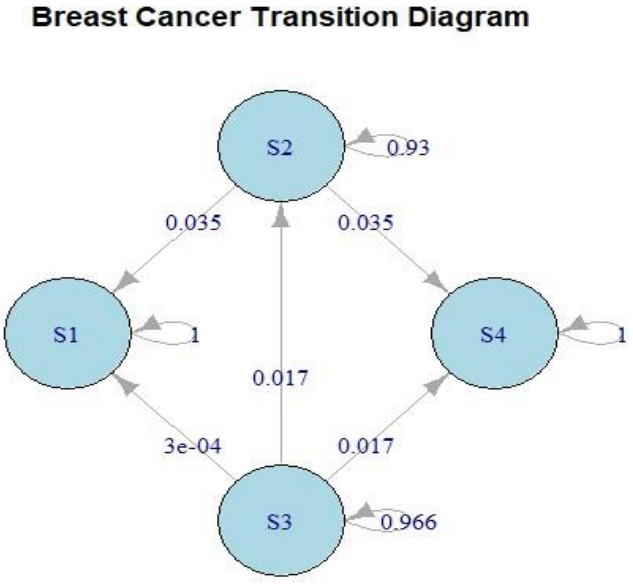


**Figure 5: Breast Cancer Transition Probability Diagram**

## 4. CONCLUSION

This study employed semi-Markov models to analyze breast cancer progression across its four possible stages. The model incorporated the use of transition probability matrices and waiting times between the diagnosis of the disease and treatment to understand the progression dynamics. One of the major findings from this study is that stage I and stage IV were found to be absorbing states based on the semi-Markovian analysis of the data of breast cancer patients obtained from the registry. Stage I, being an absorbing state, suggests that the cases may have been detected early and took time before transiting 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 backward 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 was between the four stages and not any other stage. There could be patients who might have died after stage IV or any other of the three stages, but this study does not consider the death stage and the probability of getting to the death stage.

This aligns well with the existing literature that suggests 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 backward 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. 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 use of a semi-Markov model has been proved, 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 model time-dependent transition probabilities and patterns.

## 5. RECOMMENDATIONS

Based on the findings of this study, several recommendations can be made to improve the diagnosis and treatment of breast cancer. This study recommends more emphasis on programs that promote early breast cancer detection based on the finding that stage I was found to be an absorbing state. 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. However, 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 down staging of FBC before treatment is initiated.

## STUDY LIMITATION

The completeness and accuracy of the data used in this study depended on the accuracy in recording the patient information in the cancer registry. Records with missing or incomplete information were discarded and not considered in this study.

## 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.

## ASSUMPTIONS

This study assumes that data was accurately recorded in the cancer registry. It also assumes that no external interventions affected the progression of the disease.

## CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

## REFERENCES

[1] WHO, Cancer, World Health Organization, 2022. https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] WHO, Breast Cancer, World Health Organization, 2024. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[3] N. McGarvey, M. Gitlin, E. Fadli, K.C. Chung, Increased Healthcare Costs by Later Stage Cancer Diagnosis, BMC Health Serv. Res. 22 (2022), 1155. https://doi.org/10.1186/s12913-022-08457-6.

[4] H. Sung, J. Ferlay, R.L. Siegel, et al. Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA: Cancer J Clin. 71 (2021), 209-249. https://doi.org/10.3322/caac.21660.

[5] F. Bray, D.M. Parkin, F. Gnangnon, et al. Cancer in Sub-Saharan Africa in 2020: a Review of Current Estimates of the National Burden, Data Gaps, and Future Needs, Lancet Oncol. 23 (2022), 719-728. https://doi.org/10.1016/s1470-2045(22)00270-4.

[6] O. Olaleye, U. Ekrikpo, Epidemiology of Cancers in Sub-Saharan Africa, in: O.A. Adedeji (Ed.), Cancer in Sub-Saharan Africa, Springer International Publishing, Cham, 2017: pp. 3–19. https://doi.org/10.1007/978-3-319-52554-9_1.

[7] J. Ferlay, M. Colombet, I. Soerjomataram, et al. Cancer Statistics for the Year 2020: An Overview, Int. J. Cancer 149 (2021), 778-789. https://doi.org/10.1002/ijc.33588.

[8] P.G. Jani, MBChB, MMed, et al. Cancer on the Global Stage: Incidence and Cancer-Related Mortality in Kenya, https://ascopost.com/issues/february-25-2021/cancer-on-the-global-stage-incidence-and-cancer-related-mortality-in-kenya.

[9] NCCP, The National Cancer Control Program, https://nccp.or.ke.

[10] WHO, The Global Breast Cancer Initiative, World Health Organization, 2023. https://www.who.int/initiatives/global-breast-cancer-initiative.

[11] MOH, Breast Cancer Screening and Early Diagnosis Action Plan 2021-2025, Ministry of Health, 2021.

[12] MOH, Kenya Cancer Policy 2019-2030, Ministry of Health, 2020.

[13] MOH, National Cancer Control Strategy 2023-2027, Ministry of Health, 2023.

[14] MOH, National Cancer Taskforce Report, Ministry of Health, 2022.

[15] M.S. Dalabanjan, P. Agrawal, T. Deepthi, Prognosis of Cancer - A Semi Markov Process, Int. J. Eng. Adv. Technol. 10 (2021), 146-150.

[16] C.M. Igwenagu, N.C. Egemba, A Multi-State Markov Modelling of Breast Cancer Progression, IJSAR J. Math. Appl. Stat. 8 (2021), 172-187.

[17] LD. Shockney, Stage 3 Breast Cancer Overview, National Breast Cancer Foundation, 2025 https://www.nationalbreastcancer.org/breast-cancer-stage-3.

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

[19] Y. Li, R. Hua, J. He, H. Zhang, Survival Contradiction in Stage Ii, Iiia, and Iiib Colon Cancer: a Surveillance, Epidemiology, and End Result-Based Analysis, Evidence-Based Complement. Altern. Med. 2022 (2022), 4088117. https://doi.org/10.1155/2022/4088117.

[20] Z. Alaidy, A. Mohamed, D. Euhus, Breast Cancer Progression When Definitive Surgery Is Delayed, Breast J. 27 (2021), 307-313. https://doi.org/10.1111/tbj.14177.