MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

Research Article

A Multi-State Markov Modelling of Breast Cancer Progression

Igwenagu, C. M.[1] and Egemba N.C.[2]

[1]*Department of Industrial Mathematics/ Applied Statistics, Enugu State University of Science and Technology, Enugu State, Nigeria*

ABSTRACT

This study analyzed the progression of breast cancer in patients who were in different stages of breast cancer using a multi-state Markov model to analyze the four stages of the disease in Enugu State from 2009 – 2019. The objectives of the study include: to find the probability layout and to ascertain the transition probability matrix of the breast cancer progression across the four stages of the disease, to estimate the prediction and to predict the future structure of the breast cancer progression across the four stages of the disease for 2020, 2021, and 2022.The data used for this study was a secondary data collected from Enugu State University of Science and Technology Teaching Hospital Parklane for a period of 11 years.The R-programming software version 4.1.0 was used to analyze the data. The transition probability matrix of the breast cancer progression and the prediction equation of the breast cancer progression were estimated, the future structure of the breast cancer progression across the four stages of the disease for 2020, 2021, and 2022 were predicted and the sojourn time the patients in the four stages of breast cancer were ascertained. The findings of the study showed that the distribution of breast cancer cases for Stage II is the most commonly reported stage of breast cancer, while Stage IV is the most common less reported stage. Also, the analysis of the sojourn time showed that a Stage II patient remain longer and have 93% chance of progressing to Stage III.

Keywords: *Breast Cancer, Multi-State Model, R-programming, Sojourn Time, Transition Probability Matrix.*

1. INTRODUCTION

Breast cancer is defined as a family of diseases where cells in the breast tissue grow and divide without normal growth and the growth of cells form a mass or lump called a tumor which is either benign (not cancerous) or malignant (cancerous) [1]. Breast cancer originates most commonly from breast tissue, and from the inner lining of milk ducts or the lobules that supply the ducts with milk [2]. Breast tumors can be invasive or non-invasive and the prognosis is often affected by characteristics such as the hormone receptor and human epidermal growth factor receptor 2 (Her2) status.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

There are multiple various complications around breast cancer and the response to treatment is unsatisfactory and incomplete in many cases. However, medical advances have shown that one third of all cancers are preventable and a further one third of cases if diagnosed timely are potentially curable [3]. According to [4], women with different types of breast cancer react differently to treatment. The worst breast cancer prognosis is when the cancer has already metastasized at the time of diagnosis [4].

Dual carcinoma in situ (DCIS) is a non-invasive breast cancer [1]. This is the case when the milk ducts have not spread to nearby breast tissue. This non-invasive breast cancer can develop into invasive breast cancer over time if it is not treated. Invasive breast cancer is cancer that has spread from the original location into another part of the breast tissue. If cancer cells metastasize, they spread to other parts of the body and cause tumors to grow there, as well as to the lymph nodes [5]. Consequently, invasive breast cancer has a poorer prognosis than DCIS. Hormone receptors are breast cancer cells that have special proteins inside which needs estrogen and/or progesterone to grow [1]. When breast cancers have many hormone receptors, the cancers are called hormone receptor positive cancers. Hormone receptor positive can mean either estrogen receptor (ER) positive or progesterone receptor (PR) positive. The status strongly influences the course of treatment and therefore the cost of treatment.However, the origin, progression and treatment of breast cancer are stochastic in nature. Therefore, a multi-state Markov model will be used to analyze the progression of breast cancer in this study.

The objectives of this study include: to find the probability layout of the breast cancer progression across the four stages of the disease, to ascertain the transition probability matrix of the breast cancer progression across the four stages of the disease, to estimate the prediction equation of the breast cancer progression across the four stages of the disease, to predict the future structure of the breast cancer progression across the four stages of the disease for 2020, 2021, and 2022 and to ascertain the sojourn time of the patients in the four stages of breast cancer.

II REVIEW OF RELATED LITERATURE

Multi-state models (MSM) are models for continuous time stochastic processes allowing individuals to move among a finite number of states [6]. They are statistical tools that allow medical researchers to characterize the evolution of disease natural histories through discrete states, including progressive disease like cancer. They are also flexible tool for analyzing complex time- to-event problems with multiple endpoints, especially in chronic diseases where patients move through different states [7]. Multi-state models are considered in the field of survival for modelling illness that evolve through stages over time. They can be developed by applying several techniques such as non-parametric semi-parametric and stochastic processes particularly Markov processes [8].

Multi-state models are the most common choice of model to analyze longitudinal survival data [9]. This technique is widely used in various fields such as medicine, physics, biology, economics and others.

A multi-state model is a stochastic process which occupies one of a set of discrete states, at any time point [10]. Different health states can be defined in its simplest form as healthy, sick or diseased. The states may represent different health situations of the subject [9]. A transition or event refers to a change of state which corresponds for example to an outbreak of disease or even death. The state structure and the form of the hazard function for each possible transition is specified in the full statistical model [10].

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

The possibility of projecting the number of persons who will be in a certain state of cancer, based on transition probabilities or intensity rates between states, is the greatest utility of these models when dealing with cancer.

[11] developed a multi-state model for breast cancer patients to estimate transition rates between the states in the model and later used these estimates to predict the future progression of disease for patients with given history. [12] used a multi-state model to describe invasive breast cancer progression in Canadian National Breast Screening Study and constructed progression models with and without covariates. They estimated the transition rate, the sojourn time and sensitivity of diagnostic tests for women aged 40-49 and 50-59 and the result showed that women ages 50-59 have a higher mortality rate than younger women. [7] in their work studied the significance of CA15-3 as a disease marker in monitoring and evaluating the diseases progression of breast cancer patients using multi-state Markov model. They estimated the transition intensities transition probabilities and survival time of patients in each state and found out that the elevated CA15-3 values highly associated with lower survival of the breast cancer patients. [13] used a multistate model to study prognostic factors associated with each transition in breast cancer disease. [14] provided an illustration of the application of multi-state Markov models for breast cancer progression to data from the first two rounds of the Florentine screening programme (1991-1993).

According to [15], a multi-state model (MSM) is modelling time for event data where all the individuals start in one or more states, and eventually may end up in one or several absorbing state(s). It has also been defined as a process in which an individual move through a series of states in continuous time. A longitudinal dataset or panel dataset can be observed and investigated with a MSM. A panel dataset is defined when a sample of n subjects are followed over time and multiple observations on each subject is made [15]. Some of the individuals may also be censored before they reach an absorbing state. Censored observations cause some model difficulties and therefore need to be accounted for. Censored observation is defined when the value of measurement or observation is only partially known.

[16] reported that discrete-time Markov chains have been successfully used to study treatment programs and health protocols for chronic diseases. The transition matrix, which describes the natural progression of the disease, is often estimated from a cohort observed at common intervals. However, estimating the transition matrix is often made difficult by the complex relationship between transition probabilities. In their study they summarized; Method for obtaining the maximum likelihood estimate of the transition matrix when the cycle length of the model coincides with the observation interval, the cycle length does not coincide with the observation interval and when the observation intervals are unequal. They found that a subject's medical history can then be described by moving through certain conditions over time. The discrete-time Markov model describes this movement by modelling the states at different times, which are called cycles. This model does not deal with progression between cycles and simply models the state of health at the end of each cycle. The key to the Markov model is the Markov property. These states, which given the subject's entire past, the present state depends only on the most recent state of the past. This memoryless property makes it possible to describe the model exclusively based on a single cycle transition matrix. In the construction of a homogeneous Markov chain to describe the monthly progression of HIV-infected individuals at greatest risk of developing Mycobacterium Avium Complex (MAC) infection. They stated that progression included the ability to move between three different ranges of CD4 cell counts (with and without AIDS). They concluded that the resulting matrix suggests that far too many patients will be in State 1 after six cycles. However, the present study considers the breast cancer stages and seeks to ascertain the prediction equation for the breast cancer progression in Enugu State as well as estimate the future structure of the breast cancer progression.

[15] states that when considering MSMs, it is desired to investigate the effect of different risk factors. Therefore, in a MSM, the relationship between different predictors and the outcome or variable of interest is studied. The variable of interest can be seen as the state that each individual occupies at each point in time. The transition intensities, in MSMs, provide the hazards of moving from one state to another [15].

The transition intensities can also be used to calculate the mean sojourn time in any given state.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

2.1 Transition probability matrix

Let p_{ij} be the transition probability of the system moving from state *i* to state *j*. The transition probability of moving from state *i* to state *j* at time *t* is defined as

$$p_{ij}(t) = p(X_{t+1} = j | X_t = i). \quad \textbf{(2.1)}$$

In the case where the transition probabilities are independent of time, $p_{ij}(t)$ can be written as p_{ij} and then the Markov chain is referred to as time-homogeneous.
The transition probability matrix of a multi-state process at time *t*, is an e × e matrix and can be expressed as

$$P = P(t) = \begin{bmatrix} p11(t) & p12(t) & \cdots & p1e(t) \\ p21(t) & p22(t) & \cdots & p2e(t) \\ \cdots & \cdots & \cdots & \cdots \\ pe1(t) & pe2(t) & \cdots & pee(t) \end{bmatrix} \quad \textbf{(2.2)}$$

Where E is the discrete state E = {1, 2,…, e}

The transition probability matrix (2.3), is classified as a stochastic matrix since for any row *i*, $\sum_j pij = 1$ is true [15]. Therefore, the probabilities in each of the rows of the transition probability matrix add up to one. The entries of the probability transition matrix have been defined in (2.2) and these entries define the transition or movement probabilities of individuals through states [15]. The matrix defined in (2.3), is the transition probability matrix with its elements providing the probability of being in state *j* at time *t + 1*, conditional on being in state *i* at time *t*. The transition probability matrix is time dependent and is therefore denoted as P(*t*) instead of P [15]. In time homogeneous Markov models, the dependency of *t* is omitted.
All the probabilities in the transition probability matrix must be greater than or equal to zero, that is $p_{ij} \geq 0$, ∀j, $i \epsilon \{1, ...,E\}$, and each row must sum to one
$\sum_j pij = 1$, ∀i, $j \epsilon \{1, ..., e\}$ (Mafu, 2014).

2.2 Transition intensity matrix

The intensity between two states *i* and *j*, can be defined as the rate of change of the probability p_{ij} in a small-time interval Δt[15].

The transition intensity is defined as $\lim_{\Delta t \to 0} P(X(t + \Delta t) = j \mid X(t) = i) \div \Delta t$ **(2.3)**

All possible intensities between possible states are collected in the transition intensity matrix denoted by Q [15] and given by

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

$$Q = \begin{bmatrix} q11 & q12 & \cdots & q1e \\ q21 & q22 & \cdots & q2e \\ \cdots & \cdots & \cdots & \cdots \\ qe1 & qe2 & \cdots & qee \end{bmatrix} \quad (2.4)$$

The transition intensity matrix is used to define the multi-state model and used to calculate the transition probability matrix in equation (**2.3**). The elements in each of the rows of the transition intensity matrix must also sum to zero, $\sum_j^e qij = 0$, and the off-diagonal elements of Q must be non-negative $q_{ij} \geq 0$, $i \neq j$. The diagonal elements must be negative for all values where i is not equal to j, qii = $-\sum_{i \neq j} qij$ for $i = 1, ..., e$ [15]. Therefore, the rates on the diagonal represent states that subjects remain stationary and the off-diagonal values contain rates in which the subject moves to other states [15].
The off-diagonals in this matrix are rates at which the subjects move into other states and the diagonal elements are rates at which the subjects remain in their states [15].
The transition probability matrix can be obtained by taking the matrix exponential of the scaled transition intensity matrix $P(t) = \exp(tQ)$. The exponential of a matrix C can be defined as $\exp(C) = 1 + C^2/2! + C^3/3! + ...$ using Taylor's Theorem.
The transition intensity matrix Q and transition probability matrix P can be obtained by maximizing the likelihood, L(Q). For an individual, let a series of times be $(t_1, t_2, ..., t_n)$ with corresponding states $(x_1, x_2, ..., x_n)$. A pair of successive states are observed to be i and j at time t_i and t_j.

2.3 Sojourn time

According to [15], the sojourn time of a process X in a subset of states, is an integer valued random variable which is the length of time that the process X remains in the state being occupied at time t.
The sojourn time of a continuous Markov process that is in state i is an independent and exponentially distributed random variable with mean $-\frac{1}{qii}$. The remaining elements in the i^{th} row of the transition intensity matrix is proportional to the probabilities that govern the next state after state i to which the individual makes a transition. The probability that the next transition is from state i to state j is $-\frac{qij}{qii}$[15]. The new state and the sojourn time are only dependent on state i and not on the history of the process prior to time t. Therefore, the sojourn time and the new state are independent of each other, given that the current state is state i. The mean sojourn time describes the average time period in a single stay in a state [15].

2.4 Multi-State Markov Model

A multi-state Markov model describes the process in which a patient moves through a series of states [17]. Fortunately, the *msm* package in R is one of the simpler packages that can be used to fit a multi-state model to a longitudinal dataset [17]. A longitudinal dataset consists of repeated measurements of the process at arbitrary times. The exact times of the state changes are unobserved and therefore unknown. For example, the state of a breast cancer patient may only be known when the patient consults with the oncologist.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

III. METHODOLOGY

3.1 Study Design

This study adopted a longitudinal survey design. The longitudinal survey is a research design that involves repeated observations of the same variable (e.g., people) over a short or long period, often many decades. It is often a type of observational study. The design is best suited for this study because it helps the researcher to obtain useful data for tracking particular event(s) of interest over a specified period.

3.2 Data Source

The data were collected from Enugu State Teaching Hospital Parklane, Enugu. Patient details on information such as the various stages of breast cancer recorded in the facility as well as the death status for each of the stages of breast cancer for the years under study were collected.

3.3 Population of the Study

All the records of breast cancer cases were reviewed from 2009 to 2019 with particular interest on stages of breast cancer.

3.4 Sampling Procedure

Purposive sampling technique was adopted in this study. Also, the patient's breast cancer status was categorized into (4) stages using the classification.
Stage I Breast Cancer
Stage II Breast Cancer
Stage III Breast Cancer
Stage IV Breast Cancer
The stage of breast cancer describes the size of the breast cancer and how far it has spread.

3.5 Method of Data Collection

A secondary source of data collection was adopted in this study and data was collected between 2009 and 2019. The data collected comprises of stages of breast cancer recorded at the health facility and equally death status for the various stages of breast cancer for the specified period under study.

3.6 Method of Data Analysis

The data collected over the period of 11 years (2009 – 2019) was analyzed using Markov Chain Model.

3.6.1 The Markov Chain Model

A Markov chain is a random process that goes through a transition from one state to another in state space. It has a property commonly referred to as "memoryless". The probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it. Markov chains have many uses as statistical models of real processes.
The Markov chain is a model for a finite or infinite random process sequence

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

$$X = \{X_1, X_2 \ldots X_n\} \quad (3.1)$$

Unlike the independent identical distribution (i.i.d) model that assumes the independency of a sequence of events i's, the Markov model takes into account the dependencies among the X_i's.
Consider a random process $X = \{X_t\}; t \geq 1: t = 1,2,3,\ldots,n\{X_1$, of random variables taking values in a discrete set space of stages $S = \{1, 2,3,..., s\}$ where $X_t$represents the state of the process of an individual at time t.
A random process is called a Markov Chain if the conditional probabilities between the stages at different times satisfy the Markov property: the conditional probability of future one-step-event conditioned on the entire past of the process is just conditioned on the present stage of the process. In other words, the one-step future stage depends only on the present stage.

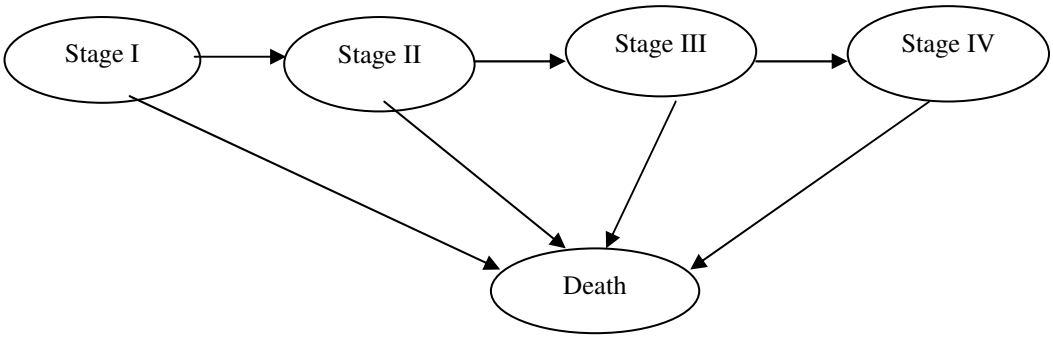

Figure 3.1 Transitions among states in the Breast Cancer Progression Markov Model

Figure 3.1 shows the transition between stage I to stage II, from stage II to stage III and from stage III to stage IV. It also shows transition from stage I, II, III, IV to Death.
A Markov chain represents a sequence of random variables $X_1, X_2, X_3,\ldots, X_n$ with the Markov property that the probability of moving to the next state depends only on the present state and not on the previous states

$$Pr(X_{n+1}=x|X_1=x_1,X_2=x_2\ldots,..,.X_n=x)_n=Pr(X_{n+1}=x|X_n=x_n) \quad (3.2)$$

If both conditional probabilities are well defined i.e. if

$$Pr\ (X_1=x_1\ldots\ldots,X_n=x_n) > 0 \quad (3.3)$$

The possible values of X_i form a countable set S called the state space of the Markov chain and are often described by a sequence of directed graphs, where the edges of graph n are labeled by the probabilities of going from one state at time n to the other states at time n+1:

$$Pr(X_{n+1}=x|X_n=x_n) \quad (3.4)$$

The same information is represented by the transition matrix from time n to time n+1.

3.7 Model Specification

The presents study considers the Markovian Model for the stages of breast cancer of the Stationary Markov Chain Model with following specification:
Let the expected number at various breast cancer events be related to the difference equation:

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

$\bar{n}_j(T+1)=\sum_{i=1}^{N} n_{ij}(T)+n_{o,j}(T+1)$, (i=1,2,3,4) (t=1,2,3…,11) **(3.5)**

Where the bar denotes the expectation (j=1,2,3,4). The above equation implies that patients in state *j* are patients who transited to state *j*.
However, the above model could be redefined by expressing the transition flow of the new entrants (the newly admitted) in terms of wastage i.e, death rate of patients.

$n_{0,j}(T+1)=W_j(T+1)$ **(3.6)**

The equation (**3.6**) becomes

$n_j(T+1)=\sum_{i=1}^{N} n_{ij}(T)+n_{0,j}(T+1)$, (i=1,2,3,4) (t=1,2,3,…,11) **(3.7)**

The common estimate of the transition probability given as;

$P_{ij}=\frac{\sum_{T=1}^{T} n_{ij}(T)}{\sum_{T=1}^{T} n_i(T)}$, *(i=1,2,3,4)* **(3.8)**

On the assumption that they are stationary over time, equation **(3.6)** becomes

$n_j(T+1)=\sum_{i=1}^{N} P_{ij}n_i(T)+W_j(T+1)$, (i=1,2,3,4) (t=1,2,3,…,11) **(3.9)**

In vector form, equation **(3.9)** becomes

$n(T+1) = n(T)P + W(T+1)$ **(3.10)**

The suitable way of using equation **(3.10)** for making forecasting is to start from the current transition counts of the breast cancer progression given the various stages of the breast cancer and then predict future transition counts for any time horizon, T, say provided that the transition probability matrix P is stationary over time or in other words is independent of time.

3.8 The Estimation of Prediction Equation for Expected number of patients in the Structure

Let $n(t) = (n_1(t), n_2(t), \cdots, n_4(t))$ be the vector of cadre sizes at the beginning of the t^{th} session.
It can be shown that

$$n(t+1) = n(t)Q \quad \textbf{(3.11)}$$

and

$$Q = P + w^T r \quad \textbf{(3.12)}$$

Where $q_{ij} = p_{ij} + w_i r_j$, i, j=1, 2, ..., 4 are elements of the matrix Q,

P = 4 X 4, overall transition probability matrix (TPM)
w = 1 X 4, row vector of wastage probabilities
r = 1 X 4, row vector of new cases probabilities

3.9 Estimation of the Sojourn time of the breast cancer progression process

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

The sojourn time in each of the stage *i* given entry in stage *j* is given as M

$$M = (I - P)^{-1} \quad (\mathbf{3.13})$$

Where, P is the (4 X 4) TPM and **I** is the identity matrix [18].
While the probability of a particular stage *j* from stage *i* can be calculated as:

$$\hat{\psi}_{ij} = \frac{\mu_{ij}}{\mu_{jj}} \quad (\mathbf{3.14})$$

Where, μ_{ij} and μ_{jj} are element of the fundamental matrix M.

IV DATA ANALYSIS AND RESULT

4. 1 The Transition Probability Matrix for the process and the wastage/Death

The transition probability matrix for the process and the wastage/Death will be extracted from **Table 13** in the Appendix 2 as follows

$$p = \begin{bmatrix} 0.3185 & 0.3268 & 0.2291 & 0.1071 \\ 0.3152 & 0.3276 & 0.2295 & 0.1103 \\ 0.2717 & 0.2922 & 0.2900 & 0.1301 \\ 0.2616 & 0.3373 & 0.2724 & 0.1142 \end{bmatrix}$$

4.2 The Transition Probability Matrix for the process

From the transition probability matrix presented above, it was found that the chances of the patient in Stage I being in Stage II are 32.68%, patients in Stage II have 22.95% chances of being in Stage III and patients in Stage III has 11.42% of being in Stage IV.

$$w = \begin{bmatrix} 0.0184 \\ 0.0174 \\ 0.0160 \\ 0.0146 \end{bmatrix}$$

4.3 The Transition Probability Matrix for the wastage

Also, the result of the wastage probability matrix reveals that patients in Stage I has 1.8% chances of death, patients in Stage II has 1.7% chances of death, patients in Stage III has 1.6% chances of death and patients in Stage IV chances of death. The slight decrease in the chances of death across the various stages can be associated with the fact that patients in Stage IV are under medication since they have been suffering from the disease for quite some time than those in Stage I who might be battling with the trauma of knowing that they have cancer.
The probability of new cases by stages of breast cancer was obtained using data presented in **Table 12** in the Appendix **1**.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

$$r=\begin{bmatrix}0.2899\\0.2770\\0.2475\\0.1856\end{bmatrix}$$

4.4 The probability of new cases

It was found from the result of the new cases probability matrix that Stage I has a 28.99% new cases rate, Stage II has a 27.77% new cases rate, Stage III has a 24.75% new cases rate while Stage IV has 18.56% new cases rate. This result indicates that Stage I has the higher probability of new cases while Stage IV has the least probability of new cases.

4.5 Estimation of Prediction Equation for the Breast Cancer Stages

The prediction equation of the structure of the breast cancer stages was computed using equation **(3.12)** as:

$$Q = P + w^T r$$

Using the R-programming 4.1.0 software to calculate the matrix multiplication

$$Q = \begin{bmatrix}0.3185 & 0.3268 & 0.2291 & 0.1071\\0.3152 & 0.3276 & 0.2295 & 0.1103\\0.2717 & 0.2922 & 0.2900 & 0.1301\\0.2616 & 0.3373 & 0.2724 & 0.1142\end{bmatrix} + \begin{bmatrix}0.0184 & 0.0174 & 0.0160 & 0.0146\end{bmatrix}^T \begin{bmatrix}0.2899\\0.2770\\0.2475\\0.1856\end{bmatrix}$$

$$Q=\begin{bmatrix}0.3238 & 0.3319 & 0.2337 & 0.1105\\0.3202 & 0.3324 & 0.2338 & 0.1135\\0.2763 & 0.2966 & 0.2940 & 0.1331\\0.2658 & 0.3413 & 0.2760 & 0.1169\end{bmatrix}$$

The future structure of the process can be obtained using equation (**3.11**):

$$n(t+1) = n(t)Q$$

Where the distribution of the total patients across the stages of breast cancer presented in **Table 11** in the Appendix 1 is used as n(t):

$$n(t+1) = n(t=11)Q$$

Table 14: Expected number of Breast Cancer Cases Structure n(t) for t= 12, 13, 14

Session	T	Stage I	Stage II	Stage III	Stage IV	n(t)
2020	12	280	307	246	112	945
2021	13	286	306	240	112	944
2022	14	287	306	240	111	944

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

The result obtained in **Table 14** found that the distribution of expected breast cancer cases for 2020 was 280 Stage I cases, 307 Stage II cases, 246 Stage III cases, and 112 stage IV cases. While for 2021 the distribution of the expected cases was obtained as 286 Stage I cases, 306 Stage II cases, 240 Stage III cases, and 112 stage IV cases. For 2022 the expected number of cases was obtained as 287 Stage I cases, 306 Stage II cases, 240 Stage III cases, and 111 stage IV cases.

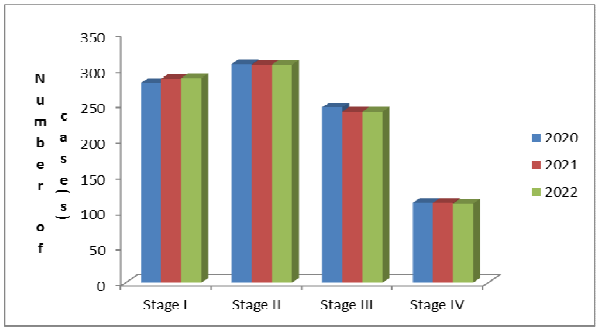

Figure 4.1: Distribution of expected breast cancer cases for 2020-2022

The result obtained in **Figure 4.1** showed that Stage II recorded the greatest number of cases followed by Stage I, Stage III and then Stage IV. This indicates that Stage II is the most reported breast cancer Stage in Enugu State while Stage IV is the least reported Stage.

4.6 Estimation of the sojourn time of the breast cancer progression

Another area of interest in this study is the estimation of the sojourn time of the breast cancer progression with regards to the four stages. The sojourn time as explained in the previous chapter is the expected time a patient is likely to spend in a particular stage of the breast cancer process. To obtain the sojourn time, we must first compute the fundamental matrix M as defined in equation (**3.13**) as:

The Fundamental Matrix of the stage of breast cancer progression in Enugu State
$M = (I - P)^{-1}$

$$M = \left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.3185 & 0.3268 & 0.2291 & 0.1071 \\ 0.3152 & 0.3276 & 0.2295 & 0.1103 \\ 0.2717 & 0.2922 & 0.2900 & 0.1301 \\ 0.2616 & 0.3373 & 0.2724 & 0.1142 \end{bmatrix} \right]^{-1}$$

$$M = \left(\hat{N}_{ij} \right) = \begin{bmatrix} Stage & \text{I} & \text{II} & \text{III} & \text{IV} \\ \text{I} & 18.53 & 18.73 & 14.61 & 6.72 \\ \text{II} & 17.55 & 19.75 & 14.63 & 6.73 \\ \text{III} & 17.53 & 18.72 & 15.72 & 6.76 \\ \text{IV} & 17.55 & 18.82 & 14.72 & 7.76 \end{bmatrix}$$

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

The result of the fundamental matrix (M) shows that a patient on stage I is expected to stay for a period of 18 years and 6 months in stage I; a patient on stage II is expected to stay for a period of 19 years and 9 months in stage II; a patient in stage III is expected to stay for a period of 15 years and 9 months in stage III; and a patient on stage IV is expected to stay for a period of 7 years and 9 months in stage IV.
However, the probability of being in a particular breast cancer stage is computed using equation (**3.14**) and obtained as:

$$\left(\hat{\psi}_{ij}\right)=\begin{bmatrix} Stage & \text{I} & \text{II} & \text{III} & \text{IV} \\ \text{I} & 1 & 0.95 & 0.93 & 0.87 \\ \text{II} & 0.94 & 1 & 0.93 & 0.87 \\ \text{III} & 0.94 & 0.95 & 1 & 0.87 \\ \text{IV} & 0.94 & 0.95 & 0.94 & 1 \end{bmatrix}$$

The matrix $(\hat{\psi}_{ij})$ shows that on the average potential patients admitted in stage I have 95% chances of being in stage II, patients in stage II have 93% chances of being in Stage III, and patients in stage III have about 87% chance of being in stage IV.

V DISCUSSION

The study looked at the progression of breast cancer in four different stages in the state of Enugu for the period 2009 -2019. The study considered the estimate of the prediction equation of the stages of breast cancer, the forecasts for the period from 2021 to 2022, as well as the estimated time of stay at each stage. This is in line with the study by Taghipour et al. (2013) who focused on breast cancer transitional states for Canadian women of reproductive age and Grover et al (2018) who considered the transition intensities and transition probabilities of chronic disease patients in each state.
The results of the study showed that the odds of a stage I breast cancer patient being stage II are 32.68 per cent, stage II patients have a 22.95 per cent chance of being stage III, and Stage III patients have 11.42% of being stage IV. The probability of wastage shows that stage I patients have a 1.8% chance of death, stage II patients have a 1.7% chance of death, stage III patients have a probability of death 1.6% and stage IV patients have a chance of death. The slight decrease in the likelihood of dying at different stages may be related to the fact that stage IV patients are on therapy, as they have had the disease for a longer time than stage I patients who may be struggling with the trauma of know they have cancer. The probability matrix for newly admitted patients in the four phases showed that stage I had new cases by 28.99%, stage II new cases by 27.77%, stage III new cases by 24.75%, and stage IV of 18.56% of new cases. This result shows that stage I has the highest probability of new cases, while stage IV has the lowest probability of new cases.
The distribution of breast cancer cases forecast for 2020 was 280 cases in Stage I, 307 cases in Stage II, 246 cases in Stage III and 112 cases in Stage IV. The forecast for 2021 was 286 cases in Stage I, 306 cases in Stage II, 240 cases in Stage III and 112 cases in Stage IV, while the forecast for 2022 was 287 cases in Stage I, 306 cases in Stage II, 240 cases in Stage III and 111 cases in Stage IV. Furthermore, it was found that Stage II records the majority of cases, followed by Stage I, Stage III and then Stage IV. This finding shows that Stage II is the most commonly reported stage of breast cancer, while Stage IV is the most common less reported stage. The analysis of the sojourn time showed that a Stage I patient is expected to remain in Stage I for 18 years and 6 months. A Stage II patient is expected to remain in Stage II for 19 years and 9 months. A Stage III patient is expected to remain in stage III for 15 years and 9 months while a Stage IV patient is expected to remain in Stage IV for 7 years and 9 months. It was also found that potential patients admitted in Stage I have a 95% chance of progressing to Stage II, patients at Stage II have a 93% chance of progressing to Stage III, and patients at Stage III have approximately 87% chance of progressing to Stage IV.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

VI. CONCLUSION

This study aimed to determine the progression of breast cancer stages using multi-state Markov model. Relevant related literatures on multi-state models, Markov process, transition probability matrix, transition intensity matrix, sojourn time, multi-state Markov model and disease progression models were reviewed. In the study, a longitudinal survey design was adopted using purposive sampling technique. The study was further guided by four stages of breast cancer progression, analyzed using the Markov chain model and the findings of the analysis were presented.

The findings deduced from the analysis were given as follows:

I. The transition probability matrix for assessing the progression of breast cancer found that the chances of the patient in stage I being in Stage II are 32.68%, patients in Stage II has 22.95% chances of being in Stage III and patients in Stage III has 11.42% of being in Stage IV.
II. The waste probability shows that patients in stage I have a probability of death of 1.8%, patients in stage II have a probability of death of 1.7%, patients in stage III have a probability of death of 1.6% and patients in stage IV have a probability of death. The slight decrease in the likelihood of dying at different stages may be related to the fact that stage IV patients are on medication, as they have had the disease for a longer time than those in stage I who may be struggling with the trauma of knowing that they have cancer.
III. The probability matrix for newly admitted patients in the four stages showed that stage I had new cases of 28.99%, stage II new cases of 27.77%, stage III new cases of 24.75% and stage IV of 18.56%. Rate of new cases. This result shows that stage I has the higher probability of new cases, while stage IV has the lowest probability of new cases.
IV. The distribution of breast cancer cases forecast for 2020 was 280 cases in stage I, 307 cases in stage II, 246 cases in stage III and 112 cases in stage IV. Forecast for 2021 was 286 cases in stage I, 306 cases in stage II, 240 cases in stage III, and 112 cases in stage IV while the forecast for 2022 was 287 cases in stage I, 306 cases in stage II, 240 cases in stage III, and 111 cases in stage IV.
V. Also, stage II was found to record most cases, followed by stage I, stage III, and then stage IV. This result shows that stage II is the most commonly reported stage of breast cancer in the state of Enugu, while stage IV is the most common least reported stage.
VI. The result of the fundamental matrix showed that a stage I patient is expected to remain in stage I for 18 years and 6 months. A stage II patient is expected to remain in stage II for 19 years and 9 months. A stage III patient is expected to remain in stage III for 15 years and 9 months while a Stage IV patient is expected to remain in stage IV for 7 years and 9 months.
VII. It was also found that, on average, potential patients admitted in stage I have 95% of the chances of progressing to stage II, patients in stage II have 93% chances of progressing to stage III and stage III patients have about 87% chances of progressing to stage IV.

REFERENCES

[1] Komen, S.G. (2017) [Online].*http://ww5.komen.org/BreastCancer/WhatisBreastCancer.html* [2017, March 13].

[2] Sharma G.N., Dave R. and Sharma K.K. (2010). Various Types and Management of Breast Cancer: An Overview. *Journal of the Advanced Pharmaceutical Technology & Research*, 1(2): 109-126.

[3] Tiengo J & Peltzer K. (2011). Knowledge Attitude and Practice of Breast Cancer Examination among Women attending a health facility in Gaborone, Botswana. *Gender &Behaviour*; 9(1)

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

[4] Grayson, M. (2012). Breast cancer. *Nature* 485, S49. https://doi.org/10.1038/485S49a

[5] National Cancer Institute.2018. [Online]. https://www.cancer.gov/publications/dictionaries/cancerterms/def/cancer [2018, April 02].

[6] Meira-Machado Luis, Una-Alvarez Jacobo and Andersonen Per K. (2009). Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res.* 2009 April; 18(2): 19-222.

[7] Grover Gurprit, Swain Prafulla Kumar, Goel Kormal, Singh Vikas, (2018). Multistate Markov Modelling for Disease Progression of Breast Cancer Patients Based on CA15-3 Marker, *Thailand Statistician* 16(2), 129-139, 2018.

[8] Ruiz-Castro Juan Eloy and Zenga Mariangela (2019). A general piecewise multi-state survival model: application to breast cancer. *Statistical Methods & Applications*, 1-31, 2019.

[9] Amorim, A.P., de Uña-Álvarez, J. and Meira-Machado, L. (2011). Pre-smoothing the Transition Probabilities in the Illness-Death Model. *Statistics & Probability Letters*, 81(7): 797-806.

[10] Hougaard, P. (1999). Multi-state Models: A Review. *Lifetime Data Analysis*, 5(3): 239-264.

[11] Putter Hein, Hage Jos, Bock Geertruida H., Elgalta Rachid and Velde Cornelis (2006). Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal: Journal of Mathematical Methods in Bioscience* 48 (30, 366-380, 2006.

[12] Taghipour S., Banjevic D., Miller AB., Montgomery N., Jardine AKS. and Harvey BJ. (2013). Parameter estimates for invasive breast cancer progression in the Canadian National Breast Screening Study. *British Journal of Cancer* 108(3), 542-548, 2013.

[13] Broet Philippe, Rochefordiere Anne, Scholl Susan M., Fourquetalain, Rycke Yann, Pouillart Pierre, Mosseri Veronque and Asselain Bernard (1999). Analyzing prognostic factors in breast cancer using multistate model. *Breast Cancer Research and Treatment* 54, 83-89, 1999.

[14] Ventura Leonardo, Carreras Giulia, Puliti Donella, Paci Eugenio, Zappa Marco and Miccinesi Guido (2014). Comparison of multi-state Markov models for cancer progression with different procedures estimation. An application to breast cancer. *Epidemiology, Biostatistics and Public Health* 11 (1), 2014.

[15] Mafu T.J. (2014). Modelling of Multi-State Panel Data: The Importance of the Model Assumptions. Doctor's dissertation. Stellenbosch: University of Stellenbosch.

[16] Craig, B. A. and Sendi, P. (2002). Estimation of the Transition Matrix of a Discrete-Time Markov Chain. *Health Economics*, 11: 33-42.

[17] Jackson, C. (201)1. Multi-state modelling with R: the msm package for R. *Journal of Statistical Software*, 38(8): 1-29.

[18] Ogbogbo, G. O., Ebuh, G. U., Aronu, C. O. (2013). Determining the Expected Length of Stay of Academic Staff in various Grade Levels in a Polytechnic Institution in Nigeria using a Markov Chain Approach. *American Journal of Computational and Applied Mathematics*, 3(4): 225-232.

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

APPENDICES

APPENDIX 1: TABLE FOR SUMMARY OF NUMBER OF NEW CASES

Table 1. Summary of the number of new cases/ admitted cases of breast cancer from 2009-2019

Year	Stage I	Stage II	Stage III	Stage IV
2009	70	12	17	10
2010	16	17	13	12
2011	73	38	69	11
2012	81	35	40	49
2013	28	32	34	17
2014	19	46	81	24
2015	12	90	16	42
2016	17	14	19	14
2017	22	31	15	10
2018	27	44	15	34
2019	5	8	7	9
Total	403	385	344	258

Source: Enugu State Teaching Hospital Parklane, Enugu, 2019

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

APPENDIX 2: TABLE FOR TRANSITION PROBABILITIES OF BREAST CANCER

Table 2: Transition Probabilities of breast cancer into various stages in Enugu for t =1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

$N_{0j}(t)i$						
T	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Death**	$n_i(t)$
	95(0.3493)	99(0.3640)	48(0.1765)	18(0.0662)	12(0.0441)	272
	110(0.4120)	90(0.3371)	50(0.1873)	15(0.-562)	2(0.0075)	267
	95(0.3654)	86(0.3308)	50(0.1923)	20(0.0769)	9(0.0346)	260
	60(0.2198)	80(0.2930)	70(0.2564)	50(0.1832)	13(0.0476)	273
	80(0.3361)	90(0.3782)	50(0.2101)	17(0.0714)	1(0.0042)	238
	90(0.3488)	80(0.3101)	47(0.1822)	40(0.1550)	1(0.0039)	258
	100(0.3876)	90(0.3488)	54(0.2093)	10(0.0388)	4(0.0155)	258
	56(0.2424)	65(0.2814)	78(0.3377)	30(0.1299)	2(0.0087)	231
	56(0.2414)	65(0.2802)	78(0.3362)	30(0.1293)	3(0.0129)	232
	83(0.3444)	75(0.3112)	50(0.2075)	32(0.1328)	1(0.0004)	241
Stage I	58(0.2397)	86(0.3554)	60(0.2479)	35(0.1446)	3(0.0124)	242
Total	883(0.3185)	906(0.3268)	635(0.2291)	297(0.1071)	51(0.0184)	2772
	100(0.3703)	96(0.3556)	50(0.1852)	20(0.07407)	4(0.01482	270
	80(0.2909)	120(0.4364)	55(0.2000)	15(0.0545)	5(0.0182)	275
	75(0.2885)	90(0.3462)	50(0.1923)	37(0.1423)	8(0.0308)	260
	60(0.2290)	70(0.2672)	65(0.2481)	57(0.2176)	10(0.0382)	262
	90(0.3475)	100(0.3861)	55(0.2124)	10(0.0386)	4(0.0154)	259
	70(0.2703)	80(0.3089)	60(0.2317)	47(0.1815)	2(0.0007)	259
	101(0.3900)	80(0.3089)	64(0.2471)	9(0.0347)	5(0.0193)	259
Stage II						

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

	85(0.3320)	55(0.2148)	86(0.3359)	24(0.0938)	6(0.0234)	256
	70(0.2789)	80(0.3187)	60(0.2390)	40(0.1594)	1(0.0004)	251
	80(0.3376)	90(0.3797)	50(0.2110)	16(0.0675)	1(0.0042)	237
	75(0.3363)	60(0.2691)	50(0.2242)	35(0.1570)	3(0.0135)	223
Total	886(0.3152)	921(0.3276)	645(0.2295)	310(0.1103)	49(0.0174)	2811
	70(0.2527)	50(0.1805)	95(0.0340)	55(0.1986)	7(0.0253)	277
	65(0.2471)	90(0.3422)	70(0.2662)	30(0.1141)	8(0.0304)	263
	70(0.2545)	95(0.3455)	80(0.2909)	24(0.0873)	6(0.0218)	275
	50(0.2193)	75(0.3289)	80(0.3509)	21(0.0921)	2(0.0088)	228
	60(0.2449)	70(0.2857)	90(0.3673)	23(0.0939)	2(0.0082)	245
	50(0.2193)	65(0.2851)	75(0.3289)	37(0.1623)	1(0.0044)	228
	70(0.3070)	50(0.2193)	60(0.2632)	43(0.1886)	5(0.0219)	228
	75(0.3289)	60(0.2632)	50(0.2193)	35(0.1535)	8(0.0351)	228
	80(0.3376)	90(0.3797)	50(0.2110)	16(0.0675)	1(0.0042)	237
	56(0.2424)	65(0.2814)	78(0.3377)	30(0.1299)	2(0.0087)	231
Stage III	83(0.3416)	74(0.3045)	50(0.2058)	35(0.1440)	1(0.0041)	243
Total	729(0.2717)	784(0.2922)	778(0.2900)	349(0.1301)	43(0.0160)	2683
	76(0.2764)	95(0.3455)	30(0.1091)	64(0.2327)	10(0.0364)	275
	60(0.2400)	84(0.3360)	85(0.3400)	20(0.0800)	1(0.0040)	250
	55(0.2200)	85(0.3400)	80(0.3200)	24(0.0960)	6(0.0240)	250
	60(0.2400)	70(0.2800)	85(0.3400)	25(0.1000)	10(0.0400)	250
	50(0.2041)	90(0.3673)	85(0.3469)	19(0.0776)	1(0.0041)	245
	50(0.2232)	70(0.3125)	60(0.2679)	43(0.1920)	1(0.0045)	224
	50(0.2232)	60(0.2679)	70(0.3125)	42(0.1875)	2(0.0089)	224
	50(0.2008)	90(0.3614)	85(0.3414)	19(0.0763)	5(0.0201)	249
Stage IV						

MDC Journals.
International Journal of Scientific and Allied Research
[IJSAR]

IJSAR Journal of Mathematics and Applied Statistics (IJSAR-JMAS)
ISSN: 2408-7637
Volume 8, Issue 3 (September, 2021), 172- 187
http://www.mdcjournals.org/ijsar-jmas...html

	90(0.3782) 80(0.3361) 80(0.3376)	80(0.3361) 90(0.3782) 90(0.3797)	50(0.2101) 50(0.2101) 50(0.2110)	17(0.0714) 17(0.0714) 16(0.0675)	1(0.0042) 1(0.0042) 1(0.0042)	238 238 237
Total	701(0.2616)	904(0.3373)	730(0.2724)	306(0.1142)	39(0.0146)	2680