

# A Simple Stochastic Stomach Cancer Model with Application

Josphat Mutwiri Ikiao<sup>1</sup>, Nyongesa Kennedy<sup>2</sup>, Robert Muriungi Gitunga<sup>1</sup>

<sup>1</sup>Department of Mathematics, Meru University of Science and Technology, Meru, Kenya

<sup>2</sup>Department of Mathematics, Masinde Murilo University of Science and Technology, Kakamega, Kenya

## Email address:

Josphatm2000@yahoo.com (J. M. Ikiao), knyongesa@hotmail.com (N. Kennedy), jgitunga2005@yahoo.com (R. M. Gitunga)

## To cite this article:

Josphat Mutwiri Ikiao, Nyongesa Kennedy, Robert Muriungi Gitunga. A Simple Stochastic Stomach Cancer Model with Application. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 3, 2018, pp. 112-117. doi: 10.11648/j.ajtas.20180703.13

**Received:** December 3, 2017; **Accepted:** December 12, 2017; **Published:** April 11, 2018

---

**Abstract:** Survival analysis majors mainly on estimation of time taken before an event of interest takes place. Time taken before an event of interest takes place is a random process that takes shape overtime. Stochastic processes theory is therefore very crucial in analysis of survival data. The study employed markov chain theory in developing a simple stochastic stomach cancer model. The model is depicted with a state diagram and a stochastic matrix. The model was applied to stomach cancer data obtained from Meru Hospice. Transition probability theory was used in determining transition probabilities. The entries of the stochastic matrix T were estimated using the Aalen-Johansen estimators. The time taken for all the people under the study to transit to death was estimated using the limiting matrix.

**Keywords:** Stochastic Stomach Cancer Model, State Diagram, Stochastic Matrix, Transition Probabilities, Limiting Matrix

---

## 1. Introduction

In this section, background of the study was discussed by outlining the history and impacts of cancer in the world in terms of morbidity and mortality. Previous studies by other researchers were also reviewed. Study problem was then formulated by stating the study objectives. The scope of the study was then discussed and the assumptions of the study outlined.

### Background of the study

American cancer society report 2012 in [1] defines cancer as a group of diseases characterized by the uncontrolled spread of abnormal cells, which if not controlled, results in the death of the person affected. Cancer is caused by both external factors such as tobacco intake, infectious organisms, chemicals and radiations and internal factors such as inherited mutations, hormones and mutations that occur from metabolism. Sirpa Heinavaara in [2] argues that survival of cancer patients is known to depend on prognostic factors such as age of the patient at diagnosis, gender of the patient and site of the cancer. According to the Kenya cancer statistics and National strategy report in [3], cancer remains a global health problem and it is estimated that globally, cancer causes more deaths than Human Immunodeficiency Virus,

Tuberculosis and Malaria combined. World Health Organization report in [4] show that cancer accounted for 7.9 million in the year 2009, which is 13% of all deaths worldwide. The annual mortality is attributed to main types of cancer which include lung cancer ( 1.3 million deaths), stomach cancer(803,000 deaths), colorectal cancer (639000 deaths), liver cancer( 610000 deaths) and oesophageal cancer (380000 deaths). According to Kenya National cancer control strategy in [5], the burden of cancer is projected to continue rising, with an estimate of 15.5 million new infections and 12 million deaths by the year 2030. Kenya cancer statistics and National strategy report in [6] ranks cancer third in terms of morbidity and mortality, accounting for 7% of all deaths annually. There are an estimated 39000 new cases of cancer each year in Kenya and 27000 deaths annually. 60% of Kenyans affected by cancer are below 70 years old and therefore productive economically. In women, breast cancer leads in morbidity and mortality followed by cervical cancer while in men prostate cancer leads in morbidity and mortality. Meru Hospice in [7] ranks stomach cancer as the highest cause of morbidity and mortality in Meru, followed by breast cancer. Cervical cancer is ranked third and prostate cancer fourth. At the moment, there are approximately 500 cancer cases reported and an average of 5 new cases reported each

day. According to Tiwari et al. in [8], cancer has a great negative effect on the economies of many countries globally, Kenya included. Thus, accurate estimation of the number of deaths caused by cancer is very crucial as it helps the government in planning, resource allocation and communication. Inaccurate predictions result in inefficient planning and budgeting resulting in ineffective government services. Kenya National cancer control strategy in [9] classifies cancer research in Kenya as non commensurate with the magnitude of the problem. This is because of the inadequate funding and training facilities in cancer research. One of the intervention measures suggested in the strategy is to strengthen research through cancer data collection, analysis, interpretation and dissemination. Besides, there is no statistical research on stomach cancer in Meru County or if any, it is little known of research, despite stomach cancer leading in morbidity and mortality in Meru County.

Various models have been used in the analysis of survival data. Moghimi et al. in [10] used Cox regression model and compared it with weibull, exponential and lognormal models to evaluate prognostic factors affecting survival of patients with stomach cancer. Age of the patients at the time of diagnosis and grade of tumour were used as the independent prognosis factors. Akaike information criterion was used to determine the best model. Results of the data analysis showed that Cox and parametric models were approximately similar. However, according to Akaike Information Criterion, weibull and Exponential models were the most favourable for survival analysis. Chun et al. in [11] used Multi-task logistic regression model to study patient specific cancer survival distributions. The study revealed that Multi-task logistic regression method predictions were more accurate than popular survival models like Cox and Aalen regression models. The study further revealed that using patient specific attributes can reduce the prediction error on survival time by as much as 20% when compared to using cancer site or stage only. Vallinayagam et al. in [12] compared the performance of different parametric models like Exponential, Weibull, Gompertz, Lognormal and Log-logistic using breast cancer patients. From the analysis results, Lognormal model recorded the lowest deviance, followed by Log-logistic and the Weibull models. Farid et al. in [13] did a review of methodological approaches that can be used to deal with violation of proportionality assumption using Cox proportional hazards model in survival analysis. The four modifications suggested in the study were stratification of covariates, partitioning of time axis, modelling time dependence of the coefficients and lastly to allow the data to select the functional form of the time dependence, for instance by using splines. Limitation of the Cox Proportional hazards model assumption of proportionality of the hazard function, which if violated renders the model inapplicable, was emphasized in the study. The study suggested the use of the four modifications or use of a different model if the Proportionality assumption is violated. Uku et al. in [14] compared the efficiency of a mixture of two distributions with a single distribution in modelling heterogeneous survival data. Single distributions like Gamma, Weibull and Exponential were compared with mixture distributions of

Exponential-Gamma, Exponential-Weibull and Gamma-Weibull. Mixture distribution of any two distributions provided a better estimate than any single distribution alone. Among the three mixture distributions, Gamma-Weibull mixture distribution gave the best model. Duke et al. In [15] used Cox Proportional model to forecast survival time for cancer patients. Weibull model was also applied to the data of patients diagnosed with lung and bronchus cancer in light of censoring. The researchers observed that survival time is non-negative and consequently its distribution is positively skewed. They argued that for positively skewed distribution, the option is to use either Exponential or Weibull model. However, they noted that Weibull model is more suitable since it provides more flexibility than only one parameter of the Exponential distribution which results in the assumption that all patients are equally likely to die regardless of how many years they have survived. Duke et al. In [16] used Cox Proportional hazards model to estimate the survival time of lung and bronchus cancer patients. To cater for the unobserved heterogeneity, a weibull mixture model was applied to the data. Age, gender, race and registries were used as independent prognostic variables. The results showed that age, gender and registry significantly affect individual survival time. The study concluded that Weibull distribution is more suitable than other long tailed distributions such as Log-logistic and Expo-power distributions. Henry A, Glick in [17] described Markov models as recursive decision trees that are used for modelling conditions that have events that May occur repeatedly overtime, or for modelling predictable events that may occur overtime like screening for a disease after some fixed intervals. The study described various ways of estimating transition probabilities. He argued that if available data are hazard rates per unit time, they can be translated into probabilities by using the formula  $P_{ij}^{(t)} = 1 - e^{-R_{ijt}}$  where  $P_{ij}^{(t)}$  is the probability of moving from state  $i$  at the beginning of a period  $t$  to state  $j$  at the beginning of a period  $t+1$ ;  $R_{ijt}$  is the instantaneous hazard rate per period and  $t$  is the length of the period. Juergen Jung in [18] estimated transition probabilities for individual health status as a function of observables characteristics. The study employed three methods. One of the methods used is the counting method in which the transition from state  $h$  to state  $j$  is estimated by  $P_{hj}^{(t)} = P(Y_t = j / Y_{t-1} = h) = \frac{N_{hj}}{\sum_{n=1}^N \sum_t I(y_{it-1}=h)}$  where  $P_{hj}^{(t)}$  is the average transition probability from state  $h$  to state  $j$ ,  $N_{hj}$  is the realization of a particular transition from state  $h$  to state  $j$ . The second method used in the study is the ordered logit and ordered probit regression models. The third method used is the semi parametric Cox Proportional hazards model. Abner et al. in [19] described markov chain process as a stochastic process that describes the movement of an individual through a finite number of states. The study identified various methods of analyzing time to event data, including Cox Proportional hazards model and Kaplan Meier estimation. Heggland in [20] argued that in multistate models, the past and the future are independent given the present. The researcher argues that multistate models are often assumed to be markovian models.

The study asserts that a process  $X(t)$  is said to be markov if  $P(X(t) = h/X(s) = g, \mathcal{F}_s) = P(X(t) = h/X(s) = g)$ , where  $\mathcal{F}_s$  is the history of the process up to time  $s$  (Information about the earlier transitions of the process). The study further argues that estimation of transition probabilities is done by solving the Chapman –Kolmogorov forward equations. The study defined  $N_{gh}(t)$  to be the number of individuals observed to move from state  $g$  to state  $h$  within an interval  $(0, t)$ . The study estimated  $N_{12}(t)$ ,  $N_{13}(t)$  and  $N_{23}$  while  $N_{21}(t)$ ,  $N_{31}(t)$  and  $N_{32}(t)$  were zero for all values of  $t$ .  $N_{21}(t)$  was zero because no transitions from state 2 to state 1 were witnessed while  $N_{31}(t)$  and  $N_{32}(t)$  were zero because state three is an absorbing state. The study then set  $N_1(t) = N_{12}(t) + N_{13}(t)$  while  $Y_1(t)$  and  $Y_2(t)$  were the number of healthy and diseased individuals respectively before time  $t$ . Solutions to Chapman Kolmogorov forward equations gave the estimators of transition probabilities as:  $\hat{P}_{11}^{(t)}(s, t) = \prod_{s < T_j \leq t} (1 - \frac{\Delta N_{1(T_j)}}{Y_1(T_j)})$ ,  $\hat{P}_{22}^{(t)}(s, t) = \prod_{s < T_j \leq t} (1 - \frac{\Delta N_{23(T_j)}}{Y_2(T_j)})$  and  $\hat{P}_{12}^{(t)}(s, t) = \sum_{s < T_j \leq t} (s, T_{j-1}) \hat{A}_{12}(T_j) \hat{P}_{22}^{(t)}(T_j, t)$  where  $\hat{A}_{12}(T_j) = \frac{\Delta N_{12}(T_j)}{Y_1(T_j)}$ . Snijders in [21] discussed different methods of estimating transition probabilities such as Kaplan-Meier estimator, Nelson-Aalen estimator and Aalen-Johansen estimator as the non parametric estimators and Cox Proportional hazards model as the semi-parametric estimator. Kaplan-Meier is used to estimate the survival function  $S(t)$ . Kaplan-Meier concept is generalized to the Aalen-Johansen estimator for the multi-state models. The study assumes an ordered list  $0 < t_1 < t_2 < \dots < t_N$ .  $S(t)$  is split into a product  $S(t) = P(T \geq t_j) = P(T \geq t_j/T \geq t_{j-1})P(T \geq t_{j-1})$ .  $d_j$  is defined as the number of observed events at time  $t_j$  and  $n_j$  as the risk set. The study then defines the estimate of  $S(t)$  as  $\hat{S}(t) = \prod_{i=1}^j (1 - \frac{d_i}{n_i})$ . The study introduces Cox

Proportional hazards model which gives an insight on how other factors such as age or sex influence the hazard rate. Luis et al. in [22] reviews modelling approaches for multi-state models and focuses on estimation of quantities like the transition probabilities and survival probabilities. The study states that multi-state models are characterized through transition probabilities between state  $h$  to state  $j$ ;  $P_{hj}(s, t) = P(X(t) = j/X(s) = h, \mathcal{H}_s)$  for  $h, j \in S, s, t \in T, s \leq t$  or through transition intensities  $\alpha_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t+\Delta t)}{\Delta t}$  which represent the instantaneous hazard of moving to state  $j$  conditional on occupying state  $h$ . The researcher argues that in markov models, transition probabilities can be estimated from the intensities by solving the forward Kolmogorov equations. The transition probabilities have the following expressions;  $P_{11}(s, t) = e^{-(A_{12}(s, t) + A_{13}(s, t))}$  (1),  $P_{22}(s, t) = e^{-A_{23}(s, t)}$  (2),  $P_{12}(s, t) = \int_s^t P_{11}(s, u) \alpha_{12}(u) P_{22}(s, t) du$  (3). The transition probabilities in equations (1), (2) and (3) can then be estimated using Kaplan-Meier estimators via the non-parametric Aalen Johansen model as follows:  $\hat{P}_{11}(s, t) = \prod_{s < t_k \leq t} (1 - \frac{d_{12k} + d_{13k}}{n_{1k}})$ ,  $\hat{P}_{22}(s, t) = \prod_{s < t_k \leq t} (1 - \frac{d_{23k}}{n_{2k}})$  and  $\hat{P}_{12}(s, t) = \sum_{s < t_k \leq t} \hat{P}_{11}(s, t_{k-1}) \frac{d_{13k}}{n_{1k}} \hat{P}_{22}(t_k, t)$ .

## Scope of the study

Simple stochastic stomach cancer model was derived and stomach cancer data from Meru County fitted on the model. This was analyzed using R software by use of the TPmsm package.

The main aim of the study was to describe the cancer progression among patients and determine the probability of moving from one state of cancer to another. A state diagram was designed to describe the movement of cancer patients from one state to another. Stochastic matrix  $T$  was designed and its entries computed by fitting the data on the model. The study assumed that future states depended only on the current state and not on past events and that all the individuals under study began from an initial state ('Healthy state').

## 2. Materials and Methods

### 2.1. Introduction

In this section, model development was discussed. Data collection and analysis was also discussed.

### 2.2. Simple Stochastic Stomach Cancer Model

The study employed the following methods: Markov chain theory was used to describe the movement of a patient through stomach cancer states. The assumption was that movement to the next state depended on the current state occupied and that all the individuals under study began from the same initial state ('Healthy state'). This was depicted with a state diagram. Transition probability theory was used in designing the stochastic matrix and deriving the transitional probabilities. R software was used in analysis of the data by use of the TPmsm package. Stomach cancer data was obtained from the Meru Hospice, who are the custodians of cancer data in Meru. Data for 274 stomach cancer patients was used in the study.

Stomach cancer model was developed by first identifying stomach cancer stages and then deriving transition probabilities. Stomach cancer has mainly three stages (Tumour grades); that is grade 1, grade 2 and grade 3. The model was limited to grade 3 stage because of the nature of secondary data available. Consequently, the model has three states namely:

H-State in which an individual is free from stomach cancer, S-state in which the individual is suffering from stomach cancer and D-State in which an individual dies of stomach cancer. This was formalized in the following transition diagram:

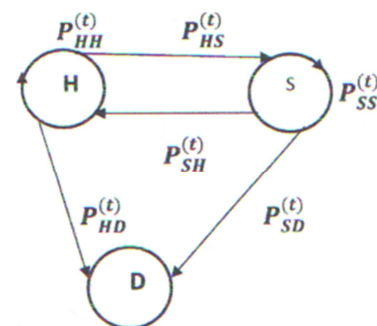


Figure 1. Model state diagram.

Where:

$P_{HS}^{(t)}$  = The probability of moving from health state to sickness state

$P_{SD}^{(t)}$  = The probability of moving from sickness state to death state

$P_{SH}^{(t)}$  = The probability of moving from sickness state to health state

$P_{HD}^{(t)}$  = The probability of moving from health state to death state

$P_{SS}^{(t)}$  = The probability of being in the sickness state and remaining there for sometime

$P_{HH}^{(t)}$  = The probability of being in the health state and remaining there for sometime

Transition probability matrix T takes the form:

$$T = \begin{bmatrix} P_{HH}^{(t)} & P_{HS}^{(t)} & P_{HD}^{(t)} \\ P_{SH}^{(t)} & P_{SS}^{(t)} & P_{SD}^{(t)} \\ P_{DH}^{(t)} & P_{DS}^{(t)} & P_{DD}^{(t)} \end{bmatrix}$$

The transition probability matrix T has entries that represent the probabilities of moving from one state of stomach cancer to another. Notice from figure 2.2.2 that D state is an absorbing state, therefore  $P_{DD}^{(t)} = 1$ . Therefore T takes the form as shown below:

$$T = \begin{bmatrix} P_{HH}^{(t)} & P_{HS}^{(t)} & P_{HD}^{(t)} \\ P_{SH}^{(t)} & P_{SS}^{(t)} & P_{SD}^{(t)} \\ 0 & 0 & 1 \end{bmatrix}$$

The transition matrix shows the actual values of the entries of the third row; i.e actual entries of  $P_{DH}^{(t)}$ ,  $P_{DS}^{(t)}$  and  $P_{DD}^{(t)}$ .

Next, the remaining  $P_{ij}^{(t)}$  for  $i = H, S$  and  $j = H, S, D$  was computed. If an individual is in state  $i$  at time  $t$ , then the probability of moving from state  $i$  to state  $j$  in a small interval  $(t, t + \Delta t)$  is given by:

$$P_{ij}^{(t)} = \lim_{\Delta t \rightarrow 0} \frac{P\{(t, t+\Delta t)/T \geq t\}}{\Delta t} \quad (1)$$

$$\text{But } f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t+\Delta t)}{\Delta t} \text{ and } S(t) = P(T \geq t)$$

$$\begin{aligned} \text{Therefore } P_{ij}^{(t)} &= \lim_{\Delta t \rightarrow 0} \frac{P\{(t \leq T < t+\Delta t)/T \geq t\}}{\Delta t} = \\ \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t+\Delta t)}{\Delta t P(T \geq t)} &= \frac{f(t)}{S(t)} = -\frac{\frac{d}{dt}S(t)}{S(t)} \end{aligned} \quad (2)$$

$$P_{ij}^{(t)} = -\frac{d}{dt} \log S(t) \quad (3)$$

$$= -\frac{d}{dt} \log S(t) = \frac{d}{dt} \log \frac{1}{S(t)} \quad (4)$$

The estimate of  $S(t)$  which is  $\hat{S}(t)$  was used to estimate  $\hat{P}_{ij}^{(t)}$  in equation 4 above.

$$\hat{P}_{ij}^{(t)} = \frac{d}{dt} \log \frac{1}{\hat{S}(t)} \quad (5)$$

The estimate of  $\hat{P}_{ij}^{(t)}$  is given by:

$$\hat{P}_{ij}^{(t)} = \frac{d}{dt} \log \frac{1}{S(t)} \quad (6)$$

From equation 6,  $\hat{P}_{ij}^{(t)}$  was estimated by use of Aalen-Johansen estimators as shown by Luis et al. (2010) to be:

$$\hat{P}_{ij}^{(t)} = \hat{P}_{11}^{(t)} = P(x(t) = 1/x(0) = 1 = \prod_{t_j \leq t} (1 - \frac{d_{12t_j} + d_{13t_j}}{n_{1t_j}}) \quad (7)$$

Where  $\hat{P}_{ij}^{(t)}$  is the probability of moving from state  $i$  to state  $j$  for  $i = 1$  and  $j = 1$  (That is the probability of remaining in state 1),  $d_{12t_j}$  is the number of transitions from state 1 to state 2,  $d_{13t_j}$  is the number of moving from state 1 to state 3 and  $n_{1t_j}$  is the number of people in the health state.

$$\hat{P}_{ij}^{(t)} = \hat{P}_{22}^{(t)} = P(x(t) = 2/x(t-1) = 2 = \prod_{t_j \leq t} (1 - \frac{d_{21t_j} + d_{23t_j}}{n_{2t_j}}) = \prod_{t_j \leq t} \frac{d_{23t_j}}{n_{2t_j}} \quad (8)$$

This is because  $d_{21t_j} = 0$ . Where  $\hat{P}_{ij}^{(t)}$  is the probability of moving from state  $i$  to state  $j$  for  $i = 2$  and  $j = 2$  ( That is the probability of remaining in state 2 for sometime),  $d_{21t_j}$  is the number of transitions from state 2 to state 1,  $d_{23t_j}$  is the number of transitions from state 2 to state 3 and  $n_{2t_j}$  is the total number of people sick.

$$\hat{P}_{ij}^{(t)} = \hat{P}_{12}^{(t)} = P(x(t) = 2/x(0) = 1) = \sum_{t_j \leq t} \hat{P}_{11}(t) \frac{d_{12t_j}}{n_{1t_j}} \hat{P}_{22}(t) \quad (9)$$

Where  $\hat{P}_{ij}^{(t)}$  is the probability of moving from state  $i$  to state  $j$  for  $i = 1$  and  $j = 2$ ,  $d_{12t_j}$  is the number of transitions from state 1 to state 2 and  $n_{1t_j}$  is the number of people healthy.

Next, the derived results in equations 7, 8 and 9 were applied on Meru County data.

### 3. Results

#### 3.1. Introduction

In this section, equations 7,8 and 9 were applied on our data to compute the entries of T. The section also discusses the results.

#### 3.2. Transition Matrix

Equations 7, 8 and 9 were used to generate the entries of T, that is the transition probabilities from state 1 to state  $j = 1, 2, 3$  and transitions from state 2 to state  $j = 1, 2, 3$  as provided below:

$$\hat{P}_{11}^{(t)} = \prod_{t_j \leq t} (1 - \frac{d_{12t_j} + d_{13t_j}}{n_{1t_j}}) = 0.630 \quad (10)$$

$$\hat{P}_{22}^{(t)} = \prod_{t_j \leq t} (1 - \frac{d_{23t_j}}{n_{2t_j}}) = 0.161 \quad (11)$$

$$\sum_{t_j \leq t} \hat{P}_{11}^{(t)} \frac{d_{12t_j}}{n_{1t_j}} \hat{P}_{22}^{(t)} = 0.027 \quad (12)$$

$\hat{P}_{13}^{(t)} = 1 - \hat{P}_{11}^{(t)} - \hat{P}_{12}^{(t)} = 1 - 0.63 - 0.027 = 0.343$ . This is because every row should total to 1.

$\hat{P}_{21}^{(t)} = 0$  since there were no transitions from state 2 to state 1 recorded

Therefore  $\hat{P}_{23}^{(t)} = 1 - \hat{P}_{21}^{(t)} - \hat{P}_{22}^{(t)} = 1 - 0 - 0.161 = 0.839$

This was confirmed through R software using the TPmsm package codes:

```
R>time1<-c(t1, t2, t3, t4, t5, t6, t7, t8, t9, t10 )
R>event1<-c(x1, x2, x3, x4, x5, x6, x7, x8, x9, x10 )
R>Stime<-c(t1, t2, t3, t4, t5, t6, t7, t8, t9, t10)
R>event<-c(x1, x2, x3, x4, x5, x6, x7, x8, x9, x10)
R>age<-c( age1 age2, age3, age4, age5 age6, age7,
age8, age9, age10)
R>Cancerdata<-
with(Cancerdata,survTP(time1,event1,Stime,event,age))
R>Cancerdata
R>Cancerdata<-transAJ(object=Cancerdata,s=949,t=1460)
R>Cancerdata
```

Where variable time1 represent observed time in state 1('healthy' state) and variable event1 the corresponding status, variable Stime represent the total survival time and variable event represent the final status of the individual.

The transition matrix therefore takes the form:

$$T = \begin{bmatrix} 0.630 & 0.027 & 0.343 \\ 0 & 0.161 & 0.839 \\ 0 & 0 & 1 \end{bmatrix}$$

The transition matrix T above shows the entries of T, which represent the probability of a person to transit from one state to another. They present the probability of a person moving to state j given that the person is in state i. For instance, the probability of a person moving to state 2(sickness state) given that he/she is in state 1(health state) is 0.027. The probability of a person moving to death state given that he/she is in health state is 0.343. The probability of a stomach cancer patient transiting to a healthy state is zero. The probability of a stomach cancer patient remaining in the same state for time t is 0.161. The probability of a stomach cancer patient transiting to death state is 0.839. Once a patient enters death state, he/she cannot go back since death is an absorbing state.

The limiting matrix from T was obtained by:

$$\lim_{n \rightarrow \infty} T^n = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

This means that all persons under study will eventually transit to death state, either as a result of stomach cancer or as a result of other causes.

## 4. Conclusions and Recommendations

### 4.1. Conclusions

The following conclusions were made

Markov chain theory was used to develop a simple stochastic stomach cancer model, which was used in describing the movement of a person from one state to another. Movement of a person from one state of cancer to another has been shown to be stochastic and depends on the state the person is in at time t. The model was depicted with a state diagram and a stochastic matrix. Transition probability theory was used to determine the probability of moving from one state of cancer to another. The model was applied to the data and Aalen-Johansen estimators were used to obtain the entries of probability matrix T. The limiting probability matrix was obtained from T, by multiplying T by itself n times. The results show that once a person enters sickness state, the probability of returning to health state is zero. The probability of a stomach cancer patient moving to the absorbing state is 0.839, which is quite high. The limiting probability matrix show that eventually, all the people under study will move to death state, either as a result of stomach cancer or as a result of other causes.

### 4.2. Recommendations

The following recommendations were made from the study:

- The study was restricted to three states of stomach cancer. Further research should be done to incorporate more cancer stages hence more states.
- The study used non parametric method in the analysis of the data. More research using parametric or semi parametric methods should be done to incorporate prognostic variables that affect survival time of stomach cancer patients.
- The study considered right censored data. More research should be done to include left censored data
- The government and other stakeholders should embrace preventive measures instead of curative measures
- People should undergo regular cancer screening to facilitate early diagnosis

## References

- [1] B. M. Dehkordi, A. Safae, M. A. Pourhoseingholi, R. Fatemi, Z. Tsbeie, M. R. Zali( 2008): Statistical Comparison of Survival Models for Analysis of Cancer Data. *Asian Pacific journal of Cancer prevention* 2008. July-September; 9(3): 417-20.
- [2] V. Vallinayagam, S. Prathap, P. Venkatesan (2014): Parametric Regression Models in the Analysis of Breast Cancer Survival Data. *International journal of Science and Technology* volume 3 No 3, March 2014.
- [3] F. E. Ahmed, P. W. Vos and D. Holbert (2007): Modelling Survival in Colon Cancer; a methodological review. *Biomed Central- Molecular cancer* 2007, di:10.1186/1476-4598-6-15.
- [4] D. B. Jun, K. Kim ( 2013): Survival analysis of lung and bronchus cancer patients segmented by demographic characteristics. *KAIST Business School Working Paper series, KCB-WP-2013-029*.

- [5] D. B. Jun, K. Kim (2014): Survival analysis of lung and bronchus cancer patients segmented by demographic characteristics. *KAIST Business School Working Paper series, KCB-WP-2014-026*.
- [6] U. Erisoglu, M. Erisoglu and H. Erol (2011): A Mixture Model of Two different Distributions Approach to the Analysis of Heterogeneous Survival Data. *International Journal of Computational and Mathematical Sciences*, 5:2 2011.
- [7] C. Yu, R. Greiner, H. Lin: Learning Patient-specific Cancer Survival Distributions as a Sequence of Dependent Regressors.
- [8] H. A. Glick 2007: Introduction to Markov Models. *Kyung Hee University July 2007*.
- [9] J. Jung 2006: Estimation of Markov Transition Probabilities between Health States in the HRS Datas. *Research gate:228419924 (2006)*.
- [10] E. L. Abner, R. J. Charnigo and R. J. Kryscio 2014: Markov chain and Semi-Markov chain models in time-to-event analysis. *Journal Of Biomedical And Biostatistics; Suppl 1(E001): 19522-doi: 10.4172/2155-6180. SI-e001*.
- [11] T. Heggland 2015: Estimating transition probabilities for the illness- death model. *Faculty of Mathematics and Natural Sciences University of Oslo*.
- [12] M. Snijders 2017: Prediction for Transition Probabilities in Multi-State models. *Utrecht University*.
- [13] L. M. Machado, J. de Una-Alvarez, Carmen Cadarso-Suarez and Per Andersen 2010: Multi-state models for analysis of time to event data. *Statistical Methods in medical Research*. 2009 April; 18(2): 195-222.
- [14] American Cancer Society (2012). Cancer Facts & Figures 2012: *American Cancer Society*.
- [15] Sirpa Heinavaara (2003). Modelling Survival of Patients with Multiple Cancers. *Statistics in Medicine* 2002 Nov 15;21(21).
- [16] Kenya Cancer Network (2013). *Kenya Cancer Statistics and National Strategy Report 2013*.
- [17] World Health Organization (2009). *Cancers*.
- [18] Ministry of Public Health and Sanitation (2014). National Cancer Control Strategy.
- [19] Tiwari R, Jernail A, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, Thun MJ(2004): *Cancer Statistics 2004. Cancer Journal for Clinicians* 2004. 54(1): 8-29.