تحليل العوامل المتنافسة في سرطان الثدي باستخدام نموذج المتعدد الحالات

خه نده غريب عزيز [1]، عباس كل مراد بك مراد [2]
[1] كلية التقنية حلبجة للعلوم التطبيقية، جامعة التقنية السليمانية، السليمانية، إقليم كوردستان، عراق
[2] قسم الاحصاء والمعلومات،كلية الادارة والاقتصاد ،جامعة السليمانية ، السليمانية، إقليم كوردستان، عراق
Corresponding author's e-mail: khanda.gharib@gmail.com

Abstract

Survival analysis can be defined as a field that studies the time period until the occurrence of an event. However, in some cases, these methods may not gain sufficient control over the disease process because disease progression may involve interesting intermediate events. Therefore, multistate model have multiple events or states, which can give greater knowledge and clarity of disease progression than a pure model for survival analysis. The main purpose of this study is to reduce the ambiguity of the multistate model theory that relies on the Markov property to estimate the competing factors that may have an impact on the amount of a patient's surgery and to estimate the severity of transmission and the probabilities of transition between different cases (transient as well as absorption) of breast cancer patients. Finally, each factor has different effects on each transition.

Key words: Multistate model, Markov Property, Cox Proportional Hazard Model, Breast Cancer, Competing factors.

پوخته

دەتوانرێت پێناسەی شیکاری هەوڵ دان بۆ مانەوە لە ژیان بکرێت بەوەی کە لقێکە گرنگی دەدات بە شیکردنەوەی کات هەتا ڕووداوێک ڕووبەدات. وە لەگەڵ ئەوەشدا لە هەندێک حاڵەتدا ئەم بیردۆزە ناتوانێت بەتەواوی کۆنترۆڵی پرۆسەی نەخۆشی بکات. کە بەرەو پێشچوونی نەخۆشی چەندین ڕووداو قۆناغی گرنگ دەگرێتە خۆی. لەبەرئەوە بیردۆزی فرە هەنگاو چەندین حاڵەت و ڕووداوی گرنگی تێدایە کە زۆرترین زانیاری و بەرچاو ڕوونی زیاتر دەدات بە بەرە وپێشچوونی نەخۆشیەکە وەک لەوەی لە شیکاری بیردۆزی هەوڵ دان بۆ مانەوە ژیان(تاك هەنگاو) دەخرێتە ڕوو. ئامانجی سەرەکی لەم توێژینەوەیە کەم کردنەوەی ناڕوونی و زیاتر ڕوونکردنەوەی بیردۆزی فرە هەنگاوە بە پشبەستن بە تایبەتمەندی مارکوف بۆ ئەوەی خەملاندن بکرێت بۆ ئەو هۆکارانەی کە لەوانەیە کاریگەری هەبێت لەسەر بڕی ئەونەشتەرگەریانەی کە دەکرێت بۆ ئەو کەسانەی شێرپەنجەی مەمکیان هەیە. هەروەها خەملاندن بکرێت بۆ توانای گواستنەوە و ئەگەری گواستنەوە لە بەینی حاڵەتە جیاوازەکان (حاڵەتی ناوەند و پڕ) بۆ نەخۆشانی شێرپەنجەی مەمک. لە کۆتاییدا هەریەکە لە هۆکارە کێبڕکێ کارەکان کاریگەری جیاوازیان هەیە لەسەرگواستنەوە لە بەینی حاڵەتەکان.

گۆڤاری زانکۆی هەڵەبجە:گۆڤارێکی زانستی ئەکادیمیە زانکۆی هەڵەبجە دەری دەکات	
بەرگ	٥ ژمارە ٣ ساڵی(٢٠٢٠)
ڕێککەوتەکان	ڕێکەوتی وەرگرتن:٢٠٢٠/٦/٢٥ | ڕێککەوتی پەسەندکردن:٢٠٢٠/٨/٢٦ | ڕێککەوتی بڵاوکردنەوە:٢٠٢٠/٩/٣٠
ئیمەیلی توێژەر	khanda.gharib@gmail.com
مافی چاپ و بڵاوکردنەوە	© ٢٠٢٠ خە نده غريب عزيز، م. د.عباس كل مراد بك مراد،گەیشتن بەم توێژینەوەیە کراوەیە لەژێر ڕەزامەندی CCBY-NC_ND 4.0

ووشەی سەرەکی:

بیردۆزی فرە هەنگاو, تایبەتمەندی مارکۆف, بیردۆزی چەماوەی کۆکس, شێرپەنجەی مەمک، هۆکارە کێبڕکێ کارەکان .

الملخص

يمكن تعريف تحليل البقاء على قيد الحياة بأنه حقل يدرس الفترة الزمنية حتى وقوع الحدث. ومع ذلك ، في بعض الحالات ، قد لا تكتسب هذه الأساليب السيطرة بشكل كافٍ على عملية المرض لأن تطور المرض قد يتضمن أحداثًا وسيطة مثيرة للاهتمام. لذلك فإن النماذج متعددة الحالات لها أحداث أو حالات متعددة ، والتي يمكن أن تعطي معرفة ووضوحًا أكبر لتطور المرض مما يوفره نموذج نقي لتحليل البقاء. الغرض الرئيسي من هذه الرسالة هو تقليل الغموض عن نظرية النموذج متعدد الحالات يعتمد على خاصية ماركوف لتقدير العوامل المتنافسة التي قد يكون لها تأثير على كمية جراحة المريض وتقدير شدة الانتقال واحتمالات الانتقال بين حالات مختلفة (عابرة وكذلك استيعاب) من مرضى سرطان الثدي. أخيرًا ، لكل عامل تأثيرات مختلفة على كل انتقال.

الكلمات الرئيسية: نموذج متعدد الحالات، خصائص ماركوف، نماذج كوكس الخطر النسبي، سرطان الثدي، عامل منافس

1. Introduction

In recent years Multistate and competing risk models are considered as one of the common models in survival analysis. This is because of their common use for cancer patients. Researchers mostly focus on this model to collect cancer data. They try to study how treatments and prognostic factors influence the course of a patient of these diseases and to quantify the hazard risks and average survival cancer patients.

Nowadays, Breast Cancer is the type of cancer that more common as compared to other types of cancer. According to many publicities and available Data, breast cancer is the most common type of cancer among women. It comprises 23% of the 1.1 million female cancers that are newly diagnosed each year. This disease became the first cause of death in third world countries. This is because of the shift in lifestyle which causes an increase in incidence, and the lack of clinical advances to fight the disease (Doutani, et al, 2012, 676-681). Furthermore, In Iraqi, and Iraqi province of Kurdistan Breast cancer ranks first among the other forms of cancer in the population and accounts for nearly one-third of reported female cancers cases among the Iraqi population, This shows a trend for the disease to affect younger women as it is the leading cause of death among Iraqi women. This belongs to the lack of investigation and detection programs Cancer early, along with insufficient facilities for diagnosis and treatment (Annual Report Iraqi Cancer Registry (2015-2016).

Multistate model can be called as prolong of survival analysis. It is an important useful way to get the answer of a wide range of questions in survival analysis that are very hard to be answered by olden models. There are many types of multi state model but the more common among other types is the illness-death model. As a model of illness-death, it has often been referred for time- to-event data in which a single terminal or more events has been measured and all people begin in one or more original states. Several intermediate states might be seen between these transitions that may or may not be visited. For variety of reasons, all individuals

of the initial state may not last until the end or absorbing state. This is the situation in which real life data is presented. Those who were able to get to the final state are called censored observations. Basic goal of such models is to define all feasible gradually and to calculate the unsimilar transition probabilities and intensities. From the viewpoint of the above descriptions and explanations, in this study, the five states multistate model has developed for patients who suffer because of breast cancer disease as a stochastic model depends on Markov property. Furthermore, the transition intensities and probabilities between variable states have estimated. It also estimates the effect of the most competing factors behind transitions using the Cox proportional hazard model. The impacts are evaluated before and after surgery for patients. In this manner, a formal test on the equality of these impacts and test this mediator's impacts on death incidence are performed. In this study, the model is used while the initial situation of the patient is divided into four or more occurrences. i.e.; "the patient with the tumor" the divided amount relies on the status of the patient. Also, the method which is used to collect the breast cancer data records the survival time of the patient before and after surgery. By using this method in the original diagnosis and using one model, the results of (no surgery, first surgery, second surgery, third surgery, and death) states during the disease are developed. In application and data description part the model is partitioned into three parts and three models to get accurate results with benefit from Markovian property.

2. Literature reviews:

[Broët, et al, (1999), 83-89] used the Multi-state model as a very helpful model, to evaluate prognostic factors behind each transition in breast cancer disease progression more precisely. The authors have shown that each of a high stage of the clinical, high grade of histological, positive lymph nodes, and age equal or less than 40 years were all factors behind heightening the risk of developing LRR, distant recurrence or death from the beginning of the disease diagnosis. Expect age, previously mentioned factors stayed as factors for metastases or death following LRR. This study reported that patients who experienced chemotherapy for first cancer, the risk for growing MCBC is very poor. The study emphasizes that as the interval between the diagnosis of the primary tumor and LRR increased, transition rate to distant recurrences or death risk decreases. After MCBC, nodal status for the first tumor and clinical stage for the contralateral tumor heightened the probability of metastasis or death. In opposite, the danger for patients who experienced hormone therapy after MCBC reduced. [Saint-Pierre, et al, (2003), 3755-3770] .Used a Markov continuous-time model using time-dependent covariates, and a Markov model with piece-by-piece steady intensities to model asthma control evolution. For this study the definition of asthma control grade was used for each visit to classify the status of the subject as optimal control, sub-optimal control and inappropriate control at the time of the visit. The author has confirmed that BMI is correlated with asthma management, and the amount of exacerbations. [Putter, et al, (2006), 366-380] . Developed a semi-Markov multistate 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 the disease for patients

with a given history. The authors have verified that a grate rate of transition by a positive tumor also the large of tumor size has the same significant but the reverse effect for the transition to death, age is known to be one of the significant risk factors to occur recurrence but sometimes it is non-significant to transition for death, Breast-conserving treatment implies a higher local recurrence rate compared to mastectomy and radiotherapy. The time at which local recurrence and distant metastasis occurred have clear effects on practically all transitions after local recurrence and distant metastasis; early local recurrences and distant metastases increase the risk of progression and death. [De Bock, et al, (2009), 401-408]. This work proposed using a multi-state model to assess whether the impacts of prognostic variables correlated with the frequency of distant metastases (DM) on the key diagnosis shift following the incidence of loco-regional recurrences (LRR) among women treated for invasive stage I or II breast cancer. It also suggests that the presence of an LRR is a major prognostic risk factor (for DM incidence); the key prognostic risk factors for a DM are diagnostic young age, greater tumor size and node positivity, whereas adjuvant chemotherapy is protective for a DM. [Meira-Machado, et al, (2009), 195-222].Discussed the use of multi-state models in the evaluation of survival data comparing multi-state assessment with the well-known Cox model with time-dependent covariates using two data sets: the Stanford heart transplant survey began in October 1967 until 1 April 1974 of the 103 patients. Furthermore, data on Galician breast cancer were collected between April 1991 and December 2003, which included 585 breast cancer patients who were treated in Galicia (Spain). The Stanford multi-state heart transplant model therefore offers new insights and confirms some of the findings obtained with the Cox model. For example, breast cancer was infringed on the Markov property (Cox semi Markov model), used to evaluate the covariate impacts and estimate the transition probabilities (non-Markov). [Grover, et al, (2018), 129-139] . Study the importance of CA15-3 as a disease marker in tracking and assessing the development of patients with breast cancer using a Markov multi-state model. Based on the identified cut-off value of tumor marker CA15-3, the study of the three disease progression states of breast cancer patients is specified. In addition, the importance of the measured CA15-3 highly correlated with the patient, reduced survival. Each age, lymph node, tumor type, and ER status were found to be significantly correlated with the risk of death in breast cancer patients, except for the effect of tumor size, which is not considered to be important in this study.

3. An introduction to Multi state model

A multistate model is a stochastic process (X (t), t ∈ T) with a finite state space S = {1, 2, 3… N}. Where, T = [0, τ], is a duration which observation of the process at time t visits states at that time, and achieve certain assumptions of simplification, such as time homogeneity, Markov or semi-Markov properties. Individuals travel through time, between states, providing data on the occupied states and the transitional period. A history (Ft-) is created by observing the process at intervals [0, t) over time t with the progress of the process. In this sense, transition probabilities and transition intensities are the two quantities that better describe a multistate pro-

cess. (Meira-Machado, et al, 2009, 195-222) .

The transition probabilities are described by

(s, t) = P(X (t) = j|X(s) = i, Ft-) for i, j S, s, t, T, s ≤ t ... (1)

And the transition intensities are determined by the derivatives

$$Pij(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t,\ t + \Delta t)}{\Delta t} \quad \ldots (2)$$

This description means that the multistate model is Markovian, meaning that the possibility of moving to a future state depends only on the present state and not on past.

According to (Ibe, O, 2013) (Ibe, O, 2014), for i, j ∈ S and, t ∈ T where is the conditional probability of entering state j at time t, provided that the system was in state i at time 0.

So, pij (t) = P(X(t) = j|X(0) = i) for all i, j ∈ S and t ≥ 0. The quantities pij (t) are called the transition probabilities. And form a P(t) matrix called the matrix of transition probability. We can therefore describe a CTMC with n possible States

$$P(t) = P\begin{pmatrix} p11(t) & p12(t) \cdots p1n(t) \\ P(t) = p21(t) & p22(t) \cdots p2n(t) \\ . & . \quad . \\ pn1(t) & pn2(t) \cdots pnn(t) \end{pmatrix}$$

Properties of transition probability matrix P (t)

1. All entries are greater than or equal to 0; pij (t) ≥ 0 for all i, j ∈ S and t ≥ 0
2. In every column the number of the entries is 1. j∈S pij (t) = 1 for all i ∈ S and t ≥ 0.
3. $pij(t + s) = \sum_{k \in s} p_{ik}(t)\, p_{kj}$ (s) for all t ≥ 0, s ≥ 0 and i, j ∈ S.)... 3)

The transition intensity can be interpreted as the instantaneous rate or hazard of making a transition from state i to state j at time t, for each state i, the amount of time spent in that state in a given visit is an exponentially distributed random variable. We define the parameter of this exponentially distributed random variable to be qi. At this point qi is referred to as the transition intensity, and by the time homogeneous assumption of the Markov process, this parameter can be regarded as a constant. Let be the rate of going from state i to j. We define the parameter to be qi for this exponentially distributed random variable.

$$Q(t) = \begin{pmatrix} -q_1 & q_{12} & \cdots & q_{1n} \\ q_{21} & -q_2 & \cdots & q_{2n} \\ . & . & . & . \\ q_{n1} & q_{n2} & \cdots & -q_n \end{pmatrix}$$

Properties of transition rate matrix Q (t)

(i) $\sum_{j \in s} q_{ij} = 0$ for all $i \in S$

(ii)$q_i = \sum_{j \neq i} q_{ij}$

(iii)$q_i = -\sum_{j \neq i} q_{ij} = -q_i$ for all $i \in S$.

4. Materials and methods

4.1 Descriptions of the data

The data are taken from the patient register at the Center for the Treatment of Breast Diseases, other medical centers. The study community is patients with breast cancer that have been diagnosed with the disease in the period from January 2011 until December 2018. Universal asymmetry, i.e. all patients with breast cancer, was used in this period were (634) of patients with breast cancer had been hospitalized and had undergone surgery or treated without surgery. The patients had records in the archives of the hospital, and in their files, their addresses and phone numbers were available for subsequent follow-ups. The survival time of patients was determined with and without surgery and those patients who were still alive at the end of the study period or the ones whose data were not available after a specific time-period was considered right-censored. But my drawback refers to the register's patient information suffer from missing values and the sample size about (319) females are accessed. Also, the use of an indicator variable indicating the patient's position takes the value one when the patient dies and takes the value zero if the patient is alive or has lost follow-up. The data are represented in age, the presence of carcinoma of the lymph nodes, tumor size in centimeters, and the margin of the surgery involved by tumor or not, type of the treatment.

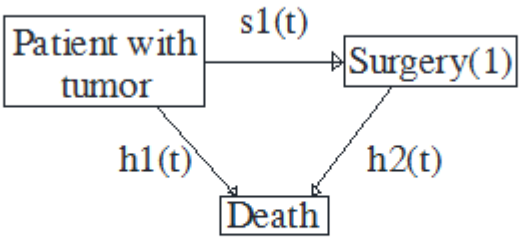
4.2 Statistical analysis

In this analysis, with the aid of R package mstate (de Wreede, et al, 2011, 1-30), a time-non homogeneous Multistate Markov model was applied to analyze the progression of breast cancer disease. The p-value < 0.01 has been accepted as statistically important in all situations. States associated with disease progression were dependent on the amount of surgery performed by patients, described as

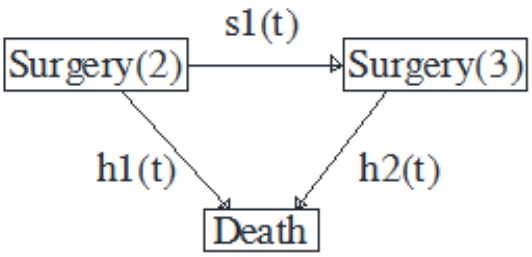
State 1 2 3 4 5

Patient with tumor first surgery second surgery third surgery Death

Nevertheless, there is no reverse transition from third surgery to second or first surgery, as for other study. The model included covariates as categorical variables – age (≤ 50 years/>50 years), tumor size (< 2cm/2-5 cm/>5 cm), nodal status (positive/negative), surgical margin (positive/negative), 'treatment' included both surgical mastectomy and breast-conserving surgery with and without treatment, no surgery with and withaout treatment. A patient's survival time, which was alive at the end of the study period, was regarded as right censored. After close analysis of the count of surgery it became apparent that reverse transitions do not need to be included in the model. States 1 to 2, 3, 4 are transient states 5 and the only consuming state is death. A patient can switch to the absorbing state continuously between different transient states and even from any of the transient state. The model can be partitioned in to three models in three parts to get accurate results and benefits from Markov property.



Figure(1)schematic representation of three states Multi-state model



Figure(3)schematic representation of three states Multi-state model

5. Cox proportional hazard regression Model

In (1972), by summing up earlier work on the life table (Kaplan and Meier, 1958), Cox presented a model in one of his famous papers, introducing his model postulating a simplified model known as the Cox regression model. This is one of the well-recognized statistical risk models in the Survival Analysis to examine the relationship between many variables overtime before the incident takes to happen in the context of an outcome such as death. This model is referred as semi parametrical model because that model has two components the first is called the baseline hazard function of time the other is an exponential term used to produce different hazard rates for each individual based on which covariate groups they belong. (Bradburn,et al, 2003, 431-436).

This model can be employed in multistate model to calculate the transition intensity and transition probability between states and to calculate the effect of covariate vector z on transition i⟶j. It is expressed by

$$\lambda_{ij}(t|z) = {}_0(t) \times \exp\{ B^T_{ij} Z\} \quad \dots (4)$$

${}_0(t)$: represents the baseline hazard function in i to j.

B^T_{ij} : is the vector of regression coefficient that describes the effect of Z on transition in i to j.

$$\lambda_{ij} \quad {}_0(t|z) = 0(t) \times \exp\{B^T_{ij} Z_{ij}\} \quad \dots (5)$$

Z_{ij} : is the vector of covariates specific to transition in i to j, defined for patients based on his / her covariates Z Provided the covariate does not change over time. .(Andersen, et al, 1991, 153-167).

In this study at first Cox proportional hazard regression model is applied to our Data to estimate cumulative hazard function with ignores influence covariates for each of three parts. The basic quantities of interest are the transition intensities or hazard rates from state i to state j for each transition and each part. This situation is illustrated with the help of three figures for each part. In the interest of space, its numeric result is not shown here. But the numeric result for prediction transition probability is presented for each of three parts. A second time Cox proportional hazard regression model is applied to calculate the effect of covariates on the transition between states for breast cancer patients the output is represented in three tables for each of three parts.

6. Estimation

6.1. Estimation baseline cumulative hazard function

For cumulative hazard analysis, we presume that we have data with independent censorship.

The cumulative hazard estimator (t) for transformation i ⟶ j is given by Nelson-Aalen

$$\hat{A}_{ij}(t) = \sum_{t_i \le t} \frac{dN_{ij}(t_i)}{Y_i(t_i)} \quad \dots (6)$$

Where t_i indicates the time of the incident, $dN_{ij(t_i)}$ is the observed number of transitions from state i to state j at time t_i, and $Y_{i(t_i)}$ is the number of subjects at risk for a transition from state i at t_i and the estimates of the diagonal elements A(t) are given by $\hat{A}_{ii}(t) = -\sum_{i \neq j} \hat{A}_{ij}(t)$.

(de Wreede, et al, 2011, 1-30).

6.2. Estimation the transition probability matrix

The Alan- Johnson estimator is used to determine the transition probability matrix, and the formula for this method is as follows:

$$\hat{p}(s,t) = \prod_{u \in (s,t)} (I + \hat{q}_j)) \quad \dots (7)$$

Where u indicates the event times, I is an n × n identity matrix, and $\hat{q}_j$ is a n × n matrix with ijth element. (de Wreede, et al, 2011, 1-30).

6.3. Estimation parameters of the model

If the ordered transition times from state i to state j are < < ... then the partial likelihood for that transition is given by

$$L(B_{ij}) = \prod_{k=1}^{m} \frac{\exp(\beta'_{ij} z_{ijk}(t_{ijk}))}{\sum_{L \in R(t_{ijk},i)} \exp(\beta'_{ij} z_j(t_l))} \quad \dots (8)$$

Where $z_{ij,k}$ is the covariate vector for patient k, and $R(t_{ij,k})$ is the risk set at time t for making transition from i⍰j.

Consequently, the pervious formulation is a generalization of the Cox (1972) partial likelihood function to multistate models.

The maximum partial likelihood estimate $\hat{b}$ = (ˆb1, ˆb2, . . . , ˆbk) 0 can be found by setting the derivative of the expression log in Equation (8) with respect to b equal to zero and iteratively resolving the resulting simultaneous equations using computational methods such as the Newton-Raphson process. (Kay , 1982, 1743-1756).

Competing factors	Class code	Frequency	Proportion
Age			
≤ 50	(1)	207	64.9
> 50	(2)	112	35.1
Tumor size			
< 2 cm	(1)	82	25.7
2-5 cm	(2)	207	64.9
> 5cm	(3)	30	9.4
Nodal status			
Positive	(1)	187	58.6
Negative	(2)	132	41.4
Surgical margin			
Positive	(1)	31	9.7
Negative	(2)	266	83.4
No surgery	(3)	22	6.9
Type of the treatment			
Mastectomy	(1)	8	2.6
Mastectomy + therapy	(2)	123	38.6
B.c.s	(3)	55	17.2
B.c.s + therapy	(4)	111	34.8
No surgery + therapy	(5)	15	4.7
No surgery + no therapy	(6)	7	2.1

Table (1): Represents descriptive features of breast cancer patients (n=319)

7. Data analysis

7.1. Data analysis for part one

The model can be represented via a transition matrix which is a 3-by-3 matrix in this case. A number at matrix entry (i;j) represents a possible change from state i to state j. Such numbers vary from 1 to 3 here, as there are 3 transitions in the pattern. These transition numbers are used during the different phases of the study. If a transition is not permitted between two States, the entry becomes zero.

	Initial	First	Death
Initial	0	1	2
First	0	0	3
Death	0	0	0

Table (2): Represents observed transition between states

The table above shows the observed transition during the follow-up visits between states (rows

	To state				
From state	Initial	First	Death	No event	Total entering
Initial	0	297	13	9	319
First	0	0	19	278	297
Death	0	0	0	32	

to column). There are 297 transitions from state 1 (patient with tumor to surgery one). In state two, 13 patients died and diagnosed with cancer without surgery, while 9 patients alive without surgery diagnosed with cancer, but they are alive with the help of other therapies like chemo, radio, hormone during the Study, also 19 deaths occurred in transition three after surgery one. There is no back transition from state 2, 3 to state 1 because the patient after surgery one cannot go back or from death, and from state 3 to state 2 after the patient has done surgery one and died the patient cannot come back. Totally in this table 9 and 278 transitions were censored from Sate1 and State 2, respectively and 32 patients died with and without surgery.

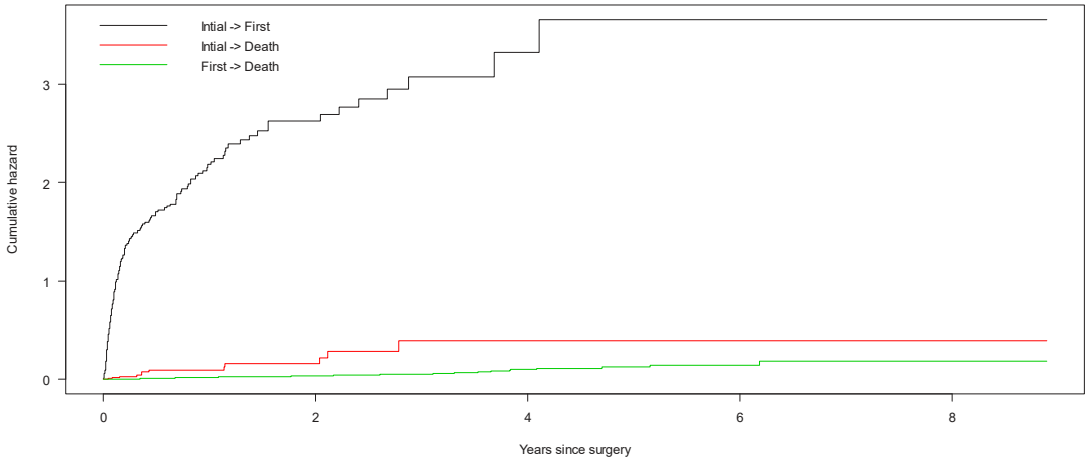


Figure (4): Represents baseline hazard curves for all transitions by cox-regression model

In Figure (4) it is clear that the cumulative hazard is increased with increasing time which leads to increasing the number of deaths for the patients under the Study. It also demonstrates the heights hazard from state one which means at first diagnosis patient with tumor to surgery one. Moreover, the hazard is high in patients that do not have done surgery compared to the patients that have done surgery one.

The model can also be represented using a transition probability that is the conditional probability of entering a patient in state j at time t, provided that the patient was in state i at time 0 with the assumption that is; The starting time is interpreted to calculate the transition probabilities for each transition that is a list of (S+1) components, where S is the number of states. Item j of the list includes all probabilities of transition starting from

state j to all states and their corresponding standard errors (j = 1,.., S).

Table (3) Represents prediction probabilities for all transitions from head and tail of state 1

	Time	pstate1	pstate2	pstate3	se1	se2	se3
1	0.000000000	1.0000000	0.000000000	0	0.000000000	0.000000000	0
2	0.002737851	0.9905956	0.009404389	0	0.005378564	0.005378564	0
3	0.005475702	0.9686520	0.031347962	0	0.009666244	0.009666244	0
4	0.008213552	0.9435737	0.056426332	0	0.012777259	0.012777259	0
5	0.010951403	0.9341693	0.065830721	0	0.013744685	0.013744685	0
6	0.013689254	0.9184953	0.081504702	0	0.015170030	0.015170030	0
149	4.084873	0.02188329	0.8402764	0.1378403	0.009087235	0.02682659	0.02567736
150	4.109514	0.01458886	0.8475708	0.1378403	0.007768499	0.02654902	0.02567736
151	4.700890	0.01458886	0.8357990	0.1496121	0.007768499	0.02863844	0.02785870
152	5.158111	0.01458886	0.8206027	0.1648085	0.007768499	0.03183105	0.03116184
153	6.190281	0.01458886	0.7912954	0.1941157	0.007768499	0.04172282	0.04125894
154	8.895277	0.01458886	0.7912954	0.1941157	0.007768499	0.04172282	0.04125894

Table (4): Represents the Likelihood ratio test of postulated model

-2Loglikelihood	Chi-square	d.f	Sig
70. 63	54.1	12	0.0000000002

The above table shows the likelihood ratio test for fitted cox-regression model which is equal to (70.63) and the result of the test clarifies that the model is statistically significant because the significant value is less than (0.01).

Table (5): Represents Parameter estimates for all covariates and all transitions with cox-regression model

Factors	$\hat{B}_i$	SE	Wald	d.f	Significant value
Age.1	-0.242	0.124	8.174	1	0.004
Age.2	0.348	0.614	2.035	1	0.153
Age.3	0.667	0.477	0.408	1	0.522
Tumor.size.1	0.198	0.090	3.745	1	0.052
Tumor.size.2	-0.475	0.637	0.084	1	0.771
Tumor.size.3	-0.203	0.408	0.230	1	0.630
Nodal.status.1	0.187	0.119	4.261	1	0.038
Nodal.status.2	0.142	1.302	5.207	1	0.022
Nodal.status.3	-1.271	0.574	4.079	1	0.043
Treatment.1	-0.159	0.041	15.370	1	0.00008
Treatment.2	1.248	0.425	20.195	1	0.000006
Treatment.3	0.940	0.308	6.8372	1	0.008

Table (5) represents the hazard ratio for each covariate (age, tumor size, nodal status and treatment) on each transition along with their 99% confidence interval (CI). In second transition from initial to death Covariate treatment, (no surgery with other treatment) is found to be statistically associated to transition from patient diagnosed with tumor associated to transition with hazard of death of breast cancer patient's that do not do surgery because they diagnosed their diseases at high stage of cancer.

Each of the age of group (≤ 50 years), positive nodal status, and treatments are significantly associated with the transition 1 → 2. More precisely, patients (≤ 50 years) in state 1 are more likely to move to state two. This interprets it as majority of patient's diagnosed with breast cancer as their age's ≤ 50 years old and have signif-

icant tumor size as well as at first before surgery type of treatment, (no surgery with and without treatment) is statistically associated with transition 1→ 2. As well as type of treatment and positive nodal is statistically associated with transition 1→ 3. In transition 2→ 3 type of treatment and positive nodal have significantly associated to transition to death for patients who underwent first surgery.

7.2. Data analysis for Part two

The model can be represented via a transition matrix which is a 3-by-3 matrix in this case. A number at matrix entry (i;j) represents a possible change from state i to state j. Such numbers vary from 1 to 3 here, as there are 3 transitions in the pattern. These transition numbers are used during the different phases of the study. If a transition is not permitted between two States, the entry becomes zero.

	First	Second	Death
First	0	1	2
Second	0	0	3
Death	0	0	0

Table (6) represents observed transition between states

	To state				
From state	First	Second	Death	No event	Total entering
First	0	67	19	196	282
Second	0	0	7	60	67
Death	0	0	0	26	26

The above table shows that the observed transition between states (rows to column) during the follow-up visits. There are 67 transitions from state 1 (first surgery to second surgery) in state two 19 patients died after surgery one, and state three has 7 transitions which represents patients died after second surgery. There is no back transition from state 2 to state 1 or from state 3 to state 2 because patients after have done surgery one and died; the patient cannot go back to surgery one or go after surgery two, as the patient cannot comeback to surgery one. Totally in this table, 196 and 60 transitions were censored after surgery one and surgery two, respectively 26 patients died.

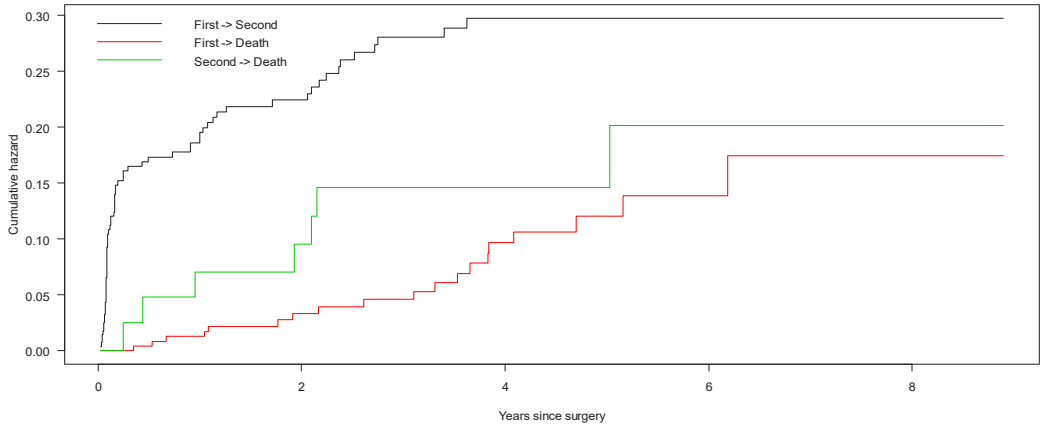


Figure (5): Represents baseline hazard curves for all transitions by cox-regression model

Figure (5) illustrates the heights hazard from state one which means at first surgery to second surgery. Moreover, it shows that the hazard of patients that have done surgery two is more compared to those patients that have done surgery one, also the number of patients that have done surgery two are less compared to those patients that done surgery one. It is also, clear that the cumulative hazard is increased which leads to increasing the number of deaths for the patients under the Study.

The model can also be represented using a transition probability that is the conditional probability of entering a patient in state j at time t, provided that the patient was in state i at time 0 with the assumption that is, the starting time is interpreted to calculate the transition probabilities for each transition that is a list of (S+1) components, where S is the number of states. Item j of the list includes all probabilities of transition starting from state j to all states and their corresponding standard errors (j = 1,.., S).

Table (7): Represents prediction probabilities for all transition from state1 (head and tail)

	Time	pstate1	pstate2	pstate3	se1	se2	se3
1	0.00000000	1.0000000	0.000000000	0	0.000000000	0.000000000	0
2	0.03011636	0.9964539	0.003546099	0	0.003533524	0.003533524	0
3	0.03285421	0.9929078	0.007092199	0	0.004988243	0.004988243	0
4	0.03832991	0.9893617	0.010638298	0	0.006098386	0.006098386	0
5	0.04380561	0.9858156	0.014184397	0	0.007029156	0.007029156	0
6	0.04928131	0.9822695	0.017730496	0	0.007844662	0.007844662	0
74	4.084873	0.6667120	0.2298280	0.1034601	0.03109734	0.02678218	0.02141305
75	4.700890	0.6574521	0.2298280	0.1127199	0.03199607	0.02678218	0.02307050
76	5.032170	0.6574521	0.2170598	0.1254882	0.03199607	0.02802172	0.02583283
77	5.158111	0.6454984	0.2170598	0.1374418	0.03353508	0.02802172	0.02808816
78	6.190281	0.6224449	0.2170598	0.1604953	0.03924141	0.02802172	0.03527410
79	8.895277	0.6224449	0.2170598	0.1604953	0.03924141	0.02802172	0.03527410

Table (8): Represents the Likelihood ratio test of postulated model

-2Loglikelihood	Chi-square	d.f	Sig
119.3	120.3	15	0.0000000000000002

The above table shows the likelihood ratio test for fitted cox-regression model which is equal to (119.3) and the result of the test clarifies that the model is statistically significant because the significant value is less than (0.01).

Table (9): Represents Parameter estimates for all covariates and all transitions with cox-regression model

Factors	$\hat{B}_i$	SE	Wald	d.f	Significant value
Age.1	-0.249	0.314	6.672	1	0.009
Age.2	0.331	0.467	0.253	1	0.614
Age.3	-1.402	1.113	0.334	1	0.562
Tumor.size.1	0.441	0.209	17.231	1	0.00003
Tumor.size.2	-0.346	0.416	0.261	1	0.609
Tumor.size.3	-0.472	0.743	0.918	1	0.337
Nodal.status.1	-0.250	0.268	1.947	1	0.162
Nodal.status.2	-1.203	0.580	4.480	1	0.034
Nodal.status.3	-17.63	2329	9.464	1	0.002
Surgical.Margin.1	-1.680	0.359	44.163	1	0.0000000000302
Surgical.Margin.2	16.33	9334	0.146	1	0.701
Surgical.Margin.3	-0.892	1.167	2.496	1	0.114
Treatment.1	0.610	0.150	19.145	1	0.00001
Treatment.2	0.647	0.251	7.745	1	0.005
Treatment.3	-0.807	0.360	4.030	1	0.044

Table (9) represents the hazard ratio for each covariate (age, tumor size, nodal status, surgical margin and treatment) on each transition along with their 99% confidence interval (CI). Each of age ≤ 50, tumor size, surgical margin which is involved by tumor or type of treatment have associated with transition 1 → 2. In transition 1 → 3 from first surgery to death positive nodal status and type of treatment is found to be statistically associated with hazard of death. In third state for breast cancer patient's that have done second surgery positive nodal status and type of treatment have statistically associated with transition 2 → 3.

7.3. Data analysis for part three

The model can be represented using a matrix for transformation, which is a matrix of 3-by-3 in this case. A number at matrix entry (i;j) represents a possible change from state i to state j. These numbers range from 1 to 3 here, since there are 3 transitions in the pattern. These transition numbers are used during the various phases of the study. If a transition is not permitted between two States, the entry becomes zero.

	Second	Third	Death
Second	0	1	2
Third	0	0	3
Death	0	0	0

Table (10) represents observed transition between states

	To state				
From state	Second	Third	Death	No event	Total entering
Second	0	15	7	60	82
Third	0	0	4	11	15
Death	0	0	0	11	11

The above table shows that the observed transition between states (rows to column) during the follow-up visits. There are 15 transitions from state 1 (second surgery to third surgery), in state two 7 patients died after surgery two, state three has 4 transitions which represents patients died after surgery three. There is no come back transition from state 2 to state 1 or from state 3 to state 2 because patients after that have done surgery three; the patient cannot come back to surgery two or die after surgery two and three as the patient cannot comeback from death. Totally in this table, 60 and 11 transitions were censored after surgery two and surgery three, respectively 11 patients died.

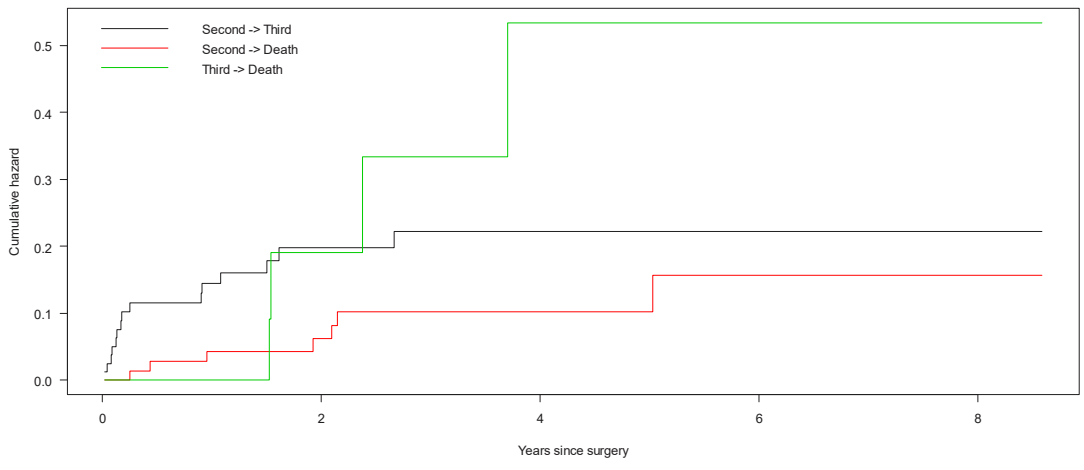


Figure (6): Represents baseline hazard curves for all transitions by cox-regression model

Figure (6) represents the highest hazard in stage three where patient after surgery three move to Death compares those patients that have done second surgery also the number of patients that have done third surgeries is less compares the patients that have done second surgery this means as the number as surgery increases the hazard increases with it.

The model can also be represented using a transition probability that is the conditional probability of entering a patient in state j at time t, provided that the patient was in state i at time 0 with the assumption that is, the starting time is interpreted to calculate the transition probabilities for each transition that is a list of (S+1) components, where S is the number of states. Item j of the list includes all probabilities of transition starting from state j to all states and their corresponding standard errors (j = 1,.., S).

Table (11): Represents prediction probabilities for all transition from state1 (head and tail)

	Time	pstate1	pstate2	pstate3	se1	se2	se3
1	0.00000000	1.0000000	0.00000000	0	0.00000000	0.00000000	0
2	0.02464066	0.9878049	0.01219512	0	0.01204640	0.01204640	0
3	0.04654346	0.9756098	0.02439024	0	0.01693004	0.01693004	0
4	0.07939767	0.9634146	0.03658537	0	0.02060417	0.02060417	0
5	0.08487337	0.9512195	0.04878049	0	0.02363962	0.02363962	0
6	0.12320329	0.9390244	0.06097561	0	0.02625884	0.02625884	0
23	2.151951	0.7402053	0.1478843	0.1119104	0.05018465	0.04026109	0.03679655
24	2.376454	0.7402053	0.1267580	0.1330367	0.05018465	0.03897199	0.04061737
25	2.663929	0.7217002	0.1452631	0.1330367	0.05215055	0.04226385	0.04061737
26	3.704312	0.7217002	0.1162105	0.1620894	0.05215055	0.04102906	0.04602337
27	5.032170	0.6816057	0.1162105	0.2021838	0.06212726	0.04102906	0.05818252
28	8.588638	0.6816057	0.1162105	0.2021838	0.06212726	0.04102906	0.05818252

Table (12): Represents the Likelihood ratio test of postulated model

-2Loglikelihood	Chi-square	d.f	Sig
44.68	17.77	15	0.00009

The above table shows the likelihood ratio test for fitted cox-regression model which is equal to (44.68) and the result of the test clarifies that the model is statistically significant because the significant value is less than (0.01).

Table (13): Represents Parameter estimates for all covariates and all transitions with cox-regression model

Factors	$\hat{B}_i$	SE	Wald	d.f	Significant value
Age.1	1.49	0.529	5.932	1	0.014
Age.2	-1.071	1.151	0.330	1	0.565
Age.3	43.06	12380	0.101	1	0.749
Tumor.size.1	0.126	0.358	0.379	1	0.537
Tumor.size.2	-0.093	0.684	0.091	1	0.762
Tumor.size.3	52.07	10490	0.335	1	0.562
Nodal.status.1	0.137	0.528	0.022	1	0.879
Nodal.status.2	19.99	9087	9.740	1	0.001
Nodal.status.3	52.76	11370	11.736	1	0.0006
Surgical.Margin.1	1.385	0.63	5.305	1	0.021
Surgical.Margin.2	-0.129	1.38	1.426	1	0.232
Surgical.Margin.3	-94.83	18170	2.951	1	0.085
Treatment.1	-0.158	0.296	0.279	1	0.597
Treatment.2	-0.788	0.396	4.656	1	0.030
Treatment.3	8.52	2506	1.386	1	0.239

Table (13) represents the hazard ratio for each covariate (age, tumor size, nodal status and treatment) on each transition along with their 99% confidence interval (CI). In first transition from second surgery to third surgery covariate age ≤ 50 and surgical margin which is involved by tumor is statistically associated with transition 1→2. On the other hand, positive nodal and type of the treatment associated with transition 1→3 as well as positive nodal associated with transition 2→ 3 push patients to death after third surgery.

8. Conclusions

The Multistate Markove model is applied for recording the complex progression of breast cancer in women. Since the multistate model is the natural option for studying the progression of disease and it also helps us estimate transition intensities and probabilities of change among states. By using the properties of Markove models, this study demonstrated the utility of the non-homogeneous multistate disease mortality model in the review of the breast cancer disease follow-up study.

The following conclusions are drawn and created to support humanity through statistics:

1. Young age is a common competing factor to occur recurrence and makes patients become surgery from patients diagnosed with tumor to surgery or from first to other surgeries, as well as the reverse effect for the transition to death.

2. It is clear that the grate rate of transition by positive nodal status also the large of tumor size has the significant effect but the reverse effect for the transition to death.

3. Surgical margin significantly increases transition rates from first surgery to other surgeries, as well as the converse influence for transition to death.

4. Type of treatment that has done for breast cancer patients significantly increases the transition rates from

first surgery to second also; it increases the transition rates after surgery to death this obviously seen from table (5 , 9, 13).

5. According to the results, the worst stage of breast cancer, which is the fourth stage and is (52) patients out of (319) patients, is reached in patients (16 per cent). This can be interpreted as having a life-long connection to the advanced stage of the illness. Nearly (16%) of patients refer to local spreading or metastatic doctors and this is the leading cause of death hazard and short lifetime.

6. The hazard function is highest in patients who have done surgery three compares with patients who have done surgery two also surgery two compares surgery one. As the number of surgeries increases the hazard function increases and the elevated time highly associated with higher hazard of the patient.

9. Recommendations

After pointing to conclusions that were obtained in this study, the following recommendation for future work are relevant:

1. It is possible to use another type of multistate model to study some chronic disease like (kidney, heart, diabetes…).

2. It is feasible to use other competing factors that maybe have an effect on breast cancer surgeries like hormone receptors (Estrogen, Progesterone, Her2 when they are triple negative) and used big sample size as much as to get more accurate conclusions.

References:-

Books:

[1] Ibe, O. (2013). Markov processes for stochastic modeling. Newnes.

[2] Ibe, O. (2014) Fundamentals of applied probability and random processes.Academic Press.

Journal Papers:

[3] Andersen, P. K., Hansen, L. S., & Keiding, N. (1991). Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. Scandinavian Journal of Statistics.

[4] Broët, P., de la Rochefordière, A., Scholl, S. M., Fourquet, A., De Rycke, Y., Pouillart, P., ... & Asselain, B. (1999). Analyzing prognostic factors in breast cancer using a multistate model. Breast cancer research and treatment.

[5] Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. British journal of cancer.

[6] Doutani, H., Doutani, J., Parveen, S., Khan, F., Baloch, D. M., & Roohullah, H. K. N. A. (2012). Clinical and histopathological characteristics of breast cancer in people of baluchistan and adjoining territory of afghanistan. Current Research Journal of Biological Sciences.

[7] De Bock, G. H., Putter, H., Bonnema, J., Van Der Hage, J. A., Bartelink, H., & Van De Velde, C. J. (2009). The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model. Breast cancer research and treatment.

[8] de Wreede, L. C., Fiocco, M., & Putter, H. (2011). mstate: an R package for the analysis of competing risks and multi-state models. Journal of statistical software.

[9] Grover, G., Swain, P. K., Goel, K., & Singh, V. (2018). Multistate Markov Modelling for Disease Progression of Breast Cancer Patients Based on CA15-3 Marker. Thailand Statistician.

[10] Kay, R. (1982). The analysis of transition times in multistate stochastic processes using proportional hazard regression models. Communications in Statistics-Theory and Methods.

[11] Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. Statistical methods in medical research.

[12] Putter, H., van der Hage, J., de Bock, G. H., Elgalta, R., & van de Velde, C. J. (2006). Estimation and prediction in a multi-state model for breast cancer. Biometrical Journal: Journal of Mathematical Methods in Biosciences.

[13] Saint-Pierre, P., Combescure, C., Daures, J.P., Godard, P., (2003).The analysis of asthma control under a markov assumption with use of covariates. Statistics in Medicine.

Websites:

[14] Annual Report Iraqi Cancer Registry (2015-2016).Republic of Health Ministry of Health\Environment. Iraqi Cancer Board.