Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Modeling the pathway of breast cancer in the Middle East

Asmaa Amer^{1,2}, Ahmed Nagah^{1,3}, Mojeeb AL-Rahman El-Nor Osman¹, Abdul Majid¹

¹ School of Mathematics and Statistics, Central China Normal University, Wuhan, 430079, P. R. China

² Faculty of Science, Mathematics Department, Menoufia University, Egypt

³ Faculty of Science, Mathematics Department, Zagazig University, Egypt

Submission Info

Communicated by Juan Luis García Guirao

Received November 20th 2019

Accepted April 11th 2021

Available online November 19th 2021

Abstract

This paper proposed an approach for the identification of mutation mechanisms of breast cancer in women in four member countries of the Middle East Cancer Consortium i.e. Egypt, Jordan, Cyprus and Israel (Arabs and Jews). We set up multi-stage models including both gene mutation and the clonal expansion of intermediate cells. We fit the data-set related to the incidence of female breast cancer in the four member countries. Our simulation results show that the maximum number of driver mutations of breast epithelium stem cells of Egyptian women is 13, whereas there are 14 driver mutations in the genome of stem cells of female patients in Jordan, Cyprus and Israel (Arabs and Jews). In addition, the 3, 10, 5, 5 and 4 stage models are the optimal ones for the tumorigenesis of females in Egypt, Jordan, Cyprus, Israel (Arabs) and Israel (Jews), respectively. The genomic instability is caused by first three driver mutations.

Keywords: breast cancer; multi-stage model; incidence rate; chi square test

1 Introduction

Cancer is the uncontrolled growth of abnormal cells that exist in a human body. This disease has become one of the major cause of death all over the world. Breast Cancer is the most common type of cancer that affect women. It constitutes 10% of all new cancer cases worldwide and 1-2% of mortality among women populations throughout the world [1, 2]. Recently, studies on female breast cancer have attracted the attention of biologists and geneticists [3–17]. Breast cancer emerges when an unregulated growth of abnormal cells begins in different parts of tissue. A tissue is a group of similar cells which have an identical general function. This may develop in milk ducts and glands of the breast. Every year, over one million new cases are diagnosed with breast cancer. However, recently many breast cancer deaths have been prevented. This is due to improvements in treatment and early detection through mammography. It is globally estimated that at least 400,000 patients are dying of this disease every year [18–21].

Breast cancer is considered a major public health problem in both developing and developed countries. The incidence of breast cancer differs in various regions. In developing countries, it is lower than in developed countries. They range between 22 and 71/100,000 of women population, but mortality rates are considerably higher [1, 2, 22–26]. Nevertheless, breast cancer is the most commonly diagnosed cancer among women in the Middle East, because many women do not seek medical care promptly [22–24, 27, 28]. Therefore, the increase in the rate of incidences slows down in this region. The average age of women affected by breast cancer in the Middle East is around a decade earlier than the average of that in the west. Comprehensive studies show that the incidence rates of breast cancer are increasing among four countries of the Middle East Cancer Consortium (MECC).

The Middle East Cancer Consortium (MECC) is an example of regional cooperation in cancer registry. The comparison of the countries in the region helps to identify common risk factors and sharing cancer prevention and control strategies. Najla et al. [2] showed age-standardized and age-specific incidence rates of breast cancer among Lebanese women and compared them with those of the regional and western countries. Sherko et al. [29] demonstrated that the incidence of breast cancer among Iraqi Kurdish women is lower than some other countries of Middle East and the west. Hanhan et al. [30] suggested eight biological capabilities that should be acquired during the multistep development of human tumors, namely, self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, resistance of programmed cell death (apoptosis), limit- less replicative potential, sustained angiogenesis, tissue invasion and metastasis, reprogramming of energy metabolism, and evading immune destruction.

The process of carcinogenesis is considered as a result of the degeneration of a cell from the normal to a malignant state through a finite number of intermediate stages. The deterioration of a normal cell leads to the proliferation of transformed cells whose descendants generate a tumor. A normal cell may generate a tumor after it has gone through two mutations. How many driver mutations are needed for human breast cancer which are necessary for tumor diagnosis and therapy? Wood et al. [31] obtained an important result that less than 15 (< 15) driver mutations were likely to be responsible for genomic landscapes of human breast and colorectal cancers. The earliest two-stage model with clonal expansion of intermediate cells were developed by Moolgavkar et al. in age-specific incidence rates of breast and colorectal cancers [32–34]. Zhang and Simon in 2005 [35] extended the two-stage model to six-stage model with clonal expansion of intermediate cells in each compartment. The two to six stage models with clonal expansion of intermediate cells fit the age-specific incidence rates of breast cancer very well. Zhang et al. [36] in 2014 used the statistical inference (Chi-square test) to test both the two to six stage models with clonal expansion of intermediate cells and the selection between mutation and clonal expansion of intermediate cells. By using the method of statistical inference, Li et al. [37] obtained the maximum number of driver mutations for the breast tumorigenesis of females in United States of America.

The data pattern for the age-specific incidence rates of breast cancer in the USA [38] is different from that in the Middle East countries. In this work, our main objective is to investigate the main differences in the mutation mechanism of breast cancer of women among four member countries (Egypt, Jordan, Cyprus and Israel) of the Middle East Cancer Consortium (MECC) and compare them with the results obtained by Li et al. [37] for the United States of America (USA). In section 2, we briefly describe the data resources. In section 3, we present the mathematical model. Section 4 shows the inference method. Section 5 provides the simulation results. Finally, section 6 describes brief discussion of the results.

2 Data Resources

The data used in this work are the age-specific incidence rates of breast cancer for female patients in four member countries of the Middle East Cancer Consortium (MECC), i.e., Egypt, Jordan, Cyprus and Israel (Arabs and Jews) [1]. The source of Egypt data is the National Cancer Registry Program (NCRP) during the period 2008-2011 [39]. The Jordan Cancer Registry covers the population from 1996-2001 [1]. The Cyprus National Cancer Registry reports the cyprus data from 1998 to 2001 [1]. Israel data for Arabs and Jews are extracted

from the Israel National Cancer Registry from the year 1996 to the year 2001 [1]. The collection of data by the MECC registries is guided by the Manual of Standards for Cancer Registration [40]. For our analysis, we used the incidence of breast cancer rates by age at diagnosis, date of birth and sex. Incidence rates are expressed as cases per 100,000 females.

Figure 1 shows the age-specific incidence rates of female breast cancer in Egypt, Jordan, Cyprus, Israel (Arabs and Jews) and the USA. Most women diagnosed with breast cancer are over 45. Female breast cancer is rare under the age of 25. In the Middle East countries, the lowest age-specific rate is from age groups of 0-4 and 20-24, then increases gradually till it reaches the highest rate and then decreases. Whereas, in the USA, incidence rate increases gradually till it reaches the highest rate without decrease. The majority of female breast cancer cases were diagnosed in the USA and Israel (Jews). The data pattern for the age-specific incidence of female breast cancer is not similar in Middle East countries, and it is different from that in the USA. The peak of the incidence rate is at the age group of 60-64 especially for Egypt, Jordan and Israel (Arabs) and 70-74 for Cyprus, and Israel (Jews). However, the peak of the incidence rates is at the age group of 75-79 for the USA data.

Fig. 1 Incidence of breast cancer in four member countries (Egypt, Jordan, Cyprus and Israel (Arabs and Jews)) of the Middle East Cancer Consortium (MECC) and USA.

3 Mathematical Model

The oncology processes should be a sequence of genetic events according to the findings of Hanahan et al. [30]. Figure 2 shows the schematic representation of multi-stage model with clonal expansion in each compartment of intermediate cells. We supposed that the growth of normal stem cells can be denoted by a logistic curve from age 0 to age 20. Our assumptions are according to ref. [35, 36]. There are ten cells in each breast tissue at birth or at age 0 and 10^7 cells at the age of 20. There are 10^6 stem cells during age 80. Since $10^7 \times e^{-0.0667(80-45)} = 10^6$, in this work, it is assumed that the number of stem cells decreases after age 45 at a rate -0.0667 per cell per year because of the expression levels of hormone and menopause.

For a k -stage model ($k = 2, 3, \dots$), let $Y_1(t)$, $Y_i(t)$ ($i = 2, 3, \dots, k$), and $Y_{k+1}(t)$ represent, respectively, the

number of normal progenitor cells, intermediate cells in compartment I_i , and fully malignant cells per breast at time t . $\Psi(y_1, y_2, \dots, y_{k+1}; t)$ introduces the probability generating function at time t starting with a single normal cell at time 0 [34, 41–43]:

$$\Psi(y_1, y_2, \dots, y_{k+1}; t) = \sum_{i_1, i_2, \dots, i_{k+1}} P(Y_1(t) = i_1, Y_2(t) = i_2, \dots, Y_{k+1}(t) = i_{k+1} | Y_1(0) = 1, Y_2(0) = 0, \dots, Y_{k+1}(0) = 0) y_1^{i_1} y_2^{i_2} \dots y_{k+1}^{i_{k+1}}$$

Fig. 2 A schematic representation of k -stage model ($k = 2, 3, \dots$) for carcinogenesis. N denotes the normal progenitor cells. I_i ($i = 2, 3, \dots, k$) denotes the compartments of intermediate cells, μ_i ($i = 1, 2, \dots, k$) are the mutation rates per cell per year of normal progenitor cells and compartment of intermediate cells, and γ_i ($i = 1, 2, \dots, k$) are the net growth rates per cell per year. M denotes the malignant cells.

Then, the Kolmogorov forward differential equation [33, 36, 43] for Ψ is defined by:

$$\begin{aligned} \Psi'(y_1, y_2, \dots, y_{k+1}; t) &= \frac{\partial \Psi}{\partial t}(y_1, y_2, \dots, y_{k+1}; t) \\ &= \{\mu_1(t)y_1y_2 + \alpha_1(t)y_1^2 + \beta_1(t) - (\alpha_1(t) + \beta_1(t) + \mu_1(t))y_1\} \frac{\partial \Psi}{\partial y_1} + \dots \\ &\quad + \{\mu_k(t)y_ky_{k+1} + \alpha_k(t)y_k^2 + \beta_k(t) - (\alpha_k(t) + \beta_k(t) + \mu_k(t))y_k\} \frac{\partial \Psi}{\partial y_k} \end{aligned} \quad (1)$$

where $\alpha_i(t)$, $\beta_i(t)$, $\mu_i(t)$ are the growth rate per cell per year, death or differentiation rate per cell per year, and mutation rate per cell per year in each compartment I_i at time t , respectively.

The hazard function (incidence function) is the instantaneous rate of the appearance of malignant tumor in a previously tumor free tissue [33, 35]. Let T represent the time of the appearance of the first tumor. Then

$$h(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \text{Prob} \{t \leq T < t + \Delta t | T \geq t\} \right) \quad (2)$$

From equation (2),

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \text{Prob} \{t \leq T < t + \Delta t | T \geq t\} \right) \\ &= \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \frac{P\{t \leq T < t + \Delta t\}}{P(T \geq t)} \right) \\ &= \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \frac{P(t + \Delta t) - P(t)}{1 - P(t)} \right) \\ &= \frac{P'(t)}{1 - P(t)} \end{aligned} \quad (3)$$

The hazard and probability are related by the given equation [33, 35]

$$P(t) = 1 - \exp \left\{ - \int_0^t h(s) ds \right\} \quad (4)$$

where $P(t)$ and $h(s)$ denote the probability of breast cancer occurring at age t and the incidence rate of breast cancer at age s , respectively.

According to ref [44],

$$P(t) = 1 - \Psi(1, 1, \dots, 1, 0; t) \quad (5)$$

Then hazard function or the incidence function should be written in terms of Ψ and Ψ' as follows:

$$h(t) = -\frac{\Psi'(1, 1, \dots, 1, 0; t)}{\Psi(1, 1, \dots, 1, 0; t)} \quad (6)$$

where $\Psi(1, 1, \dots, 1, 0; t)$ is the survival function and $\Psi'(1, 1, \dots, 1, 0; t)$ is the derivative of survival function. By using equation (1), we get

$$\frac{\partial \Psi}{\partial t}(1, 1, \dots, 1, 0; t) = -\mu_k \frac{\partial \Psi}{\partial y_k}(1, 1, \dots, 1, 0; t) \quad (7)$$

Furthermore the conditional expectation is given by

$$E[Y_k(t)|Y_{k+1}(t) = 0] = \frac{\frac{\partial \Psi}{\partial y_k}(1, \dots, 1, 0; t)}{\Psi(1, \dots, 1, 0; t)} \quad (8)$$

By substituting from equations (7) and (8) in (6), we obtained

$$h(t) = \mu_k E[Y_k(t)|Y_{k+1}(t) = 0] \quad (9)$$

Since cancer is a rare disease, the probability $P(t)$ of the presence of a malignant cell at time t is close to zero. Thus,

$$E[Y_k(t)|Y_{k+1}(t) = 0] \approx E[Y_k(t)] \quad (10)$$

Then, we have the approximation for one breast tissue:

$$h(t) = \mu_k E[Y_k(t)] \quad (11)$$

In this section, we proposed the multi-stage model of ordinary differential equations [35, 36] that we used in our calculations, in order to estimate the expectation $E[Y_i(t)]$ ($i = 1, 2, \dots, k$), to estimate the number of $Y_i(t)$ before and after age 45.

For $t \leq 45$, $Y_i(t)$ ($i = 1, 2, \dots, k$) should satisfy the following ordinary differential equations:

$$\begin{aligned} \frac{dY_1(t)}{dt} &= \gamma_1 Y_1(t) \left(1 - \frac{Y_1(t) + Y_2(t) + \dots + Y_k(t)}{K}\right) - \mu_1 Y_1(t) \\ \frac{dY_2(t)}{dt} &= \mu_1 Y_1(t) + \gamma_2 Y_2(t) - \mu_2 Y_2(t) \\ \frac{dY_3(t)}{dt} &= \mu_2 Y_2(t) + \gamma_3 Y_3(t) - \mu_3 Y_3(t) \\ \frac{dY_4(t)}{dt} &= \mu_3 Y_3(t) + \gamma_4 Y_4(t) - \mu_4 Y_4(t) \\ &\vdots \\ \frac{dY_k(t)}{dt} &= \mu_{k-1} Y_{k-1}(t) + \gamma_k Y_k(t) - \mu_k Y_k(t) \end{aligned} \quad (12)$$

At $t > 45$, $Y_i(t)$ ($i = 1, 2, \dots, k$) should satisfy the following ordinary differential equations:

$$\begin{aligned} \frac{dY_1(t)}{dt} &= \gamma'_1 Y_1(t) - \mu_1 Y_1(t) \\ \frac{dY_2(t)}{dt} &= \mu_1 Y_1(t) + \gamma'_2 Y_2(t) - \mu_2 Y_2(t) \\ \frac{dY_3(t)}{dt} &= \mu_2 Y_2(t) + \gamma'_3 Y_3(t) - \mu_3 Y_3(t) \\ \frac{dY_4(t)}{dt} &= \mu_3 Y_3(t) + \gamma'_4 Y_4(t) - \mu_4 Y_4(t) \\ &\vdots \\ \frac{dY_k(t)}{dt} &= \mu_{k-1} Y_{k-1}(t) + \gamma'_k Y_k(t) - \mu_k Y_k(t) \end{aligned} \quad (13)$$

where K is the maximal number of normal progenitor cells so $K = 10^7$, $\gamma_1 = 1.0074$, and $\gamma'_1 = -0.0667$. In addition, γ_i and γ'_i denote the growth rates of the intermediate compartments I_i ($i = 2, 3, \dots, k$) before and after age 45. The growth rates change after age 45 in response to the changing hormonal environment of a women.

4 Chi-Square Test

In order to study some biological mutation mechanisms of female breast cancer such as impact of genetic mutation and clonal expansion of cells on cancer evolution, we have to hypothesize two of the following:

Hypothesis one: There is a mutator phenotype in the progression of the tumor of breast cancer for multi-stage models, that is, the mutation rates per cell per year in i -stage model ($i = 2, 3, \dots, k$) should satisfy the relation $\mu_1 \leq \mu_2 \leq \dots \leq \mu_k$.

Hypothesis two: There is no mutator phenotype in the tumorigenesis of breast cancer for multi-stage models, that is, the mutation rates per cell per year in the i -stage ($i = 2, 3, \dots, k$) model should satisfy $\mu_1 = \mu_2 = \dots = \mu_k$.

To study the mutation mechanism of female breast cancer, we use the statistical inference to test whether the optimal fittings for the 2-15 stage model are available for each country. In this work, Chi-square test is employed to test those optimal fittings. For these countries, there are insufficient data available for the age of 79 years. However, there are only 16 values for the age-specific incidence rates of women breast cancer patients in Egypt, Jordan, Cyprus and Israel. For the 15-stage model, there are 43 unknown parameters. Therefore, it is impossible for us to examine the 15-stage model because the amount of data is small. To produce enough data to implement the Chi-square test, we use the previous formula (4) to convert the data set of incidence data into data set for probability of female breast cancer. Since the probability values are always small, we handle this problem by amplifying $P(t)$ with a suitable multiple (i.e. 10^3); we get 49, 47, 48, 46 and 49 data points for Egypt, Jordan, Cyprus, Israel (Arabs) and Israel (Jews), respectively, which are greater than the one which is adequate for performing the Chi-Square test. The goodness of fit is measured by Chi-square test which is applied within this relation

$$Chi2 = \sum_i (P_i - P_i^*)^2 * 10^3 / P_i^* \quad (14)$$

where P_i and P_i^* are the probability of breast cancer derived from the real data and that derived from numerical simulation, respectively. The chi-square goodness of fit test is used to accept or reject the null hypothesis. Our calculations are compared with those obtained by statistical distribution at 5% significant level.

5 Results

The key question in this study is how many driver mutations are needed in the oncology processes for women patients in the four countries of the MECC, i.e., Egypt, Jordan, Cyprus and Israel. A useful tool to answer this question is the use of statistical inference. The mortality may decrease if the breast cancer is diagnosed at early stages, by the implementation of proper awareness and screening programmes. In this paper, Chi-square test is employed to test whether the 2-15 stage models are available. The calculations were obtained using the numerical optimization routine, `fminsearch` in MATLAB. Table 1 exhibits the results of Chi-square test for the four countries calculated using equation (14), and comparing them with the values at 5% significant level for the 2-15 stage models under the first hypothesis. The fitting of models more than 15 stages given worse results than the 15-stage model.

In Egypt data, the Chi-square values of the multi-stage models starting from 2 to 13 are less than the values at 5% significant level, whereas the 14 and 15 stage models give Chi-square values which are greater than the values at 5% significant level. Therefore, we reject 14 and 15 stage models and accept 2-13 stage models. However, we reject only the 15 stage model and accept 2-14 stage models for Jordan, Cyprus, and Israel, which are similar to that of the USA [37]. It is clear that the 2-14 stage models fit the data very well and these models provide a very small error. The error is the difference between real data and simulated data. It is obvious from the optimal parameters for the net proliferation rates that the net proliferation rates of premalignant cells of the 2-14 stage models before age 45 are larger than those after 45 years of age, which are similar to that in the USA [37]. Net proliferation rates of intermediate cells with one mutation approach zero. This confirms the fact that normal stem cells need two mutations in tumor suppressor genes to become the intermediate cells that

Table 1 Minimum Chi-square values for multi-stage models in the four member countries (Egypt, Jordan, Cyprus and Israel) of the Middle East Cancer Consortium (MECC); Chi2 denotes the value of Chi-square; df(=number of data-number of parameters-1) denotes the degrees of freedom; and S.5% denotes the value at 5% significant level.

Hypothesis One						
Multi-stage model	Chi-square test	Egypt	Jordan	Cyprus	Israel(Arabs)	Israel(Jews)
Stage 2	Chi2	1.5901	1.9497	3.9351	1.4681	2.9320
	df	44	42	43	41	44
	S.5%	60.481	58.124	59.304	56.942	60.481
Stage 3	Chi2	0.2860	1.0191	1.0778	0.9077	0.9929
	df	41	39	40	38	41
	S.5%	56.942	54.572	55.758	53.384	56.942
Stage 4	Chi2	3.1342	1.2497	0.8470	5.3882	0.0289
	df	38	36	37	35	38
	S.5%	53.384	50.998	52.192	49.802	53.384
Stage 5	Chi2	0.4743	25.2820	0.0609	0.2932	0.3452
	df	35	33	34	32	35
	S.5%	49.802	47.400	48.602	46.194	49.802
Stage 6	Chi2	2.8163	0.7947	0.2816	0.8875	0.6357
	df	32	30	31	29	32
	S.5%	46.194	43.773	44.985	42.557	46.194
Stage 7	Chi2	1.3874	0.3559	0.1831	1.1421	1.4832
	df	29	27	28	26	29
	S.5%	42.557	40.113	41.337	38.885	42.557
Stage 8	Chi2	2.1740	0.4034	0.3379	1.1119	0.8344
	df	26	24	25	23	26
	S.5%	38.885	36.415	37.652	35.172	38.885
Stage 9	Chi2	1.1726	0.4277	0.6023	1.4395	1.0687
	df	23	21	22	20	23
	S.5%	35.172	32.671	33.924	31.410	35.172
Stage 10	Chi2	2.3850	0.3292	3.4140	1.6899	1.3919
	df	20	18	19	17	20
	S.5%	31.410	28.869	30.144	27.587	31.410
Stage 11	Chi2	4.7264	1.0679	1.3535	3.0451	2.3307
	df	17	15	16	14	17
	S.5%	27.587	24.996	26.296	23.685	27.587
Stage 12	Chi2	8.5654	2.7151	1.4607	3.9033	4.6111
	df	14	12	13	11	14
	S.5%	23.685	21.026	22.362	19.675	23.685
Stage 13	Chi2	10.1917	4.5112	1.7721	5.8185	5.1914
	df	11	9	10	8	11
	S.5%	19.675	16.919	18.307	15.507	19.675
Stage 14	Chi2	63.2724	6.8347	6.8223	6.3965	11.1831
	df	8	6	7	5	8
	S.5%	15.507	12.592	14.067	11.070	15.507
Stage 15	Chi2	92.6729	71.2838	53.7725	91.2415	82.2308
	df	5	3	4	2	5
	S.5%	11.070	7.815	9.488	5.991	11.070

expand clonally to destroy the homeostasis of the cells, or the first mutation occurs in genetic instability genes. By the optimal parameters for the mutation rates, it is found that only the first three mutation rates are small; however, the mutation rates approach to one after three stages. Therefore, the genomic instability is caused by the first three driver mutations (see supplementary materials). Our results are consistent with results obtained

by Li et al. for USA [37]. They have used stochastic multi-stage models to determine how many mutations are needed for the breast cancer processes. In addition, 3, 10, 5, 5 and 4-stage models are optimal models for Egypt, Jordan, Cyprus, Israel (Arabs) and Israel (Jews), respectively. This means that for tumorigenesis, only Egypt has the three-stage model which is the optimal stage for USA.

We have given the optimal fittings for the 2-15 stage models in Figures 3-7. It is clear that models from stage 2 to 14 can fit all data very well except the data from Egypt. The models which are more than 13-stage cannot fit the data from Egypt. Figure 3 displays the optimal fitting for the 2-15 stage models of Egypt for the period from 2008 to 2011. We can see that 14 and 15 stage models cannot fit the data, so we can only accept the mathematical models with 2-13 mutations at the 5% significant level. The optimal fittings for the 2-15 stage models in Jordan from 1996 to 2001 are shown in Figure 4. We accept the 2-14 stage models at the 5% significant level and reject the 15 stage model at the 5% significant level. The age-specific incidence rates per 100,000 females for breast cancer in Cyprus during 1998-2001 are given in Figure 5. Two to fourteen stage models fit the data very well, while the 15 stage model provides a bad fitting with a large error. We plot the age-specific incidence rates for Israel (Arabs) during 1996-2001 in Figure 6. The 15 stage has led to a very large error, so we reject it and accept the 2-14 stage models. Figure 7 shows the age-specific breast cancer incidence for Israel (Jews) in 1996-2001. The 2-14 stage models fit the data very well, and thus we reject the 15 stage model.

We applied hypothesis two of constant mutation rate per cell per year to check the genetic instability. This hypothesis means that there is no mutator phenotype but only the clonal expansion of intermediate cells in the process of breast carcinogenesis. Mutator phenotype refers to the increase in mutation rates of breast cancer cells [45]. Table 2 shows the minimum Chi-square values of the goodness fit of the data for the 2-6 stage models. The fitting of the models in more than 6 stages given worse results than the 6-stage model. We found that models with two to four stages can fit Egypt, Jordan and Cyprus data and reject all other models that are more than 4-stage of the selection of clonal expansion of intermediate cells in the process of breast carcinogenesis, because it gives a large error. Models with two to five stages can fit the data from Israel. However, models with more than three stages cannot fit the SEER data [37]. We conclude that the first two or three stages lead to genetic instability before generating a cancer cell. We plot the probability under hypothesis two for models with two to six stages for the four countries as shown in Figures 8-12. In figures (8-10), it can be observed that two, three, and four stage models provided good simulations with data, while five and six stage models cannot fit the data well. Therefore, we should accept the two-four stage models and reject the models with more than four stages. In Figures 11 and 12, the 2-5 stage models fit the data very well. So, we reject all other models of the selection of clonal expansion of intermediate cells in the process of breast carcinogenesis. This hypothesis gives the best results for the data from Israel (Arabs and Jews) because two to five stage models give us good simulation for the data from Israel.

6 Discussion

The purpose of this work is to investigate the main differences in the mutation mechanism of breast cancer of women in the four member countries (Egypt, Jordan, Cyprus and Israel) of the Middle East Cancer Consortium (MECC) and compare them with the results obtained by Li et al. [37] for the United States of America (USA). Li et al. [37] have used the stochastic multistage models to fit the data set from America on age-specific incidence rates of breast cancer by using several coupled ordinary differential equations derived from the Kolmogorov backward equations. We used different mathematical models to estimate the age-specific incidence rates of breast cancer for women in four member countries (Egypt, Jordan, Cyprus and Israel) of the Middle East Cancer Consortium (MECC). In addition, we have used statistical inference to test the number of mutation stages that are required for female breast cancer. We estimated the optimal parameters by using the numerical optimization routine, `fminsearch` in MATLAB, to calculate the minimum Chi-square value. We made comparisons among four countries to know the differences in the process of breast carcinogenesis. The regional comparisons can shed light on the health care factors and population, to identify common risk factors and share cancer prevention

Table 2 Chi-square values for the biological hypothesis two.

Hypothesis Two						
Multi-stage model	Chi-square test	Egypt	Jordan	Cyprus	Israel(Arabs)	Israel(Jews)
Stage 2	Chi2	1.5984	1.9152	3.9383	53.384	2.9285
	df	45	43	44	42	45
	S.5%	61.656	59.304	60.481	58.124	61.656
Stage 3	Chi2	0.3994	0.5781	1.1131	0.2319	0.8121
	df	43	41	42	40	43
	S.5%	59.304	56.942	58.124	55.758	59.304
Stage 4	Chi2	6.9790	6.5856	3.1302	3.3066	4.3960
	df	41	39	40	38	41
	S.5%	56.942	54.572	55.758	53.384	56.942
Stage 5	Chi2	59.9711	74.4047	55.3319	10.6929	51.2783
	df	39	37	38	36	39
	S.5%	54.572	52.192	53.384	50.998	54.572
Stage 6	Chi2	72.6603	164.2708	143.3481	82.7561	159.2855
	df	37	35	36	34	37
	S.5%	52.192	49.802	50.998	48.602	52.192

and control strategies. The differences between countries reflect the variation in lifestyle, genetic influences, environmental exposure, and the availability of mammography screening.

The maximum number of driver mutations is 13 in Egypt. However, by goodness fit of data to the probabilities of female breast cancer occurrence in Jordan, Cyprus, and Israel (Arabs and Jews), the maximum number of driver mutations for three countries is found to be 14. These results support the results of Wood et al. [31] that are less than 15 mutations required to develop breast cancer and colorectal cancer. The optimal models are 3, 10, 5, 5 and 4-stage for Egypt, Jordan, Cyprus, Israel (Arabs) and Israel (Jews), respectively. The maximum number of driver mutations is 14 in the USA, and the three-stage model is the optimal model [37]. Therefore, Egypt has the same optimal model as United States of America (USA), but differs in the maximum number of driver mutations for breast cancer. Genomic instability is a characteristic of almost all human cancers. By optimal parameter values for mutation rates of multi-stage models, it is found that only the first three mutation rates are small. Therefore, the mutation of instability gene occurred due to the first three mutations (see supplementary materials). This confirms the fact that normal stem cells need two mutations to become intermediate cells that expand clonally to destroy the homeostasis of the cells, or the first mutation occurs in genetic instability genes. The mutations rates and clonal expansion of intermediate cells depend on menopause. Therefore, hormone therapy such as tamoxifen has a good effect on female patients before menopause. By testing hypothesis two, we have checked the selection between the mutator phenotype and the clonal expansion of intermediate cells. We conclude in the case of three countries (Egypt, Jordan and Cyprus), that if the driver mutations are less than five stages, we cannot reject the selection of both the mutator phenotype and the clonal expansion of intermediate cells. If the driver mutations are larger than four stages, we should reject the selection of the clonal expansion of intermediate cells. As a matter of fact, if the driver mutations are larger than four stages, carcinogenesis does not need the mutator phenotype because all mutation rates after four stages approach to one. However, in Israel (Arabs and Jews), we should accept the two-five stage models at the 5% significant level and reject models that are more than 5-stage. According to our results, hypothesis one gives the best results for the four countries compared to hypothesis two because 2 to 13 or 14 stage models are available and give us a good simulation and a small error.

Fig. 3 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Egypt from 2008 to 2011 under hypothesis one. The blue line represents the model and red stars represent the data from Egypt.

Fig. 4 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Jordan from 1996 to 2001 under hypothesis one. The blue line represents the model and red stars represent the data from Jordan.

Fig. 5 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Cyprus from 1998 to 2001 under hypothesis one. The blue line represents the model and red stars represent the data from Cyprus.

Fig. 6 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Israel (Arabs) from 1996 to 2001 under hypothesis one. The blue line represents the model and red stars represent data from Israel (Arabs).

Fig. 7 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Israel (Jews) from 1996 to 2001 under hypothesis one. The blue line represents the model and red stars represent data from Israel (Jews).

Fig. 8 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Egypt under hypothesis two that all mutation rates are equal. The blue line represents the model and red stars represent data from Egypt.

Fig. 9 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Jordan under hypothesis two that all mutation rates are equal. The blue line represents the model and red stars represent the data from Jordan.

Fig. 10 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Cyprus under hypothesis two that all mutation rates are equal. The blue line represents the model and red stars represent the data from Cyprus.

Fig. 11 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Israel (Arabs) under hypothesis two that all mutation rates are equal. The blue line represents the model and red stars represent data from Israel (Arabs).

Fig. 12 The simulation of the probability of age-specific incidence rates of all races per 100,000 females for breast cancer in Israel (Jews) under hypothesis two that all mutation rates are equal. The blue line represents the model and red stars represent data from Israel (Jews).

References

- [1] Rennert G.: Breast cancer, In Cancer Incidence in the Four Member Countries (Cyprus, Egypt, Israel, and Jordan) of the Middle East Cancer Consortium (MECC) Compared with US SEER. Volume (Chapter 8). Edited by L. S. Friedman, B. K. Edwards, L. A. G. Ries, J. L. Young. National Cancer Institute. NIH Pub No. 06-5873. Bethesda, MD;:73-81, (2006).
- [2] N. A. Lakkis, S. M. Adib, M. H. Osman, U. M. Musharafieh, and G. N. Hamadeh, Breast cancer in Lebanon: Incidence and comparison to regional and western countries, *Cancer Epidemiology*, vol. 34, no. 3, 221-225, (2010).
- [3] T. V. Clendenen, W. Ge, K. L. Koenig, Y. Afanasyeva, C. Agnoli, L. A. Brinton, F. Darvishian, J. F. Dorgan, A. H. Eliassen, R. T. Falk, et al., Breast cancer risk prediction in women aged 35-50 years: impact of including sex hormone concentrations in the gail model, *Breast Cancer Research*, vol. 21, no. 1, pp. 42, (2019).
- [4] A. M. Falstie-Jensen, A. Kjarsgaard, E. L. Lorenzen, J. D. Jensen, K. V. Reinertsen, O. M. Dekkers, M. Ewertz, and D. P. Cronin-Fenton, Hypothyroidism and the risk of breast cancer recurrence and all-cause mortality-a danish population-based study, *Breast Cancer Research*, vol. 21, no. 1, pp. 44, (2019).
- [5] C. M. Vachon, C. G. Scott, R. M. Tamimi, D. J. Thompson, P. A. Fasching, J. Stone, M. C. Southey, S. Winham, S. Lindström, J. Lilyquist, et al., Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk, *Breast Cancer Research*, vol. 21, no. 1, pp. 68, (2019).
- [6] M. T. Grinde, B. Hilmarsson, H. M. Tunstet, I. M. Henriksen, J. Kim, M. H. Haugen, M. B. Rye, G. M. Mlandsmo, and S. A. Moestue, Glutamine to proline conversion is associated with response to glutaminase inhibition in breast cancer, *Breast Cancer Research*, vol. 21, no. 1, pp. 61, (2019).
- [7] R. Demicheli, H. Dillekas, O. Straume, and E. Biganzoli, Distant metastasis dynamics following subsequent surgeries after primary breast cancer removal, *Breast Cancer Research*, vol. 21, no. 1, pp. 57, (2019).
- [8] J. Santucci-Pereira, A. Zeleniuch-Jacquotte, Y. Afanasyeva, H. Zhong, M. Slifker, S. Peri, E. A. Ross, R. L. de Cicco, Y. Zhai, T. Nguyen, et al., Genomic signature of parity in the breast of premenopausal women, *Breast Cancer Research*, vol. 21, no. 1, pp. 46, (2019).
- [9] Y. Shieh, C. G. Scott, M. R. Jensen, A. D. Norman, K. A. Bertrand, V. S. Pankratz, K. R. Brandt, D. W. Visscher, J. A. Shepherd, R. M. Tamimi, et al., Body mass index, mammographic density, and breast cancer risk by estrogen receptor subtype, *Breast Cancer Research*, vol. 21, no. 1, pp. 48, (2019).
- [10] H. Samavat, X. Xun, A. Jin, R. Wang, W.-P. Koh, and J.-M. Yuan, Association between prediagnostic leukocyte telomere length and breast cancer risk: the singapore chinese health study, *Breast Cancer Research*, vol. 21, no. 1, pp.

- 50, (2019).
- [11] X. Zhang, Y. Wang, H.-C. Chiang, Y.-P. Hsieh, C. Lu, B. H. Park, I. Jatoi, V. X. Jin, Y. Hu, and R. Li, BRCA1 mutations attenuate super-enhancer function and chromatin looping in haploinsufficient human breast epithelial cells, *Breast Cancer Research*, vol. 21, no. 1, pp. 51, (2019).
 - [12] C. Petridis, I. Arora, V. Shah, A. Megalios, C. Moss, A. Mera, A. Clifford, C. Gillett, S. E. Pinder, I. Tomlinson, et al., Frequency of pathogenic germline variants in BRCA1, BRCA2, PALB2, CHEK2 and TP53 in ductal carcinoma in situ diagnosed in women under the age of 50 years, *Breast Cancer Research*, vol. 21, no. 1, pp. 58, (2019).
 - [13] L. Huss, S. T. Butt, S. Borgquist, K. Elebro, M. Sandsveden, A. Rosendahl, and J. Manjer, Vitamin D receptor expression in invasive breast tumors and breast cancer survival, *Breast Cancer Research*, vol. 21, no. 1, pp. 84, (2019).
 - [14] E. Linnenbringer, S. Gehlert, and A. T. Geronimus, Black-white disparities in breast cancer subtype: the intersection of socially patterned stress and genetic expression, *AIMS public health*, vol. 4, no. 5, pp. 526, (2017).
 - [15] X. Nian, Y. Nagai, C. Jeffers, K. N. Maxwell, and H. Zhang, Dietary influence on estrogens and cytokines in breast cancer, *AIMS MOLECULAR SCIENCE*, vol. 4, no. 3, pp. 252-270, (2017).
 - [16] A. S. Kashani and M. Packirisamy, Cellular deformation characterization of human breast cancer cells under hydrodynamic forces, *AIMS Biophysics*, vol. 4, pp. 400-414, (2017).
 - [17] H.-C. Wei, Mathematical modeling of tumor growth: the MCF-7 breast cancer cell line, *Mathematical Biosciences and Engineering*, vol. 16, no. 6, pp. 6512-6535, (2019).
 - [18] D. M. Parkin and L. M. Fernandez, Use of statistics to assess the global burden of breast cancer, *The breast journal*, vol. 12, pp. S70-S80, (2006).
 - [19] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, Global cancer statistics, 2002, CA: a cancer journal for clinicians, vol. 55, no. 2, pp. 74-108, (2005).
 - [20] B. O. Anderson, C.-H. Yip, S. D. Ramsey, R. Bengoa, S. Braun, M. Fitch, M. Groot, H. Sancho-Garnier, V. D. Tsu, G. S. H. C. Systems, and P. P. Panel, Breast cancer in limited-resource countries: health care systems and public policy, *The breast journal*, vol. 12, pp. S54-S69, (2006).
 - [21] F. Bray, P. McCarron, and D. M. Parkin, The changing global patterns of female breast cancer incidence and mortality, *Breast cancer research*, vol. 6, no. 6, p. 229, (2004).
 - [22] S. M. Mousavi, A. Montazeri, M. A. Mohagheghi, A. M. Jarrahi, I. Harirchi, M. Najafi, M. Ebrahimi, Breast Cancer in Iran: an epidemiological review, *The breast journal*, vol. 13, no. 4, 383-391, (2007).
 - [23] I. G. Alghamdi, I. I. Hussain, M. S. Alghamdi, M. A. El-Sheemy, The incidence rate of female breast cancer in Saudi Arabia: an observational descriptive epidemiological analysis of data from Saudi cancer registry 2001-2008, *Breast Cancer: Targets and Therapy*, vol. 5, 103-109, (2013).
 - [24] V. Ozmen, B. Ozcina, H. Karanlik, et al., Breast cancer risk factors in Turkish women-a university hospital based nested case control study, *World Journal of Surgical Oncology*, vol. 7, no.1, 37, (2009).
 - [25] N. S. El Saghir, M. K. Kahlil, T. Eid, et al., Trends in epidemiology and management of breast cancer in developing Arab countries: a literature and registry analysis, *International journal of surgery*, vol. 5, no.4, 225-233, (2007).
 - [26] M. Corbex, S. Bouzbid, P. Boffetta, Features of breast cancer in developing countries, examples from North-Africa, *European Journal of Cancer*, vol. 50, no.10, 1808-1818, (2014).
 - [27] K. A. Hirko, A. S. Soliman, A. Hablas, et al., Trends in breast cancer incidence rates by age and stage at diagnosis in Gharbiah, Egypt over 10 years (1999-2008), *Journal of cancer epidemiology*, vol. 2013, Article ID 916394, 7 pages, (2013).
 - [28] D. M. McCredie, M. Coates, A. Grulich, Cancer incidence in migrants to new South Wales (Australia) from the Middle East 1972-91, *Cancer Causes & Control*, vol. 5, no. 5, 414-421, (1994).
 - [29] S. A. M. Karim, H. H. A. Ghalib, S. A. Mohammed, and F. H. R. Fattah, The incidence, age at diagnosis of breast cancer in the iraqi kurdish population and comparison to some other countries of middle-east and west, *International Journal of Surgery*, vol. 13, pp. 71-75, (2015).
 - [30] D. Hanahan and R. A. Weinberg, Hallmarks of cancer: the next generation, *Cell*, vol. 144, no. 5, pp. 646-674, (2011).
 - [31] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjoblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, et al., The genomic landscapes of human breast and colorectal cancers, *Science*, vol. 318, no. 5853, pp. 1108-1113, (2007).
 - [32] S.H. Moolgavkar, A.G. Knudson, Mutation and cancer: a model for human carsinogenesis, *JNCI: Journal of the National Cancer Institute*, vol. 66, no.6, 1037-1052, (1981).
 - [33] S. H. Moolgavkar, A. Dewanji, D. J. Venzon, A stochastic two-stage model for cancer risk assessment. I. the hazard function and the probability of tumor, *Risk Analysis*, vol. 8, no.3, 383-392, (1988).
 - [34] E. G. Luebeck, S. H. Moolgavkar, Multistage carcinogenesis and the incidence of colorectal cancer, *Proceedings of the National Academy of Sciences*, vol. 99, no. 23, 15095-15100, (2002).
 - [35] X. Zhang and R. Simon, Estimating the number of rate limiting genomic changes for human breast cancer, *Breast cancer research and treatment*, vol. 91, no. 2, pp. 121-124, (2005).
 - [36] X. Zhang, Y. Fang, Y. Zhao, and W. Zheng, Mathematical modeling the pathway of human breast cancer, *Mathematical biosciences*, vol. 253, pp. 25-29, (2014).
 - [37] L. Li, T. Tian, and X. Zhang, Mutation mechanisms of human breast cancer, *Journal of Computational Biology*, vol.

- 25, no. 4, pp. 396-404, (2018).
- [38] L. A. G. Ries, M. P. Eisner, C. L. Kosary, et al. SEER Cancer Statistics Review, 1973-1999. National Cancer Institute, Bethesda, MD, (2002).
- [39] A. S. Ibrahim, H. M. Khaled, N. N. Mikhail, H. Baraka, and H. Kamel, Cancer incidence in Egypt: results of the national population-based cancer registry program, *Journal of cancer epidemiology*, vol. 2014, Article ID 437971, 18 pages, (2014).
- [40] M. E. C. Consortium et al., Manual of standards for cancer registration. 2005, Available at: <http://mecc.cancer.gov/MECC Manual of Standards. PDF>. [Last Accessed: 1/06].
- [41] N. T. J. Bailey, *The Elements of Stochastic Processes* (Wiley, New York), (1964).
- [42] C. L. Chiang, *Introduction to Stochastic Processes in Biostatistics* (Wiley, New York), (1968).
- [43] G. Serio, Two-stage stochastic model for carcinogenesis with time-dependent parameters, *Statistics & probability letters*, vol. 2, no. 2, pp. 95-103, (1984).
- [44] C. J. Portier, A. Kopp-Schneider, and C. D. Sherman, Calculating tumor incidence rates in stochastic models of carcinogenesis, *Mathematical biosciences*, vol. 135, no. 2, pp. 129-146, (1996).
- [45] L. A. Loeb, Human cancers express a mutator phenotype: hypothesis, origin, and consequences, *Cancer research*, vol. 76, no. 8, pp. 2057-2059, (2016).