Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 16(6): 7771–7788.
DOI: 10.3934/mbe.2019390
Received: 01 July 2019
Accepted: 08 August 2019
Published: 23 August 2019

Research article

The effect of cigarette smoking on lung cancer evolution

Ahmed Nagah[1,2,*], **Asmaa Amer**[1,3] **and Xinan Zhang**[1]

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

[2] Mathematics Department, Faculty of Science, Zagazig University, Egypt

[3] Mathematics Department, Faculty of Science, Menoufia University, Egypt

* **Correspondence:** Email: a.nagah89@yahoo.com; Tel/Fax: +8613277060260.

Abstract: The aim of this paper is to elucidate the risk of lung carcinogenesis from cigarette smoking among current and former smokers. To achieve this goal, we have set up a stochastic three-stage model to fit the data of Surveillance, Epidemiology, and End Results (SEER) program besides the data set of smoking derived from the Nurses' Health Study cohort of females (NHS) and the Health Professionals Follow up Study cohort of men (HPFS). The calculations are performed by considering both mutation and clonal expansion rates as parameters in each compartment. For current smokers, our findings show that cigarette smoking has more significant impact on the mutation rates of cells than the clonal expansion rates of premalignant cells among men and women. In particular, for male patients, cigarette smoking affects the mutation in normal cells and the transformation from premalignant cells to malignant ones in the optimal model. In addition, cigarette smoking induces only the initial mutation rates in normal cells among American women. For current and former smokers, cigarette smoking stimulates only the clonal expansion rate of the first premalignant cells in both sexes. However, we find that the impact of cigarette smoking is minimal in former smokers who have stopped smoking for over ten years among men and women in US.

Keywords: lung cancer; cigarette smoking; multi-stage models ; clonal expansion; chi-square test

1. Introduction

Cancer develops when changes, or driver mutations, in a cell's DNA cause the cell to grow abnormally. Lung cancer of males and females is the most frequent tumor all over the world, and is a major cause of increased mortality rates, accounting for about 27 % of total cancer deaths in US in 2015 and 20 % of total deaths in Europe (EU) in 2016 [1, 2, 3]. In 2015, Feraly et al. [4] showed that 19 % of mortality rates caused by cancer among men and females are attributed to lung cancer. The rise

in lung cancer rates (incidence and mortality) derives heavily from factors that impact the risk of lung cancer formation in both sexes such as genetic susceptibility, poor diet, occupational exposures, tobacco smoking, and air pollution [5]. In addition, the control of tobacco smoking is considered a power tool for lung cancer prevention. It is clear that there are significant risk factors for lung cancer progression other than tobacco smoking that contribute substantially to lung cancer death rates in non-smokers [6, 7, 8]. These risks include occupational and environmental factors such as radiation exposure. Evidence suggests that environmental causes such as radiation combined with cigarette smoking can drastically increase the production of lung tumors among men and women [9]. Wu et al. [10] used statistical analysis to find that these factors significantly increase the rates of lung cancer by more than 90 % .

Cigarette smoking has been established to contribute the most to the growth of lung cancer in the world. Ray et al. [11] have tried to elaborate the relation between cigarette smoking and cancers other than lung cancer by using incidence rates, calculated from the SEER data. They found out that there is an intense connection between cigarette smoking and several other cancers such as laryngeal, esophageal, and colon cancer. However, the correlation between cigarette smoking and cancers, such as pancreatic and liver cancers is weak. In order to fully validate the extent of the impact of cigarette smoking on lung cancer, we should know the effect of certain attributes regarding smoking such as smoking status (current or former smokers) [12, 13], intensity or dose (the number of cigarettes smoked per day) [14, 15], age at starting smoking [16, 17], duration of smoking, and number of years since cessation [18, 19]. In 2002, Leffondr et al. [20] investigated the effects of these factors within time-dependent variables of Cox's model to case-control study occurred in Canada in the period from 1979 to 1985. They concluded that the two features (intensity and duration) must be separate variables to better fit the data. This result is supported by [21] who considered that smoking duration is more important than smoking intensity in the progression of lung cancer. Flanders et al. [22] suggested using Poisson models with the Armitage-Doll multistage model of carcinogenesis. They then used the previously described models to fit the data in Cancer Prevention Study II, with subjects aged 40-79 years old and intensity less than 40, and confirmed that the duration of smoking has a big influence on lung cancer rates of men in the US.

The mathematical study of cancer that considers both mutation and clonal expansion rates as parameters in each compartment has been of growing interest. The multi-stage models of cancer have had a long history dating back to the 1950s [23, 24]. In addition, these models did not give any attention to the clonal expansion of intermediate cells. The genomic instability in stem cells is confirmed by the fact that there are driver mutations in genes besides the mutations in other types of genes (oncogenes and tumor suppressor genes), which also validates the mutator phenotype hypothesis. Tomlinson and Bodmer [25] examined the mutator phenotype hypothesis and claimed that selection for clonal expansion is sufficient for the evolution of a tumor. Thus, taking mutations and clonal expansion into account is helpful for studying carcinogenesis. Normal stem cell transformation from a healthy cell into a malignant cell needs a number of gene changes (mutations), which are fundamental for carcinogenesis [26, 27]. We do not yet know the number of gene mutations that are required for lung cancer occurrence in both sexes [28]. For instance, statistical analysis shows that fewer than 15 of mutations are adequate for breast and colorectal cancers [29]. Byrne et al. [30] showed that four to six gene mutations are sufficient to form a cancerous colon cell.

There are many studies on the influence of cigarette smoking on lung cancer development using

multistage mathematical models to fit the incidence rates of lung cancer [31, 32]. These scientific works did not provide the details regarding parameters changes. Gaffney et al. [33] proposed a two stage model of Armitage and Doll to analyze the British physician lung cancer data of Doll and Hill under the condition that cigarette smoking stimulates the initial and penultimate mutations. They also assumed that smoking initiates normal target cells and promotes the clonal growth of just the smoke-initiated cells. Moolgavkar and Knudson [34] found that the major role of cigarette smoking is its proliferative effect. The two-stage clonal expansion (TSCE) model was used by Meza et al. [35] to fit lung cancer incidence among non-smokers, current smokers, and former smokers in the Nurses' Health Study (NHS) and the Health Professionals Follow up Study (HPFS). They concluded that cigarette smoking promotes partially altered cells on the pathway to cancer.

In this work, we hope to determine the role of cigarette smoking on the growth of lung cancer within the three-stage stochastic model of cancer. Moreover, we study the effects of smoking features (smoking status/age at start of smoking/intensity/time since cessation) on lung cancer causation among men and women in US. For this goal, we use the 3-hit stochastic model, which permits the property of clonal expansion in each compartment, and analyze the data set of the probabilities of human lung cancer in the SEER registry, the Nurses' Health Study cohort of females (NHS), and the Health Professionals Follow up Study cohort of men (HPFS). The paper is formatted as follows: In section 2, we present the mathematical development of the stochastic model with three-cancer mutations. In section 3, we discuss our numerical results regarding the fitting of the model to data. The conclusion is given in section 4.

2. Materials and method

2.1. The SEER data

Incidence data for lung cancer, rates per 100,000 person-years and 95% confidence intervals (CIs), were calculated from the Surveillance, epidemiology, and end result (SEER) registry from 1992 to 2010 [36, 37]. For our study, we use the probabilities of lung cancer incidence, which is reported by sex, race, age, and calendar year in the thirteen SEER registries. The population bases come from SEER population files (based on the data from the U.S. Census Bureau) by gender, race and are cross-tabulated by calendar years 1992–2010 and 5 year age groups (age 0–85+).

2.2. The smoking data

In 1976, the Nurses' Health Study (NHS) was launched when 121,700 female registered nurses, aged 30–55 years, returned a mailed questionnaire. After ten years, the Health Professionals Follow-up Study (HPFS) was established with 51,529 male health professionals (dentists, optometrists, osteopaths, podiatrists, pharmacists, and veterinarians) aged 40–75 years. The Nurses' Health Study was approved by the Internal Review Board of the Brigham and Womens Hospital in Boston, and the Health Professionals Follow-up Study received Internal Review Board approval from the Harvard School of Public Health in Boston [38]. From 1986 to 2000, the number of lung cancer cases was 994 among the NHS women, and 319 among the HPFS men. For our analysis, we computed the probabilities of lung cancer incidence rates by including current age (40–79 years), gender (male and female), smoking status (current/former smokers), intensity or dose (<25 and ≥ 25 cigarettes per day), age at start of

smoking (≥20 and ≤19 years), and time since quitting (≥10 and ≤10 years). Smoking intensity and duration of smoking were considered as separate variables.

2.3. Mathematical model

The oncology processes should be a sequence of genetic events according to the findings of Hanhan et al. [39]. The mathematical MSCE models are mainly based on the clonal expansion in each compartment. The intermediate cells are described by the symmetric division and differentiation or death processes. The schematic representation of the 3-stage model with clonal expansion of premalignant cells in each compartment is displayed in figure 1. In this study, We suppose that there are N(t) normal stem cells, set to be unchanged $N(t) = N$ in each lung tissue, and ν be the initial mutation rate per cell per year such that $\mu_0 = \nu * N$.

Let α_i , β_i, and μ_i are the birth rate per cell per year, death or differentiation rate per cell per year, and mutation rate per cell per year in the compartment $I_i(i = 1, 2)$, respectively. For the 3-hit model, We choose $X_i(t)(i = 1, 2)$ to symbolize the number of premalignant cells in the compartment $I_i(i = 1, 2)$ per lung at time t, $X_3(t)$ the number of malignant cells per lung at time t. For t $\geq \tau$, we introduce the probability generating functions as follows:

$$\begin{aligned}\Psi(x_1, x_2, x_3; \tau, t) = &\sum_{i_1,i_2,i_3} P(X_1(t) = i_1, X_2(t) = i_2, X_3(t) = i_3|X_1(\tau) \\ &= 0, X_2(\tau) = 0, X_3(\tau) = 0)x_1^{i_1}x_2^{i_2}x_3^{i_3}.\end{aligned} \tag{2.1}$$

and

$$\begin{aligned}\Phi_1(x_1, x_2, x_3; \tau, t) = &\sum_{i_1,i_2,i_3} P(X_1(t) = i_1, X_2(t) = i_2, X_3(t) = i_3|X_1(\tau) \\ &= 1, X_2(\tau) = 0, X_3(\tau) = 0)x_1^{i_1}x_2^{i_2}x_3^{i_3}.\end{aligned} \tag{2.2}$$

and

$$\Phi_2(x_2, x_3; \tau, t) = \sum_{i_2,i_3} P(X_2(t) = i_2, X_3(t) = i_3|X_2(\tau) = 1, X_3(\tau) = 0)x_2^{i_2}x_3^{i_3}. \tag{2.3}$$

where Ψ and Φ_i, $(i = 1, 2)$ represent the probability generating functions regarding the number of premalignant stem cells and totally malignant cells at time t beginning from zero intermediate cell in the compartment I_i to one intermediate cell in the compartment I_i at time τ, respectively. All these generating functions are subject to the Kolmogorov backward equations [40].

$$\frac{d\Psi(x_1, x_2, x_3; \tau, t)}{d\tau} = -\nu * N * \Psi(x_1, x_2, x_3; \tau, t)[\Phi_1(x_1, x_2, x_3; \tau, t) - 1]. \tag{2.4}$$

$$\begin{aligned}\frac{d\Phi_1(x_1, x_2, x_3; \tau, t)}{d\tau} = &[\alpha_1 + \beta_1 + \mu_1]\Phi_1(x_1, x_2, x_3; \tau, t) - \alpha_1{\Phi_1}^2(x_1, x_2, x_3; \tau, t) \\ &- \mu_1\Phi_1(x_1, x_2, x_3; \tau, t)\Phi_2(x_2, x_3; \tau, t) - \beta_1.\end{aligned} \tag{2.5}$$

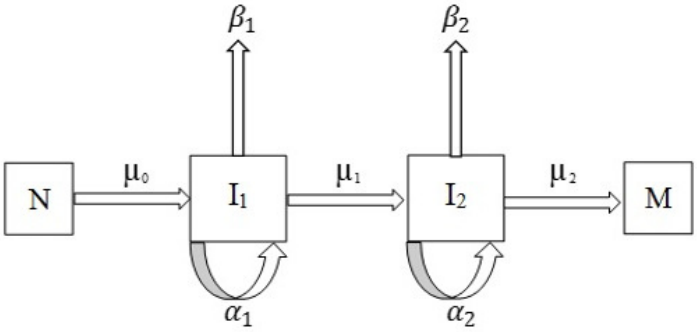

Figure 1. The schematic representation of stochastic 3-stage model for carcinogenesis. Let $I_i(i = 1, 2)$ denote the compartments of intermediate cells, $\mu_i(i = 1, 2)$ represent the mutation rates per cell per year, and $\mu_0 = \nu * N$ such that N be the total number of normal stem cells and ν denotes the initial mutation rate per cell per year. α_i and β_i are the birth rates and death or differentiation rates per cell per year in the compartment $I_i(i = 1, 2)$, respectively.

and

$$\frac{d\Phi_2(x_2, x_3; \tau, t)}{d\tau} = [\alpha_2 + \beta_2 + \mu_2]\Phi_2(x_2, x_3; \tau, t) - \alpha_2 {\Phi_2}^2(x_2, x_3; \tau, t) - \mu_2 x_3 \Phi_2(x_2, x_3; \tau, t) - \beta_2. \tag{2.6}$$

We define the following probability generating functions $\Psi(x_1, x_2, x_3; \tau, t)$, $\Phi_1(x_1, x_2, x_3; \tau, t)$, and $\Phi_2(x_2, x_3; \tau, t)$ at the point $(x_1, x_2, x_3) = (1, 1, 0)$, as follows $\Psi(\tau, t)$, $\Phi_1(\tau, t)$, and $\Phi_2(\tau, t)$ respectively. By doing so, we can obtain the following equation system at the point $(x_1, x_2, x_3) = (1, 1, 0)$,

$$\frac{d\Psi(\tau, t)}{d\tau} = -\nu * N * \Psi(\tau, t)[\Phi_1(\tau, t) - 1]. \tag{2.7}$$

$$\frac{d\Phi_1(\tau, t)}{d\tau} = [\alpha_1 + \beta_1 + \mu_1]\Phi_1(\tau, t) - \alpha_1 {\Phi_1}^2(\tau, t) - \mu_1 \Phi_1(\tau, t)\Phi_2(\tau, t) - \beta_1. \tag{2.8}$$

and

$$\frac{d\Phi_2(\tau, t)}{d\tau} = [\alpha_2 + \beta_2 + \mu_2]\Phi_2(\tau, t) - \alpha_2 {\Phi_2}^2(\tau, t) - \beta_2. \tag{2.9}$$

with the following boundary conditions:
$\Psi(t, t) = 1$, $\Phi_i(t, t) = 1$ $(i = 1, 2)$,

The cancer incidence is characterized by hazard function h(t), which can be written as $h(t) = \Psi'(0, t)/\Psi(0, t)$, but in the 3-stage stochastic model, we are interested in determining the probability p(t). By defining the probability generating function $\Psi(0, t)$, the probability of at least one

malignant cell by the time t, starting with only normal cells at time 0, can be formulated as

$$p(t) = 1 - \Psi(0, t) \tag{2.10}$$

In order to comprehensively calculate the probability function p(t), we must find a solution of the generating function $\Psi(0, t)$. The function $\Psi(0, t)$ does not have a closed-form solution by equations (2.7), (2.8), and (2.9). To solve the system of equations numerically, we assume $s = t - \tau$, $A(s, t) = \Psi(\tau, t)$, $B_1(s, t) = \Phi_1(\tau, t)$, and $B_2(s, t) = \Phi_2(\tau, t)$.

Then, for $(i = 1, 2)$, the equations (2.7), (2.8), and (2.9) satisfy following set of differential equations:

$$\begin{aligned}
\frac{dA(s,t)}{ds} &= \nu * N * A(s,t)[B_1(s,t) - 1] \\
\frac{dB_1(s,t)}{ds} &= -[\alpha_1 + \beta_1 + \mu_1]B_1(s,t) + \alpha_1 {B_1}^2(s,t) + \mu_1 B_1(s,t)B_2(s,t) + \beta_1 \\
\frac{dB_2(s,t)}{ds} &= -[\alpha_2 + \beta_2 + \mu_2]B_2(s,t) + \alpha_2 {B_2}^2(s,t) + \beta_2
\end{aligned} \tag{2.11}$$

with initial conditions:

$$\begin{aligned}
A(0,t) &= 1 \\
B_1(0,t) &= 1 \\
B_2(0,t) &= 1
\end{aligned} \tag{2.12}$$

However, the probability function can be written as $p(t) = 1 - A(t, t)$.

2.4. Computer simulation

In our study, we apply a mathematical model that only takes into account three gene mutations for the progression of lung cancer [41, 42]. We use the fourth-order Runge-Kutta method to numerically solve the differential equation system (2.11) with respect to s (for fixed t) directly. By using the numerical optimization routine *f*minsearch in MATLAB, we estimate the optimal parameters using the minimum sum of error squares, which is a very strong tool for estimating the optimal parameters. This method is very similar to the global optimum of the genetic algorithm. Here the error defined as the difference between the data and the simulated values from the probability function calculated in the model.

2.5. Statistic analysis

In order to check some biological mutation mechanisms of human lung cancer such as the impact of cigarette smoking on lung cancer development, we assumed that the principle concerning mutator phenotype is mandatory to achieve good numerical results where the mutation rates per cell per year in the 3-hit stochastic model should satisfy the formula $\nu \leq \mu_1 \leq \mu_2$ [43]. The mutator phenotype hypothesis has an outstanding performance in carcinogenesis [44]. In addition, we considered the growth (birth) cell rates, the death or differentiation rates, and the mutations rates as parameters in each compartment. However, epidemiological data do not give us full details about all the model parameters. There are two popular options we can use to handle this non-identifiability problem. First,

we can set some parameters equal to each other and suppose a plausible number of normal stem cells, $N = 1 * 10^7$ [45]. Second, we can use a new collection of parameters [46] and assume a fixed value for the parameter. Since the model is not sensitive to the birth and death rates, we deal with the clonal expansion rates. Then, we adopted the value for birth rates $\alpha_i (i = 1, 2)$ proposed by Meza et al. [47] at about 9 per year.

To study the mechanisms of lung cancer in the US and the effects of cigarette smoking in advancing lung tumorigenesis, by analyzing age-specific incidence rates in the SEER registry and the smoking data (NHS cohort of women and the HPFS cohort of men), we need to use statistical inference by executing the Chi-square test which is chosen to examine our optimal fitting. In our work, we represent the probabilities of lung cancer incidence using the following formula:

$$P(t) = 1 - \exp[-\int_0^t h(s)ds] \tag{2.13}$$

where $h(s)$ represents the incidence function (hazard function) at age s. Since the probability values are always small; we handle this problem by amplifying $P(t)$ with a suitable multiple (i.e., 10^4). We get 45 probability data points that are greater than one in the SEER data, and 40 probability data points of smoking data among current and former smokers. The optimal parameters can be calculated by using the numerical optimization routine $f$minsearch in MATLAB within the Chi-square test between the real data and simulated ones. The goodness of fit is determined by the Chi-square test which can be implemented within this relation:

$$Chi2 = \sum_j (data_{p(j)} - simulation_{p(j)})^2 * 10^4 / simulation_{p(j)} \tag{2.14}$$

where the index j represents the age such that $j = 35, ..., 79$ in the case of SEER data, and $j = 40, ..., 79$ for smoking data. The chi-square goodness of fit is applied for approving or refusing the null hypothesis by comparing our calculations with those obtained by statistical distribution at 5% significant level.

3. Results

Results from the three-hit stochastic model in the SEER program are presented. The optimal model with smoking data is discussed in more detail and the significance of the different pathways in the model is investigated. Furthermore, the effects of smoking characteristics in lung carcinogenesis are described.

3.1. The SEER fitting

In the present study, chi-square test was employed to examine whether outcomes from the optimal model were available. The calculations were performed by the numerical optimization routine $f$minsearch in MATLAB. The optimal fitting for the three-stage model using the age probability data of lung cancer incidence in the SEER program from 1992 to 2010 is demonstrated in Figure 2 (for male patients) and Figure 3 (for female patients). It is clear that the models with three mutations fit the

SEER data well among male and female patients, and these models provide a very small error. The error is the difference between real data and simulated data.

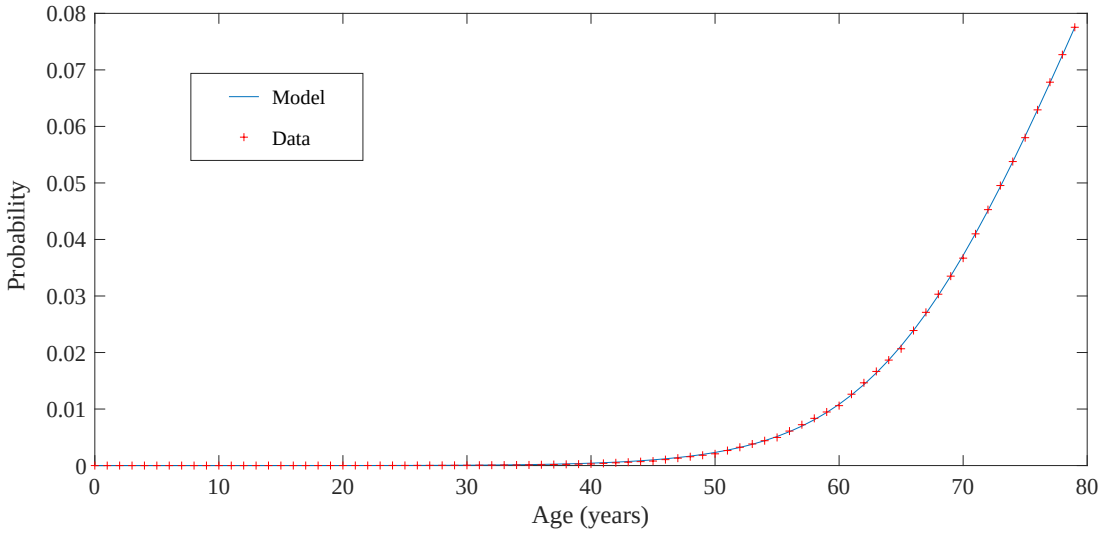

Figure 2. The three stage simulation of the probability of lung cancer incidence rates from the SEER registry between 1992 and 2010 years for male patients.

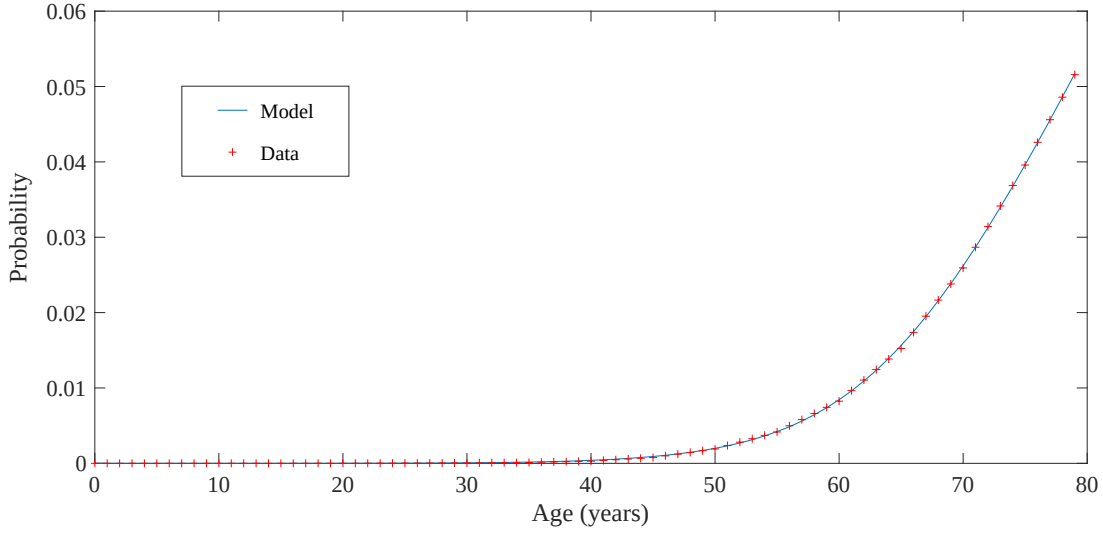

Figure 3. The same as figure 2 but for female patients.

Table 1. The values of optimal parameters calculated by using the SEER data for male patients and female patients in the 3-hit model.

Parameters	Definition	Male	Female
$\alpha_1 - \beta_1$	Net proliferation rate in premalignant cells, I_1	0.1646	0.1574
$\alpha_2 - \beta_2$	Net proliferation rate in premalignant cells, I_2	0.2307	0.4718
ν	Initial mutation rate in the normal stem cells, N	$3.3253 * 10^{-8}$	$2.0932 * 10^{-8}$
μ_1	Mutation rate in the first intermediate cells	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$
μ_2	Mutation rate in the second intermediate cells	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$

The optimal parameters of clonal expansion rates and mutation rates for the three hit model are shown in Table 1. We can observe that the intermediate cells do not exhibit very large values of clonal expansion rates for both men and women in US, and the initial mutation rate in normal cells is small (multiple of 10^{-8}). It was found that the intermediate cells with two mutations have higher clonal expansion rates than those with one mutation in both genders. The clonal expansion rates of cells with two mutations for women are greater than those for men, and the reverse is also true of cells with one mutation. However, the mutation rates of normal cells and intermediate cells for American women are smaller than those for American men. Thus, gender differences have a great ability in promoting the progression of lung cancer.

3.2. The impact of cigarette smoking

It is widely known that environmental factors such as cigarette smoking play a major role in increasing the risk of lung carcinogenesis among male and female patients. However, detailed effects of cigarette smoking on lung cancer are not clear, especially from the mathematical point of view. So for our study, we supposed that cigarette smoking affects the mutation rates and clonal expansion rates. In order to study the influence of cigarette smoking on lung cancer evolution, we assumed that cigarette smoking changes the mutations rates and the net proliferation rates. we studied these effects by using a three-stage stochastic model, which allows the clonal expansion rates to enter in each compartment, to fit the smoking data (NHS cohort of women and the HPFS cohort of men) from 1986 to 2000. We chose the SEER data and smoking data as test systems. In these systems, the fitting of the SEER data for the 3-hit model is considered as the control group. We divided the smoking data based on smoking features such as smoking status (current/former smokers), intensity or dose (<25 and ≥ 25 cigarettes per day), age at start of smoking (≥ 20 and ≤ 19 years), and time since quitting (≥ 10 and ≤ 10 years).

To check the impact of cigarette smoking on genetic mutation and clonal expansion of cells, the data from NHS cohort of women and the HPFS cohort of men were examined by modifying values of model parameters in the control group. To fit the smoking data to a model, we consider the following two alternative hypotheses:

Hypothesis one: Cigarette smoking influences mutation rates of cells in the process of carcinogenesis. Thus, the growth rates and the death or differentiation rates of cells in the model are constant (determined by the control group), but the mutation rates of cells should be greater than those of the control group.

Hypothesis two: Cigarette smoking has an effect on the clonal expansion of cells in advancing lung cancer. Thus, the mutation rates of cells in the model are constant (determined by the control group),

but the net proliferation rates of cells should be greater than those of the control group.

For current smokers, Table 2 offers the values of Chi-square test among men and women by using smoking data defined by smoking intensity (< 25 and ≥ 25 cigarettes per day) and age at start of smoking (≥ 20 and ≤ 19 years) , and these results showed that the two hypotheses are acceptable at 5 % significant level. In addition, hypothesis one has smaller Chi-square results than those under hypothesis two in all instances of smoking data among men and women in US. Then, the 3-hit mathematical model under Hypothesis one provides a better description of the smoking data than the model under hypothesis two among current smokers for both male and female patients. In the case of age at start of smoking ≥ 20 years, the Chi-square statistics results of the two hypotheses for smoking intensity (< 25 cigarettes per day) are lower than those of smoking intensity (≥ 25 cigarettes per day) among men and women. The same pattern occurred in the attribute for age at start of smoking ≤ 19 years of the two hypotheses among men and female patients, but the situation is different for male patients under hypothesis one.

The values of parameters of the 3-hit stochastic model under hypothesis one for both male and female patients among current smokers are displayed in Table 3. By comparing the optimal results of parameters in Table 1 and Table 3, it is apparent that the initial mutation rate in normal cells ν and mutation rate in intermediate cells I_2, μ_2 are higher than those obtained by the control group while the mutation rate in intermediate cells I_1, μ_1 remains almost equal to the control group of male patients in all cases of smoking intensity < 25 cigarettes per day (age at start of smoking ≥ 20 and ≤ 19 years). Regarding smoking intensity ≥ 25 cigarettes per day, only the initial mutation rate in normal cells ν is increased for male patients in the case of age at start of smoking ≥ 20 years. In addition, all the mutation rate values are higher than those found in the control group for age at start of smoking ≤ 19 years among American men.

On the other hand, in female patients, the only rise was in all mutation rates that occur in initial mutation rate in normal cells ν, while the other mutation rates μ_1 and μ_2 remained unchanged in all other cases of smoking data. It was observed that the initial mutation rate ν increased for both male and females in US, and the following smoking data (smoking intensity ≥ 25 and age at start of smoking ≤ 19 years) have the highest effect in inducing mutation rates. Therefore, under the exposure of cigarette smoking, mutation is more likely to happen in normal cells, which may cause lung cancer.

Table 4 exhibits the optimal parameters of the three stage model under hypothesis two for both American men and women among current smokers. By taking Tables 1 and 4 into consideration, we found that the only effect of cigarette smoking in the progression of lung cancer appears in the values of net proliferation rates of intermediate cells in the compartment I_1 while the clonal expansion rates of intermediate cells in the compartment I_2 remain fixed (determined by the control group) for both male and female patients in all cases of smoking data. In addition, the optimal values of net proliferation rate of intermediate cells in the compartment I_1 for smoking intensity ≥ 25 cigarettes per day are greater than those of smoking dose < 25 among men and women. To be more specific, these values for smoking data (age at start of smoking ≥ 20 years) are higher than those for smoking data (age at start of smoking ≤ 19 years) among male and female patients. In general, the values of clonal expansion rates of premalignant cells in the compartment I_1 for male patients are smaller than those for female patients. Thus, the role of cigarette smoking on clonal expansion of stem cells is quite clearly only in the event of female patients.

Table 2. The chi-square values of the two hypotheses for both male and female among current smokers. Chi2 denotes the value of Chi-square; df(=40-number of parameters-1) denotes the degrees of freedom; S.5% denotes the value at 5% significant level.

	Smoked cigarettes per day < 25								Smoked cigarettes per day ≥ 25							
	Age at start ≥ 20				Age at start ≤ 19				Age at start ≥ 20				Age at start ≤ 19			
	Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Chi2	0.4452	0.8609	0.6386	4.6293	1.8867	0.8293	2.1529	5.3004	5.6945	2.2761	5.7021	34.8465	0.5701	1.4231	19.9276	35.9186
df	36	36	37	37	36	36	37	37	36	36	37	37	36	36	37	37
S.5%	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192

Table 3. The optimal parameters of 3-hit model under hypotheses one for male and female among current smokers. The values of net proliferation rates are the same as those in Table 1.

	Smoked cigarettes per day < 25				Smoked cigarettes per day ≥ 25			
	Age at start ≥ 20		Age at start ≤ 19		Age at start ≥ 20		Age at start ≤ 19	
	Mutation rates		Mutation rates		Mutation rates		Mutation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
ν	$4.3547 * 10^{-8}$	$4.8742 * 10^{-8}$	$5.0136 * 10^{-8}$	$5.4665 * 10^{-8}$	$8.5741 * 10^{-8}$	$1.0453 * 10^{-7}$	$1.2049 * 10^{-7}$	$1.0638 * 10^{-7}$
μ_1	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7390 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$4.3228 * 10^{-5}$	$1.7389 * 10^{-5}$
μ_2	$4.4372 * 10^{-5}$	$1.7391 * 10^{-5}$	$4.3564 * 10^{-5}$	$1.7393 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$	$4.4762 * 10^{-5}$	$1.7391 * 10^{-5}$

Table 4. The optimal parameters of 3-hit model under hypotheses two for male and female among current smokers. The values of cell mutation rates are the same as those in Table 1.

	Smoked cigarettes per day < 25				Smoked cigarettes per day ≥ 25			
	Age at start ≥ 20		Age at start ≤ 19		Age at start ≥ 20		Age at start ≤ 19	
	Net proliferation rates		Net proliferation rates		Net proliferation rates		Net proliferation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
$\alpha_1 - \beta_1$	0.1786	0.2001	0.1850	0.2077	0.2087	0.2577	0.2437	0.2597
$\alpha_2 - \beta_2$	0.2307	0.4718	0.2307	0.4718	0.2307	0.4718	0.2307	0.4718

We have checked the impact of cigarette smoking on lung cancer development among former smokers for both male and female patients. For this goal, the smoking characteristic (time elapsed since quitting smoking) is strongly related to risk of cancer, and was therefore added as another layer of categorization. The chi-square test values of the two hypotheses among former smokers for both male and female patients for smoking doses < 25 and ≥ 25 are shown in Tables 5 and 6, respectively. For smoking frequency < 25, it is clear that the male patients have Chi-square results larger than those of female patients in the two hypotheses regardless of the values of age at start of smoking and time since cessation. The same situation happened for smoking intensity ≥ 25 only if the time since quitting smoking is greater than ten years.

Table 5. The chi-square values of the two hypotheses for both male and female among former smokers for dose < 25. Chi2 denotes the value of Chi-square; df(=40-number of parameters-1) denotes the degrees of freedom; S.5% dentes the value at 5% significant level.

	Age at start ≥ 20								Age at start ≤ 19							
	Quit ≥ 10 years ago				Quit < 10 years ago				Quit ≥ 10 years ago				Quit < 10 years ago			
	Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Chi2	41.9032	23.6001	41.9032	23.6001	19.4419	1.8991	19.4419	1.9768	32.5641	19.9401	32.5641	19.9402	5.5380	2.1414	5.5380	3.1067
df	36	36	37	37	36	36	37	37	36	36	37	37	36	36	37	37
S.5%	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192

Table 6. The same as Table 5 but for dose ≥ 25.

	Age at start ≥ 20								Age at start ≤ 19							
	Quit ≥ 10 years ago				Quit < 10 years ago				Quit ≥ 10 years ago				Quit < 10 years ago			
	Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two		Hypothesis one		Hypothesis two	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Chi2	27.7449	7.4621	27.7449	7.4617	2.1877	2.4749	5.0233	3.1214	6.2457	1.1432	6.2458	1.1432	3.4858	7.1057	9.5542	26.5997
df	36	36	37	37	36	36	37	37	36	36	37	37	36	36	37	37
S.5%	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192	50.998	50.998	52.192	52.192

In the case of time since quitting ≥ 10 years ago, the Chi-square values under hypothesis one for both male and female patients are equal to Chi-square values under hypothesis two, whether the doses are less than 25 or greater than 25. Based on smoking intensity < 25 and quitting time < 10 years ago, the Chi-square results for American men under hypothesis one are similar to ones under hypothesis two, while the Chi-square values for American women under hypothesis two are higher than those obtained by hypothesis one. On the opposite side, for smoking dose ≥ 25 and cessation time < 10 years ago, the Chi-square test shows statistical results for both male and female patients in US under hypothesis one as lower than those calculated under hypothesis two. Tables 7 and 8 display the parameters of the three-stage model under hypothesis one for both male and female patients among former smokers for smoking intensity < 25 and ≥ 25, respectively. By observing these values with results of parameters in Table 1, we can see that for smoking dose < 25, there is only a slight increase in the initial mutation rate ν in normal cells for female patients if the time since quitting is less than ten years, while all the mutation rates in premalignant cells remain as they are calculated in Table 1 for all durations of time since quitting smoking (< 10 or ≥ 10 years ago). Looking at the other side, dose ≥ 25, the change is very small compared to the previous situation where both males and females have higher initial mutation rate in normal cells than shown in Table 1 only if the time since quitting is smaller than ten years. Furthermore, the values of initial mutation rate in the smoking data (age at start ≤ 19 years and quit < 10 years ago) are greater than those determined by considering age at start of smoking ≥ 20 years ago. Thus, the effect of cigarette smoking on mutation rates of cells for both male and female patients among former smokers appears only in the initial mutation rate in normal cells when the duration of smoking cessation is lower than ten years. If the duration is greater than ten years we can say that the effect is very weak.

On the other hand, the optimal values of the 3-hit model under hypothesis two for both male and female patients among former smokers for doses < 25 and ≥ 25 are presented in Tables 9 and 10, respectively. For smoking doses < 25 and ≥ 25, it is easy to observe that the values of net proliferation rates of intermediate cells in the compartment I_2 stay unchanged (determined by the control group)

among men and women, while the clonal expansion rates of premalignant cells in the compartment I_1 increase only when the time since quitting smoking is less than ten years for both male and female patients. We can see that the impact of cigarette smoking on clonal expansion and mutation rates of cells among current smokers is very noticeable, especially the age at start of smoking $\leq$ 19 years for both male and female patients. On the opposite side, this effect is weak among former smokers.

Table 7. The optimal parameters of 3-hit model under hypotheses one for male and female among former smokers for dose < 25. The values of net proliferation rates are the same as those in Table 1.

	Age at start $\geq$ 20				Age at start $\leq$ 19			
	Quit $\geq$ 10 years ago		Quit < 10 years ago		Quit $\geq$ 10 years ago		Quit < 10 years ago	
	Mutation rates		Mutation rates		Mutation rates		Mutation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
ν	$3.3253 * 10^{-8}$	$2.0932 * 10^{-8}$	$3.3253 * 10^{-8}$	$2.5601 * 10^{-8}$	$3.3253 * 10^{-8}$	$2.0932 * 10^{-8}$	$3.3253 * 10^{-8}$	$3.5277 * 10^{-8}$
μ_1	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7421 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$
μ_2	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7528 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$

Table 8. The same as Table 7 but for dose $\geq$ 25.

	Age at start $\geq$ 20				Age at start $\leq$ 19			
	Quit $\geq$ 10 years ago		Quit < 10 years ago		Quit $\geq$ 10 years ago		Quit < 10 years ago	
	Mutation rates		Mutation rates		Mutation rates		Mutation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
ν	$3.3253 * 10^{-8}$	$2.0932 * 10^{-8}$	$6.0718 * 10^{-8}$	$2.6943 * 10^{-8}$	$3.3253 * 10^{-8}$	$2.0932 * 10^{-8}$	$7.4699 * 10^{-8}$	$7.3755 * 10^{-8}$
μ_1	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$	$3.7072 * 10^{-5}$	$1.7389 * 10^{-5}$
μ_2	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$	$3.7115 * 10^{-5}$	$2.21671 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$	$3.7115 * 10^{-5}$	$1.7391 * 10^{-5}$

Table 9. The optimal parameters of 3-hit model under hypotheses two for male and female among former smokers for dose < 25. The values of cell mutation rates are the same as those in Table 1.

	Age at start $\geq$ 20				Age at start $\leq$ 19			
	Quit $\geq$ 10 years ago		Quit < 10 years ago		Quit $\geq$ 10 years ago		Quit < 10 years ago	
	Net proliferation rates		Net proliferation rates		Net proliferation rates		Net proliferation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
$\alpha_1 - \beta_1$	0.1646	0.1574	0.1646	0.1662	0.1646	0.1574	0.1646	0.1814
$\alpha_2 - \beta_2$	0.2307	0.4718	0.2307	0.4718	0.2307	0.4718	0.2307	0.4718

Table 10. The same as Table 9 but for dose ≥ 25.

	Age at start ≥ 20				Age at start ≤ 19			
	Quit ≥ 10 years ago Net proliferation rates		Quit < 10 years ago Net proliferation rates		Quit ≥ 10 years ago Net proliferation rates		Quit < 10 years ago Net proliferation rates	
	Male	Female	Male	Female	Male	Female	Male	Female
$\alpha_1 - \beta_1$	0.1646	0.1574	0.1911	0.1806	0.1646	0.1574	0.2022	0.2255
$\alpha_2 - \beta_2$	0.2307	0.4718	0.2307	0.8109	0.2307	0.4718	0.2307	0.4718

4. Conclusion

The impact of cigarette smoking on clonal expansion of premalignant cells and mutations rates of cells in causing lung cancer was presented in this study. All numerical calculations were performed by using the 3-hit model based on the mutator phenotype hypothesis. The calculations were obtained by the numerical optimization routine *f*minsearch in Matlab. The smoking data from the Nurses' Health Study cohort of females (NHS) and the Health Professionals Follow up Study cohort of men (HPFS) were used to infer the key parameters in the optimal model. Furthermore, cigarette smoking data were divided into four categories (smoking status/age at start of smoking/intensity/time since cessation). We have used the SEER data as the control group to indicate the parameters of the optimal model for both male and female patients in the absence of cigarette smoking. Therefore, the NHS cohort of females and HPFS cohort of men were used to infer another group of optimal model parameters. our simulations showed that there is a gender difference in the evolution of lung cancer. The inferred key parameters of the optimal model give insightful predictions regarding the impact of cigarette smoking on the progression of lung cancer among current and former smokers.

To study the effect of smoking on the progression of lung cancer, we have tested the influence of cigarette smoking on the mutation rates of cells or on the clonal expansion rates of premalignant cells among current and former smokers in US. We applied the three-hit model with cigarette smoking data to calculate the optimal parameters under different hypotheses. The Chi-square statistics results in Tables 2, 5, and 6 proposed that cigarette smoking affects not only the mutation rates of cells but also the net proliferation of cells in lung cancer.

According to the mutation rates of cells, our findings suggested that cigarette smoking induces the initial mutation rate in the normal cells and the mutation rate in the premalignant cells I_2 for American men, while only the initial mutation rate in the normal cells increased for American women among current smokers. For former smokers, the effect of cigarette smoking on mutation rates of cells is absent for male patients and very weak for female patients in the case of smoking frequency < 25. Moreover, cigarette smoking has a slight effect on initial mutation rates of cells for both male and female patients only if the smoking intensity ≥ 25 and the time since quitting ≥ 10 years ago. Regarding the clonal expansion of cells, cigarette smoking has no effect on the net proliferation rates of cells in the compartment I_2 for both male and female patients among current and former smokers, which supports the results of some previous works [33, 34]. The impact of cigarette smoking on the net proliferation rates of cells in the compartment I_1 for female patients is greater than that of male patients among current smokers. For former smokers, this effect is low for women only when smoking intensity is less than 25 cigarettes per day. When smoking intensity increases to more than 25 cigarettes

per day, the effect becomes apparent for both sexes only if the the time since quitting < 10 years ago. However, there are no effects of cigarette smoking on mutation rates of cells or on clonal expansion of intermediate cells among former smokers for both male and female patients if they have quit smoking for over ten years. In general, the effect of cigarette smoking on lung cancer is directly proportional to the smoking intensity (cigarettes smoked per day). The lower the age at start of smoking, the greater the impact of cigarette smoking on lung cancer progression. Finally, we conclude that gender significantly impacts the evolution of lung carcinogenesis.

Acknowledgments

This research was partially supported by the National Nature Science Foundation of China (No.11871238, No.11371161), by self-determined research funds of CCNU from the colleges' basic research and operation of MOE (Grant No.CCNU16JCZX10, No.CCNU16A02009).

Conflict of interest

The authors declare no potential conflicts of interest.

References

1. D. M. Parkin, F. Bray, J. Ferlay, et al., Global cancer statistics, 2002 , *CA Cancer J. Clin.*, **55** (2005), 74–108.
2. R. L. Siegel, K. D. Miller and A. Jemal, Cancer statistics, 2015, *CA Cancer J. Clin.*, **65** (2015), 5–29.
3. M. Malvezzi, G. Carioli, P. Bertuccio, et al., European cancer mortality predictions for the year 2016 with focus on leukaemias, *Ann. Oncol.*, **27** (2016), 725–731.
4. J. Ferlay, I. Soerjomataram, Dikshit, et al., Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, *Int. J. Cancer*, **136** (2015), E359–E386.
5. J. Malhotra, M. Malvezzi, E. Negri, et al., Risk factors for lung cancer worldwide, *Eur. Resp. J.*, **48** (2016), 889–902.
6. S. Sun, J. H. Schiller and A. F. Gazdar, Lung cancer in never smokersa different disease, *Nat. Rev. Cancer*, **7** (2007), 778–790.
7. A. Blair and L. B. Freeman, Lung cancer among nonsmokers, *Epidemiology*, **17** (2006), 601–613.
8. R. W. Field and B. L. Withers, Occupational and environmental causes of lung cancer, *Clin. Chest Med.*, **33** (2012), 681–703.
9. C. S. D. Cruz, L. T. Tanoue and R. A. Matthay, Lung cancer: Epidemiology, etiology, and prevention, *Clin. Chest Med.*, **32** (2011), 605–644.
10. S. Wu, S. Powers, W. Zhu, et al., Substantial contribution of extrinsic risk factors to cancer development, *Nature*, **529** (2016), 43–47.
11. G. Ray, D. E. Henson and A. M. Schwartz, Cigarette smoking as a cause of cancers other than lung cancer: An exploratory study using the Surveillance, Epidemiology, and End Results Program, *Chest*, **138** (2010), 491–499.

12. S. D. Coello, A. C. De Len, F. B. Ojeda, et al., High density lipoprotein cholesterol increases with living altitude, *Int. J. Epidemiol.*, **29** (2000), 65–70.

13. V. L. Roger, M. E. Farkouh, S. A. Weston, et al., Sex differences in evaluation and outcome of unstable angina, *JAMA*, **283** (2000), 646–652.

14. M. Mori, M. Hara, I. Wada, et al., Prospective study of hepatitis B and C viral infections, cigarette smoking, alcohol consumption, and other factors associated with hepatocellular carcinoma risk in Japan, *Am. J. Epidemiol.*, **151** (2000), 131–139.

15. K. Husgafvel-Pursiainen, P. Boffetta, A. Kannio, et al., p53 mutations and exposure to environmental tobacco smoke in a multicenter study on lung cancer, *Cancer Res.*, **60** (2000), 2906–2911.

16. T. L. Holmen, E. Barrett-Connor, J. Holmen, et al., Health problems in teenage daily smokers versus nonsmokers, Norway, 1995–1997: The Nord-Trndelag Health Study, *Am. J. Epidemiol.*, **151** (2000), 148–155.

17. G. Launoy, C. Milan, J. Faivre, et al., Tobacco type and risk of squamous cell cancer of the oesophagus in males: A French multicentre case-control study, *Int. J. Epidemiol.*, **29** (2000), 36–42.

18. F. Barbone, M. Bovenzi, F. Cavallieri, et al., Cigarette smoking and histologic type of lung cancer in men, *Chest*, **112** (1997), 1474–1479.

19. L. M. Nelson, V. McGuire, W. T. Longstreth Jr, et al., Population-based case-control study of amyotrophic lateral sclerosis in western Washington State. I. Cigarette smoking and alcohol consumption, *Am. J. Epidemiol.*, **151** (2000), 156–163.

20. K. Leffondr, M. Abrahamowicz, J. Siemiatycki, et al., Modeling smoking history: A comparison of different approaches, *Am. J. Epidemiol.*, **156** (2002), 813–823.

21. R. Doll and R. Peto, Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers, *J. Epidemiol. Commun. Health*, **32** (1978), 303–313.

22. W. D. Flanders, C. A. Lally, B. P. Zhu, et al., Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II, *Cancer Res.*, **63** (2003), 6556–6562.

23. C. O. Nordling, A new theory on the cancer-inducing mechanism, *British J. Cancer*, **7** (1953), 68–72.

24. P. Armitage and R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis, *British J. Cancer*, **8** (1954), 1–12.

25. I. Tomlinson and W. Bodmer, Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog, *Nat. Med.*, **5** (1999), 11–12.

26. K. Anderson, C. Lutz, F. W. Van Delft, et al., Genetic variegation of clonal architecture and propagating cells in leukaemia, *Nature*, **469** (2011), 356–361.

27. M. Gerlinger, S. Horswell, J. Larkin, et al., Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing, *Nat. Genet.*, **46** (2014), 225–233.

28. M. R. Stratton, P. J. Campbell and P. A. Futreal, The cancer genome, *Nature*, **458** (2009), 719–724.

29. L. D. Wood, D. W. Parsons, S. Jones, et al., The genomic landscapes of human breast and colorectal cancers, *Science*, **318** (2007), 1108–1113.

30. H. M. Byrne, Dissecting cancer through mathematics: from the cell to the animal model, *Nat. Rev. Cancer*, **10** (2010), 221–230.

31. S. H. Moolgavkar, A. Dewanji and G. Luebeck, Cigarette smoking and lung cancer: reanalysis of the British doctor's data, *J. Nat. Cancer Inst.*, **81** (1989), 415–420.

32. A. S. Whittemore, Effect of cigarette smoking in epidemiological studies of lung cancer, *Statist. Med.*, **7** (1988), 223–238.

33. M. Gaffney and B. Altshuler, Examination of the Role of Cigarette Smoke in Lung Carcinogenesis Using Multistage Models, *J. Nat. Cancer Inst.*, **80** (1988), 925–931.

34. S. H. Moolgavkar and A. G. Knudson, Mutation and cancer: A model for human carcinogenesis, *J. Nat. Cancer Inst.*, **66** (1981), 1037–1052.

35. R. Meza, W. D. Hazelton, G. A. Colditz, et al., Analysis of lung cancer incidence in the nurses health and the health professionals follow-up studies using a multistage carcinogenesis model, *Cancer Causes Control*, **19** (2008), 317–328.

36. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence-SEER 13 Regs Research Data, Nov 2012 Sub (1992-2010) <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2013, based on the November 2012 submission.

37. D. R. Lewis, D. P. Check, N. E. Caporaso, et al., US lung cancer trends by histologic type, *Cancer*, **120** (2014), 2883–2892.

38. C. Bain, D. Feskanich, F. E. Speizer, et al., Lung cancer rates in men and women with comparable histories of smoking, *J. Nat. Cancer Inst.*, **96** (2004), 826–834.

39. D. Hanahan and R. A. Weinberg, Hallmarks of cancer: The next generation, *Cell*, **144** (2011), 646–674.

40. T. E. Harris, *The theory of branching processes*, Prentice-Hall Applied Mathematics Series, Englewood CLiffs, NJ, 2002.

41. C. Tomasetti, L. Marchionni, M. A. Nowak, et al., Only three driver gene mutations are required for the development of lung and colorectal cancers, *Proceed. Nat. Aca. Sci.*, **112** (2015), 118–123.

42. L. Li, T. Tian and X. Zhang, The impact of radiation on the development of lung cancer, *J. Theor. Biol.*, **428** (2017), 147–152.

43. L. A. Loeb, K. R. Loeb and J. P. Anderson, Multiple mutations and cancer, *Proceed. Nat. Aca. Sci.*, **100** (2003), 776–781.

44. L. A. Loeb, Mutator phenotype may be required for multistage carcinogenesis, *Cancer Res.*, **51** (1991), 3075–3079.

45. H. Schllnberger, N. Beerenwinkel, R. Hoogenveen, et al., Cell selection as driving force in lung and colon carcinogenesis, *Cancer Res.*, **70** (2010), 6797–6803.

46. W. F. Heidenreich, E. G. Luebeck and S. H. Moolgavkar, Some properties of the hazard function of the twomutation clonal expansion model, *Risk Anal.*, **17** (1997), 391–399.

47. R. Meza, J. Jeon, S. H. Moolgavkar, et al., Age-specific incidence of cancer: Phases, transitions, and biological implications, *Proceed. Nat. Aca. Sci.*, **105** (2008), 16284–16289.

AIMS Press