Stochastic Modeling of Carcinogenesis

Rafael Meza

February 2, 2010

Abstract

Carcinogenesis is the transformation of normal cells into cancer cells. This process has been shown to be of a multistage nature, with stem cells that go through a series of genetic and epigenetic changes that eventually lead to a malignancy. Since the origins of the multistage theory in the 1950s, mathematical modeling has played a prominent role in the investigation of the mechanisms of carcinogenesis. In particular, two stochastic (mechanistic) models, the Armitage-Doll and the two-stage clonal expansion (TSCE) model, are commonly used for cancer risk assessment and the analysis of cancer epidemiology and experimental data. In this mini-course, I will introduce some of the basic biological, epidemiological and mathematical concepts behind the theory of multistage carcinogenesis, and discuss in detail the Armitage-Doll model, the TSCE model and some generalizations. The use of these models for the analysis of cancer epidemiology and experimental data will be described in detail, and some examples will be discussed.

1 Introduction

1.1 Mathematical Modeling of Carcinogenesis

Carcinogenesis is the transformation of normal cells into cancer cells. This process has been shown to be of a multistage nature, with stem cells that go through a series of (stochastic) genetic changes that eventually lead them to malignancy. In general, it is quite hard to give a precise definition of the steps of the process. However, experimental evidence suggests that most malignancies can be analyzed under the initiation-promotion-malignant conversion paradigm. Initiation refers to the emergence of altered cells prone to clonal expansion. Promotion is the outgrowth of initiated cells into premalignant lesions. Malignant conversion refers to the growth of malignant cells into tumors and the onset of the clinical disease.

Multistage models of carcinogenesis constitute a powerful quantitative framework that has played a major role in the advancement of cancer research [72, 39]. Being biologically based, multistage models allow the investigation of the effects of carcinogens on cancer initiation, promotion and malignant conversion. Hence, multistage models provide a natural framework to evaluate the potential benefits of prevention and intervention strategies designed to reduce cancer risk, like screening for premalignant lesions or the use of chemo-preventive agents. Mathematical models of carcinogenesis are stochastic in nature, and their analysis requires expertise in several areas of applied mathematics, like stochastic processes, differential equations, numerical analysis, probability and statistics.

2 Mathematical Models of Carcinogenesis

2.1 Hazard or Age-Specific Incidence Function

The standard measure for cancer risk is the hazard or age-specific incidence function of cancer. Let T be the time to malignant transformation of a particular tissue. Let $P(t) = P[T \leq t]$ be the probability

that cancer occurs before age t. Then, the hazard or incidence function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t < T \leq t + \Delta t | T > t]}{\Delta t}. \tag{1}$$

So, the hazard measures the instantaneous rate of change in cancer probability. The hazard can be expressed also in terms of the survival function $S(t) \equiv P[T > t] = 1 - P(t)$,

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{P[T > t] - P[T > t + \Delta t]}{P[T > t]} = \frac{-S'(t)}{S(t)}. \tag{2}$$

The hazard is a theoretical representation of the observed incidence or mortality of cancer in the population (# of cases / population at risk). As an example, Fig 1 shows the observed incidence of colorectal cancers among white males in the US during 1995 [67].

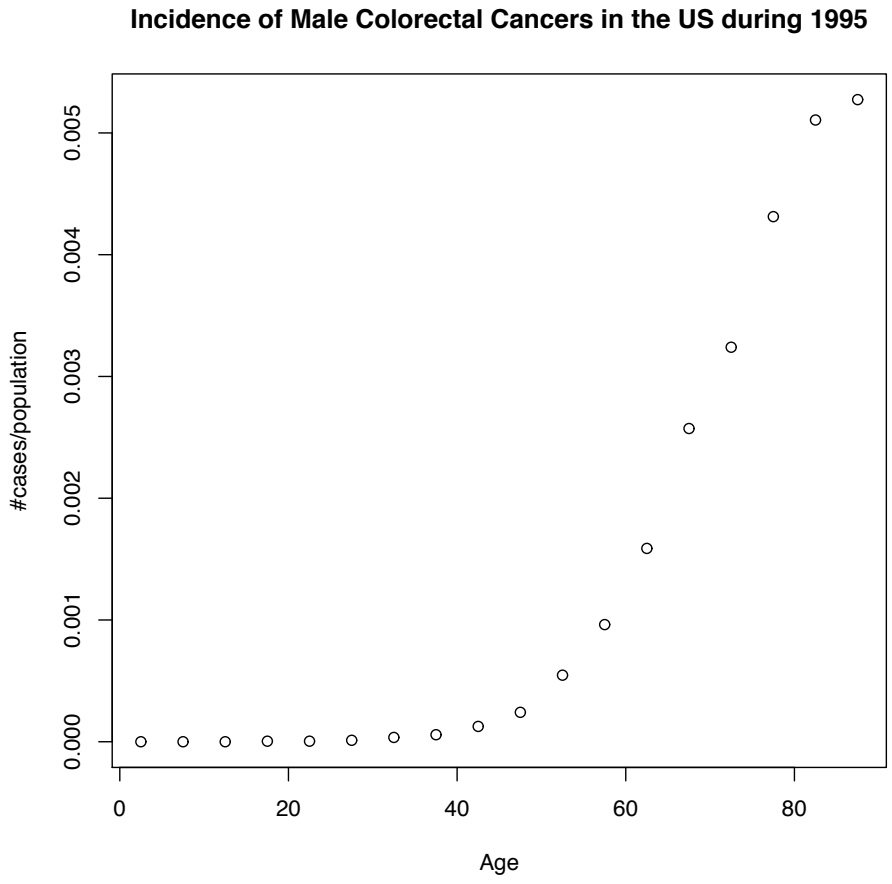

Figure 1: Incidence of colorectal cancer in the SEER database during 1995. Cases are grouped in 5-yr intervals.

3 History of Multistage Carcinogenesis Models

As mentioned in the previous chapter, multistage carcinogenesis theory states that cancer is the consequence of a series of genetic transformations that accumulate in a single cell's progeny. The multistage

theory was originally suggested independently by Muller [64] and Nordling [66]. In particular, Nordling proposed that several mutations were needed for malignant transformation in order to explain the observation that the incidence of several cancers increases according to the sixth power of age. Armitage and Doll [1] incorporated the ideas of Nordling into a stochastic model of carcinogenesis and showed that the power of age observed in cancer mortality may be related to the number of transformations required to produce a malignant cell. Armitage and Doll's is certainly one of the landmark papers in cancer epidemiology and in the mathematical modeling of carcinogenesis.

Although the Armitage–Doll model is consistent with some of the epidemiological characteristics of several cancers, it predicts a higher number of mutations required for malignant transformation than the number supported by experimental evidence[72]. Moreover, the model does not incorporate cell growth dynamics, which are known to be a key element of carcinogenesis. For these reasons, several alternative models were developed. In particular, Armitage and Doll proposed a two-stage model that assumed that intermediate-stage cells had exponential growth [2]. A similar model, proposed by Fisher [20], assumed that the growth of the intermediate cells was proportional to the square of the age of the clone. A few years later, Kendall developed a two-stage model where premalignant and malignant lesions were modeled as sub-critical and supercritical birth and death processes, respectively [36]. In his paper, Kendall derived expressions for the probability generating functions of the number of premalignant and malignant cells and discussed the problem of extrapolating the carcinogenic effects from high to lose dose exposures. A similar two-stage model was also suggested by Neymann and Scott [65] to represent the number of hyper-plastic foci originated from urethane exposures in rats. However, the model was not consistent with the experimental data and was abandoned in favor of an alternative three-stage model. None of these models gained general acceptance and the original Armitage–Doll's model remained as the main reference.

Whittemore and Keller provided a survey of the mathematical models of carcinogenesis developed until the early seventies [72]. Their work remains today as one of the main references in this research area. In their review, Whittemore and Keller described all the principal theories of carcinogenesis existing up to that moment and expressed them in a unified mathematical framework. They also introduced their own theory of carcinogenesis, which was the first one to incorporate explicitly the initiation-promotion paradigm into the mathematical theories of carcinogenesis.

A few years before the publication of Whittemore and Keller's review, Knudson proposed a two-stage statistical model to explain the incidence of retinoblastoma in children [37]. From his model, Knudson derived the two-hit hypothesis, which states that the two-alleles of a single gene have to be inactivated to produce a tumor in the retina. This idea lead to the concept of tumor suppressor genes, which was corroborated in the laboratory a few years later [21]. Knudson's work is without a doubt one of the most important contributions of mathematical and statistical modeling to cancer research. A few years later, Moolgavkar and Venzon incorporated the ideas of Knudson into a mathematical model, which is consistent with the incidence of most cancers in both children and adults [62]. Similarly to Kendall and Neymann and Scott, Moolgavkar and Venzon modeled the clonal expansion of premalignant cells as a birth-death process. However they also incorporated the dynamics of normal stem cells into the picture, a feature that turned out to be essential for the model to be able to describe the incidence of cancers in children [62, 56]. Using the theory of filtered Poisson processes, Moolgavkar and Venzon were able to obtain general expressions for the hazard and survival curves of their model. As shown in Moolgavkar and Knudson [56], the now called two-stage clonal expansion (TSCE) model was the first carcinogenesis model to be consistent with the epidemiology of most cancers and also with the data from carcinogen experiments.

Since the original work of Moolgavkar and Venzon and Moolgavkar and Knudson, the TSCE model has been widely used for the analysis of epidemiological [58, 43, 35, 28, 6, 49, 32, 10, 30, 25, 27, 26, 68, 52, 13, 22] and experimental [54, 60, 47, 23, 46] data. In addition, several mathematical extensions of the model have been proposed to accommodate the particular characteristics of different cancers. In the following sections, we describe some of the mathematical details of multistage models and, in particular, derive some of the principal mathematical results of the Armitage–Doll and the TSCE model.

4 The Armitage–Doll Model

Fifty years after its original publication, the Armitage–Doll model [1] remains one of the standard references in cancer epidemiology. In this section, we briefly describe the model and derive the hazard and survival functions of cancer under this model.

Following the ideas of Nordling, Armitage and Doll postulated that a malignant tumor arises in a particular tissue when a *single* susceptible cell undergoes a series of sequential transformations that eventually lead to malignancy. A schematic representation of the model is given by

$$E_0 \overset{\lambda_0}{\to} E_1 \overset{\lambda_1}{\to} E_2 \overset{\lambda_2}{\to} ... \overset{\lambda_{n-1}}{\to} E_n, \tag{3}$$

where E_0 represents the normal stage, E_n the malignant stage, E_k the intermediate stages and λ_k the transition rates between stages, $k = 0, ..., n-1$. Armitage and Doll assumed that the time spent by the susceptible cell in each stage is an exponential random variable with mean $1/\lambda_k, k = 1, ..., n-1$.

We derive now the hazard and survival function of the Armitage–Doll model. Let $p_k(t) \equiv$ Prob[the cell is in stage k by time t]. We are particularly interested in $p_n(t)$, the probability that the cell has gone through malignant transformation by time t. The Armitage–Doll model can be seen as a pure-birth or Yule process with an absorbing boundary at stage n. This implies that the probability of being in state k at time $t+h$, $p_k(t+h)$, can be expressed, using the Markov property and the Chapman-Kolmogorov equations, as

$$\begin{aligned} p_0(t+h) &= (1-h\lambda_0)p_0(t) + o(h), \\ &\vdots \\ p_k(t+h) &= (1-h\lambda_k)p_k(t) + h\lambda_{k-1}p_{k-1}(t) + o(h), \\ &\vdots \\ p_n(t+h) &= p_n(t) + h\lambda_{n-1}p_{n-1}(t) + o(h). \end{aligned} \tag{4}$$

Dividing by h and taking the limit when h goes to zero, we obtain the system of differential equations:

$$\begin{aligned} p_0'(t) &= -\lambda_0 p_0(t), \\ &\vdots \\ p_k'(t) &= -\lambda_k p_k(t) + \lambda_{k-1}p_{k-1}(t), \\ &\vdots \\ p_n'(t) &= \lambda_{n-1}p_{n-1}(t). \end{aligned} \tag{5}$$

The solution of system (5) can be easily computed by successive integrations or by calculating the exponential of the matrix Q, where we write the system (5) as $P' = QP$.

Let us assume that the number of cells that can go through malignant transformation is N and that $T_1, T_2, .., T_N$, are independent random variables, representing the time to malignancy of each cell. Clearly, the time to malignant transformation of the tissue, T, is the minimum of the malignant transformation times of all cells, i.e., $T = \min[T_1, ...T_N]$. Assuming independence among all cells in the tissue, we have that the survival function of the model satisfies

$$S(t) = P(T > t) = P(\min[T_1, ..., T_N]) = (1 - p_n(t))^N \tag{6}$$

and, substituting this relationship into definition (1) for the hazard function, we get that the hazard is given by

$$h(t) = \frac{Np_n'(t)}{1 - p_n(t)}. \tag{7}$$

If the probability that a cell becomes malignant is small (i.e. if $p_n(t) \approx 0$ for all t), then we can approximate the hazard by

$$h(t) \approx N p_n'(t). \tag{8}$$

A further approximation is obtained if we calculate the Taylor series of (8) and keep only the first non-zero term of the series. This leads to the well-known approximation to the hazard function of the Armitage–Doll model

$$h(t) \approx \frac{N\lambda_0\lambda_1...\lambda_{n-1}t^{n-1}}{(n-1)!}. \tag{9}$$

This approximation has been extensively used in the past by epidemiologists. However, its long-term (large-age) behavior is completely different from the long-term behavior of the exact solution, which converges to $N\lambda_{min}$. The approximation is particularly inappropriate if we need to analyze laboratory experiments, where the probability of malignancy is usually significant [57]. On the other side, the "approximate" hazard behaves exactly as a power of age (the number of intermediate cells in the model), which is consistent with the observed mortality of several cancers [1]. So the Armitage–Doll model gives a very simple explanation to the observation that several cancers behave approximately as a power of age. For this reason and for its simplicity, the Armitage–Doll model remains one of the basic tools in cancer epidemiology.

4.1 Exact Solution of the Armitage–Doll Model with Constant Rates

As mentioned above, the exact Armitage–Doll model hazard and survival function can be computed by solving the system of differential equations (5). The solution depends on the eigenvalues of the system, which coincide exactly with the transition rates of the model, $\lambda_0, ..., \lambda_{n-1}$. The functional form of the solution depends on the equality or inequality of the transition rates. We present the exact hazard and survival function in case of equality or inequality of all the transition rates.

If the transition rates between all stages are equal, then the malignant transformation time of a cell is a Gamma random variable with parameters (n, λ). In this case, the hazard and survival function of the model (formulas 6 and 7) can be evaluated explicitly:

$$S(t) = \frac{e^{-\lambda t}}{(n-1)!}\sum_{k=0}^{n-1}\frac{(n-1)!}{k!}(\lambda t)^k, \tag{10}$$

$$h(t) = \frac{Nt^{n-1}\lambda^n}{\sum_{k=0}^{n-1}\frac{(n-1)!}{k!}(\lambda t)^k}. \tag{11}$$

Alternatively, if we assume that the transition rates are all different, then the exact hazard and survival function of the model are given by

$$S(t) = \left[\frac{\lambda_0\cdots\lambda_{n-1}}{L^*}\sum_{k=0}^{n-1}\frac{e^{-\lambda_k t}}{\lambda_k}L_k\right]^N, \tag{12}$$

$$h(t) = \frac{N\sum_{k=0}^{n-1}e^{-\lambda_k t}L_k}{\sum_{k=0}^{n-1}\frac{e^{-\lambda_k t}}{\lambda_k}L_k}, \tag{13}$$

where

$$L^* = \prod_{i<j}(\lambda_i - \lambda_j),$$

$$L_k = \prod_{\substack{i<j\\ i,j\neq k}}(\lambda_i - \lambda_j).$$

From (13) it is straightforward to show that the Armitage–Doll hazard converges to $N\lambda_{min}$ as $t \to \infty$, which is a very important qualitative difference between the exact solution and its approximation, as was mentioned before.

4.2 Armitage–Doll Hazard after Acute Exposures to Mutagens

It is difficult to obtain closed form expressions of the Armitage–Doll hazard when the model parameters are age-dependent. However, Heidenreich *et al.* derived expressions for the Armitage–Doll hazard when an acute exposure to a mutagen occurs [30].

Assume that the exposure occurs at age τ. After the exposure, a fraction f_k of the cells in stage k is assumed to jump to stage $k+1$, $k = 0, ..., n-1$. This can be modeled mathematically by assuming that the system (5) is restarted at age τ with the initial conditions

$$\begin{aligned} p_0(\tau) &= (1-f_0)p_0(\tau), \\ &\vdots \\ p_k(\tau) &= (1-f_k)p_k(\tau) + f_{k-1}p_k(\tau), \\ &\vdots \\ p_n(\tau) &= f_{n-1}p_{n-1}(\tau), \end{aligned} \tag{14}$$

where the values of $p_0(\tau), ..., p_{n-1}(\tau)$ on the right hand side are the values right before the exposure.

If the fractions and the background transition rates between all stages are assumed to be equal and constant, then the hazard after exposure is given by [30]

$$h(t) = \frac{N[1 + f(\frac{n-1}{\lambda t} - 1)](t^{n-1}\lambda^n)}{\sum_{k=0}^{n-1} \frac{1}{k!}(\lambda t)^k - f\frac{(\lambda t)^{n-1}}{(n-1)!}}, \tag{15}$$

for $t > \tau$.

5 The Two-Stage Clonal Expansion Model

The two-stage clonal expansion model (TSCE) was originally developed by Moolgavkar and Venzon [62] and Moolgavkar and Knudson [56]. This model assumes that malignant transformation of susceptible stem cells occurs as a result of two specific, irreversible and hereditary events. In its most general form, the number of normal (susceptible) stem cells as a function of age, $X(t)$, is modeled as an stochastic process. However, for simplicity we concentrate only in the case when the normal stem cells are assumed to follow a deterministic process [62, 56, 55]. The first event in the TSCE occurs when a normal stem cell becomes "initiated". Initiation is modeled as a non-homogeneous Poisson process of intensity $\nu(t)X(t)$, where $X(t)$ is the number of susceptible stem cells at age t. Initiated cells, also called intermediate cells, are assumed to expand clonally; this expansion is modeled as a birth-death and mutation process. This means that initiated cells can divide into two initiated cells with rate $\alpha(t)$, die or differentiate with rate $\beta(t)$ and divide into one intermediate and one malignant cell with rate $\mu(t)$. Figure 2 gives a schematic representation of this model.

Following the derivation of the TSCE model in [55], we calculate the hazard and survival functions of the model. Let $Y(t)$ and $Z(t)$ be the number of intermediate and malignant cells at age t, respectively and let $P_{j,k}(t) \equiv \text{Prob}[Y(t) = j, Z(t) = k]$. The vector $\{Y(t), Z(t)\}$ constitutes a two-dimensional continuous-time Markov chain. Hence, the $P_{j,k}(t)$'s satisfy the Chapman-Kolmogorov equations

$$\begin{aligned} P_{j,k}(t+h) &= [\nu(t)X(t)P_{j-1,k}(t) + (j-1)\alpha(t)P_{j-1,k}(t)]h \\ &\quad +[(j+1)\beta(t)P_{j+1,k}(t) + j\mu(t)P_{j,k-1}(t)]h \\ &\quad +[1 - jh(\alpha(t) - \beta(t) - \mu(t)) - hX(t)\nu(t)]P_{j,k}(t) + o(h). \end{aligned} \tag{16}$$

This equation is a consequence of the Markovian property. It expresses the probability of the process at time $t+h$, using the probability at time t. Dividing by h and taking the limit when h goes to zero

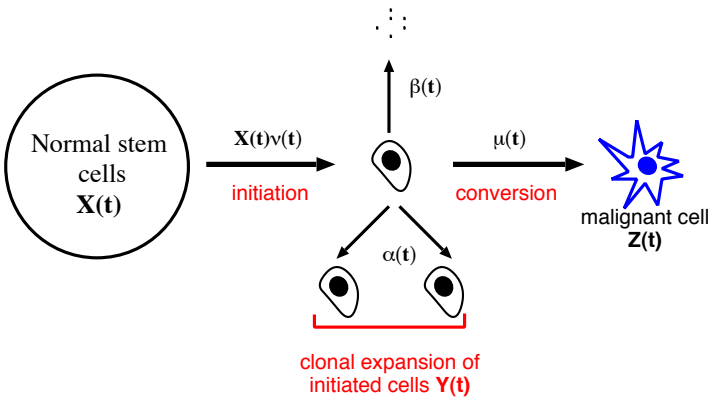

Figure 2: Two-stage clonal expansion model.

in (16), we get the Kolmogorov differential equations [55]

$$\begin{aligned}\frac{d}{dt}P_{j,k}(t) &= [(j-1)\alpha(t) + X(t)\nu(t)]P_{j-1,k}(t) \\ &\quad + (j+1)\beta(t)P_{j+1,k}(t) + j\mu(t)P_{j,k-1}(t) \\ &\quad - \{j[\alpha(t)+\beta(t)+\nu(t)] + X(t)\nu(t)\}P_{j,k}(t).\end{aligned} \tag{17}$$

Let $\Psi(y,z,t)$ denote the probability generating function (PGF) of the process, i.e. let $\Psi(y,z,t) \equiv E[y^{Y(t)}z^{Z(t)}] = \sum_{j,k} P_{j,k}(t)y^j z^k$. The probability generating function of a discrete stochastic process summarizes all its probabilistic properties. Hence it can be used, at least in theory, to calculate all the quantities of interest. Multiplying (17) by $y^j z^k$ and summing over all j and k, we get the forward Kolmogorov differential equation for the PGF [55],

$$\begin{aligned}\frac{\partial\Psi(y,z,t)}{\partial t} &= (y-1)X(t)\nu(t)\Psi(y,z,t) \\ &\quad + \{[\mu(t)z + \alpha(t)y - \alpha(t) - \beta(t) - \mu(t)]y + \beta(t)\}\frac{\partial\Psi(y,z,t)}{\partial y}.\end{aligned} \tag{18}$$

If we assume that all stem cells are normal at the time of birth, then $\Psi(y,z,0) = 1$.

Given the probability generating function, the survival function of the time to the appearance of the first malignant cell, $S(t)$, is given by

$$\Psi(1,0,t) = \sum_j P_{j,0}(t) = P[T > t] = S(t),$$

and the corresponding hazard is then [55]

$$h(t) = -\frac{\Psi'(1,0,t)}{\Psi(1,0,t)}. \tag{19}$$

This implies that we only need to solve for $\Psi(1,0,t)$ and its derivative with respect to t to obtain the hazard and survival function. However, more can be said before solving directly any equation. Using (18) we get that $\Psi'(1,0,t) = -\mu(t)\frac{\partial\Psi(1,0,t)}{\partial y}$. This plus the fact that

$$E[Y(t)|Z(t)=0] = \sum_j \frac{P_{j,0}}{\sum_i P_{i,0}} j = \frac{\frac{\partial\Psi(1,0,t)}{\partial y}}{\Psi(1,0,t)} \tag{20}$$

gives us the hazard

$$h(t) = -\frac{\Psi'(1,0,t)}{\Psi(1,0,t)} = \mu(t)E[Y(t)|Z(t)=0]. \tag{21}$$

As pointed out in Section 4, cancer is a rare event, so it can be assumed that the probability of malignant transformation is very small. Therefore, the conditional expectation in (21) can be approximated by the unconditional expectation $E[Y(t)]$. Moreover, we see that $\frac{\partial \Psi}{\partial y}(1,1,t) = E[Y(t)]$, and using (18), we obtain [55]

$$\frac{d}{dt}E[Y(t)] = \nu(t)X(t) + [\alpha(t) - \beta(t)]E[Y(t)], \tag{22}$$

which can easily be solved for $E[Y(t)]$. Thus, a general approximation to the TSCE hazard is given by [55]

$$h(t) \approx \mu(t)E[Y(t)] = \mu(t)\int_0^t \nu(s)X(s)\exp\left[\int_s^t [\alpha(u) - \beta(u)]du\right]ds. \tag{23}$$

A detailed discussion of the validity of this approximation is given in [55, 57]. However it is important to point out that the use of (23) instead of the exact solution may give incorrect results when analyzing experimental and epidemiological data. In experimental studies, the probability of malignancy is usually significant. Thus, the approximation of the conditional expectation by the unconditional one may not be accurate in this case. Moreover, the effects of carcinogen exposures on cancer risk predicted by the TSCE model differ depending upon whether you use the exact hazard or its approximation. For this reason, it has been pointed out by several authors that in most cases it is better not to make use of the approximation [55, 61, 57].

5.1 Constant Number of Susceptible Stem Cells and Constant Model Parameters

If we assume that the number of susceptible stem cells is constant during life and that the model parameters are constant, then we can find a close form expression for the hazard and survival function under the TSCE model.

In this case, equation (18) can be easily solved using the method of characteristics and the expressions for the hazard and survival function are [49]

$$S(t) = \left(\frac{q-p}{qe^{-pt} - pe^{-qt}}\right)^{\frac{\nu X}{\alpha}} \tag{24}$$

and

$$h(t) = \frac{\nu X}{\alpha}\frac{pq(e^{-qt} - e^{-pt})}{qe^{-pt} - pe^{-qt}}, \tag{25}$$

where

$$p = \frac{1}{2}\left[(-\alpha + \beta + \mu) - \sqrt{(\alpha - \beta - \mu)^2 + 4\alpha\mu}\right], \tag{26}$$

$$q = \frac{1}{2}\left[(-\alpha + \beta + \mu) + \sqrt{(\alpha - \beta - \mu)^2 + 4\alpha\mu}\right]. \tag{27}$$

5.2 Age-Dependent Piecewise Constant Parameters

Multistage models can be used to evaluate the effects of carcinogens on cancer risk, by assuming that the model parameters are functions of the exposure dose at any particular age. Similarly, multistage models can also be used to evaluate the effects of preventive measures on cancer risk, for example, the use of non-steroidal anti-inflammatory drugs as a preventive measure to reduce the risk of polyps in the colon.

To evaluate the effects of carcinogens with age-dependent exposure, we need expressions of the hazard and survival functions when the model parameters depend on age. Closed form expressions

for the hazard and survival functions of the TSCE model are not available in the case of general age-dependent parameters, so numerical methods are necessary. However, closed form expressions for the hazard and survival function of the TSCE model in the case of piecewise constant parameters were derived by Heidenreich *et al.* [31].

Let's assume that the parameters of the TSCE are constant between the time points $0, t_1, t_2, ..., t_K$. The number of time intervals is K, and in the i–th interval, the parameters of the model are α_i, β_i, ν_i, for $i = 1...K$.

The characteristic equations of (18) are [31]

$$\begin{aligned}
\frac{dy}{du} &= -\{[\mu(t)z + \alpha(t)y - \alpha(t) - \beta(t) - \mu(t)]y + \beta(t)\}, \\
\frac{dz}{du} &= 0, \qquad \frac{dt}{du} = 1, \\
\frac{d\Psi}{du} &= (y-1)X(u)\nu(u)\Psi.
\end{aligned} \tag{28}$$

As shown in the previous sections, we need only to solve to calculate $\Psi(1, 0, t)$ to obtain the hazard function. Thus we need only to find Ψ along the characteristic through $(y(0), 0, 0)$ with $y(t) = 1$. Once $y(t, u)$ is known, we can easily solve [31] for

$$\Psi(y(t), z, t) = \Psi_0 \exp\left[\int_0^t [y(u,t) - 1]X(u)\nu(u)du\right]. \tag{29}$$

The equation for $y(u, t)$ is a Ricatti differential equation with piecewise constant parameters, so it can be solved analytically. Heidenreich *et al.* [31] showed that under this assumption, the solution to the characteristic equations is given by

$$\alpha_i(y(u,t) - 1) = \frac{\partial \log f_i(u,t)}{\partial u}, \tag{30}$$

where

$$\begin{aligned}
f_i(u,t) &= (\tilde{y_i} - p_i)\exp\left[q_i(u - t_i)\right] + (q_i - \tilde{y_i})\exp\left[p_i(u - t_i)\right], \\
p_i, q_i &= \frac{1}{2}\left((-\alpha_i + \beta_i + \mu_i) \mp \sqrt{(\alpha_i - \beta_i - \mu_i)^2 + 4\alpha_i\mu_i}\right), \\
f_i(t_i, t) &= q_i - p_i, \\
\tilde{y}_{i-1} &= \frac{\alpha_{i-1}}{\alpha_i}\frac{\partial \log f_i(u,t)}{\partial u}|_{u=t_{i-1}}.
\end{aligned} \tag{31}$$

The final expressions for the hazard and survival function in the case of age-dependent piecewise-constant parameters are [31]

$$S(t) = \exp\left\{\sum_{i=1}^{K} \frac{v_i}{\alpha_i} \ln \frac{q_i - p_i}{f_i(t_{i-1}, t)}\right\}, \tag{32}$$

$$h(t) = \sum_{i=1}^{K} \frac{v_i}{\alpha_i}\frac{\partial}{\partial t}\log\left(f_i(t_{i-1}, t)\right), \tag{33}$$

where all the summands can be calculated recursively using (31).

5.2.1 Time between Malignant Conversion and Clinical Detection or Death

In its simplest form, the TSCE model does not account for the time between the appearance of the first malignant cell and cancer incidence or mortality (usually termed "lag time"). Several approaches can be taken to address this issue. The simplest one is to assume that the lag time is constant. Although

somewhat unrealistic, this approach adds only one more parameter to the model and the computation of the hazard or survival function at age t from the corresponding TSCE model functions is straightforward. A more realistic approach is to assume a distribution for the lag time, so inter-individual variability can be taken into consideration. A gamma distribution has often been used. This adds two more parameters to the model, the mean and the standard deviation of the gamma distribution, and calculation of the hazard and survival functions involves the convolution of the gamma distribution with the TSCE survival and hazard functions. Specifically,

$$S(t) = \begin{cases} S_2(t - t_{lag}) & \text{if lag time is constant,} \\ 1 - \int_0^t (1 - S_2(u)) f(t-u) du & \text{if lag time is gamma-distributed,} \end{cases} \quad (34)$$

$$h(t) = \begin{cases} h_2(t - t_{lag}) & \text{if lag time is constant,} \\ (\int_0^t h_2(u) S_2(u) f(t-u) du)/S(t) & \text{if lag time is gamma-distributed,} \end{cases} \quad (35)$$

where $h_2(t)$ and $S_2(t)$ represent the TSCE model hazard and survival, respectively, and $f(\cdot)$ is the gamma density. In general, the integrals cannot be evaluated in closed-form and numerical methods for integration are required.

Another approach is to incorporate the dynamics of malignant cells explicitly into the model and assume that the malignant compartment follows also a birth-death-mutation process. The "mutation rate" of the malignant compartment is then interpreted as the rate of cancer detection or death (per malignant cell). Although this approach can be thought as more realistic and complete, it has some disadvantages. Having models with more than one clonal expansion stage makes impossible the calculation of closed-form expressions for the hazard and survival function, so numerical methods are required [41, 44, 9, 51]. In addition, the parameters of the malignant compartment may not be identifiable from incidence or mortality data alone, so explicit modeling of the malignant cell dynamics may just overparameterize the model unnecessarily. Nevertheless, models with more than one clonal expansion have been implemented successfully and used to analyze the incidence and mortality of several cancers [41, 44]. However, no significant improvement in fit has been obtained when compared with models with a single clonal expansion stage, suggesting that more data on intermediate lesions is required [40, 49, 53].

6 Analysis of Premalignant Lesions using the TSCE

In several cancers, premalignant lesions appear long before the clinical onset of the disease. Examples of these are the aberrant crypts foci (ACF) and the adenomatous polyps that appear in the colon and which are considered to be precursors of colorectal carcinoma. Premalignant lesions provide important information about carcinogenesis. They clearly show the multistage nature of the process and their study has provided a lot of insight about the mechanisms of malignant transformation.

Since the experiments of Berenblum and Shubik [4], the analysis of premalignant lesions in animal experiments constitutes one of the principal frameworks to investigate the carcinogenesis process. In particular, initiation-promotion animal experiments where tumors are induced by chemical agents have been particularly useful for elucidating the specific effects of carcinogens on the occurrence and growth dynamics of premalignant lesions [57]. More recently, the analysis of premalignant lesions in humans has provided significant information about the genetic transformations required for malignant transformation in different tissues [19, 3]. Moreover, epidemiological analysis of the characteristics of premalignant lesions found during screening has also been crucial in the design of prevention, and intervention strategies against cancer [18, 50, 5, 34]. One of the big advantages of using multistage models is that they provide a unified framework for analyzing epidemiological and experimental data.

Cancer incidence (or mortality) data alone may not be sufficient to evaluate in great detail the mechanisms of carcinogenic agents. For example, the cell division and cell death rates cannot be estimated independently when using cancer incidence data alone. However, information about the number and size distribution of premalignant clones can elucidate the specific role of some agents for cancer development [61]. For example, in rat hepatocarcinogenesis experiments, good quantitative information is

available on the number and size of enzyme-altered foci, which are considered premalignant clones in liver cancer. And this information may allow the estimation of both the cell division and cell death rates in the TSCE model.

Several expressions for the number and size of premalignant lesions under different scenarios have been derived in the past [17, 61, 15, 34]. We briefly describe some of the mathematical formulas for the number and size distributions of premalignant lesions and the total number of premalignant cells under the TSCE model.

6.1 The Number and Size Distribution of Premalignant Lesions Using the Two-Stage Clonal Expansion Model

Dewanji *et al.* [17] and Luebeck and Moolgavkar [48] derived expressions for the number and size distributions of premalignant lesions under the TSCE. Let $Y(t,s)$ be defined as the size of a premalignant lesion at time t, given it was originated at time s. $Y(t,s)$ follows a birth and death process with birth rate $\alpha(t-s)$ and death rate $\beta(t-s)$. Let $P_i(t,s) = P[Y(t,s) = i]$. Then, the probability generating function of $Y(t,s)$ is given by [70, 48]

$$\begin{aligned} Q(z;t,s) &\equiv \sum_{i=0}^{\infty} P_i(t,s) z^i \\ &= 1 - \frac{z-1}{(z-1)G(t,s) - g(t,s)}, \end{aligned} \tag{36}$$

where

$$g(t,s) = \exp\left[-\int_s^t (\alpha(u) - \beta(u)) du\right]$$

and

$$G(t,s) = \int_s^t \alpha(u) g(u,s) du.$$

Substituting $z = 0$ in (36), we get that the probability of extinction of a premalignant lesion by time t, given it was originated at time s is

$$P_0(t,s) = 1 - \frac{1}{G(t,s) + g(t,s)}.$$

The expected size of a premalignant lesion at time t, given it was originated at time s is [48]

$$E[Y(t,s)] = \frac{\partial Q(z;t,s)}{\partial z}|_{z=1} = \exp\left[\int_s^t (\alpha(u) - \beta(u)) du\right]. \tag{37}$$

Initiation follows a non-homogeneous Poisson process with rate $X(s)\nu(s)$, so the number of non-extinct premalignant lesions ($\equiv N(t)$) will follow also a Poisson process with rate [48]

$$\Lambda(t) = \int_0^t X(s)\nu(s)(1 - P_0(t,s)) ds = \int_0^t \frac{X(s)\nu(s)}{G(t,s) + g(t,s)} ds. \tag{38}$$

The last equation represents also the mean number of premalignant lesions existing at time t.

Now, we analyze the size of a particular premalignant lesion, given that it hasn't become extinct by time t. The probability generating function of a premalignant lesion originated at time s, given that it still exists at time t, $E[z^{Y(t,s)}|Y(t,s) > 0]$, satisfies

$$E[z^{[Y(t,s)|Y(t,s)>0]}] = \frac{Q(z;t,s) - Q(0;t,s)}{1 - Q(0;t,s)} = 1 - \frac{[G(t,s) + g(t,s)](z-1)}{(z-1)G(t,s) - g(t,s)}. \tag{39}$$

So by taking the derivatives of (39) and evaluating them at $z = 0$, we obtain a general expression for the probability that the number of cells in a non-extinct premalignant lesion at time t is equal to $m > 0$ [48]:

$$P[Y(t,s) = m | Y(t,s) > 0] = \frac{g(t,s)}{G(t,s)} \left(\frac{G(t,s)}{G(t,s) + g(t,s)} \right)^m. \tag{40}$$

The number of non-extinct premalignant lesions by time t follows a non-homogeneous Poisson process with rate $\Lambda(t)$. This implies that the occurrence time of a premalignant lesion, given it appeared before age t, is a random variable on $0 < s < t$ with density function given by [69, 48]

$$\frac{\lambda(t,s)}{\Lambda(t)} \equiv \frac{X(s)\nu(s)[1 - P_0(t,s)]}{\Lambda(t)}. \tag{41}$$

So, the unconditional probability of having a non-extinct premalignant lesion of size m at time t, $p_m(t)$, is given by [48]

$$\begin{aligned} p_m(t) &= \frac{1}{\Lambda(t)} \int_0^t \lambda(s,t) P[Y(t,s) = m | Y(t,s) > 0] ds \\ &= \frac{1}{\Lambda(t)} \int_0^t \frac{X(s)\nu(s)}{G(t,s) + g(t,s)} \left[\frac{g(t,s)}{G(t,s)} \left(\frac{G(t,s)}{G(t,s) + g(t,s)} \right)^m \right] ds. \end{aligned} \tag{42}$$

Although these expressions can be used with any age-dependent parameters, it may not be possible to evaluate (38) and (42) in closed form. However, closed form expressions for the number and size distributions of premalignant lesions in case of piecewise constant parameters are available in the literature [38, 17].

Constant parameters.

The number of non-extinct premalignant lesions at time t, $N(t)$, is a Poisson random variable with expectation $\Lambda(t)$, where [17]

$$\Lambda(t) = \frac{\nu X}{\alpha} \left[\ln \left(\frac{\beta}{\beta - \alpha p(t)} \right) \right] \tag{43}$$

and

$$p(t) = \begin{cases} \frac{\beta - \beta e^{-(\alpha-\beta)t}}{\alpha - \beta e^{-(\alpha-\beta)t}}, & \text{if } \alpha \neq \beta, \\ \frac{\alpha t}{1 + \alpha t}, & \text{if } \alpha = \beta. \end{cases} \tag{44}$$

For a non-extinct premalignant lesion, say $W(t)$, at time t, the probability of it having $m \geq 1$ cells is [17]

$$\Pr[W(t) = m | W(t) > 0] = \frac{\left(\frac{\alpha}{\beta} p(t) \right)^m}{m \ln \left(\frac{\beta}{\beta - \alpha p(t)} \right)}, \tag{45}$$

$$E[W(t) | W(t) > 0] = \frac{\alpha}{\alpha - \beta} \left[\frac{e^{(\alpha-\beta)t} - 1}{\ln(\beta / (\beta - \alpha p(t)))} \right]. \tag{46}$$

Minimum detection threshold n_0.

Since in reality, premalignant lesions can be detected only when they are larger than a certain size, Dewanji et al. also derived expressions for the number and size of detectable premalignant lesions assuming a detection threshold size n_0 [17]. Assuming that α and β are constant, the number of detectable premalignant lesions at time t, say $N^*(t)$, is a Poisson random variable with expectation $\Lambda^*(t)$:

$$\Lambda^*(t) = \int_0^t \nu(u) X(u) (1 - p^*(t-u)) du, \tag{47}$$

where $p^*(t-u)$ is the probability that the size of the lesion is less than or equal to n_0 at time t, and is given by

$$p^*(t-u) = 1 - (1-p(t-u))\left(\frac{\alpha}{\beta}p(t-u)\right)^{n_0}, \tag{48}$$

where $p(t-u)$ is the same as in (44).

The distribution of the size of a detectable premalignant lesion at time t is given by

$$\Pr[W(t) = m | W(t) > n_0] = \frac{1}{\Lambda^*(t)}\int_0^t \lambda^*(u)\left(\frac{\alpha}{\beta}p(t-u)\right)^{m-n_0-1}\left(1-\frac{\alpha}{\beta}p(t-u)\right)du, \tag{49}$$

where $\lambda^*(u) = \nu(u)X(u)(1-p^*(t-u))$. The expected size of a detectable premalignant lesion is given by

$$E[W(t)|W(t) > n_0] = n_0 + \frac{1}{\Lambda^*(t)}\int_0^t \nu(u)X(u)\left(\frac{\alpha}{\beta}p(t-u)\right)^{n_0} e^{(\alpha-\beta)(t-u)}du. \tag{50}$$

Piecewise constant parameters.

If we assume that the model parameters are piecewise constant, the number of non-extinct premalignant lesions, $N(t)$, is Poisson distributed with mean [17, 38, 46]

$$\Lambda(t) = \sum_{j=1}^{n}\frac{\nu_{0,j}X}{\alpha_j}\ln\left[\frac{\alpha_j e^{\{(\alpha_j-\beta_j)(t_j-t_{j-1})\}} - \tilde{\beta}_j}{\alpha_j - \tilde{\beta}_j}\right], \tag{51}$$

where n is the number of age-periods with different parameter values before age t; $[t_{j-1}, t_j], j = 1, ..., n$ denote the end-points of the $j-th$ age-period; $t_0 = 0$, and $\nu_{0,j}$, α_j, β_j denote the parameter values during the $j-th$ age-period.

The probability of having $m \geq 1$ cells in a nonextinct premalignant lesion at time t satisfies [17, 38, 46]

$$\Pr[W(t) = m | W(t) > 0] = \frac{1}{\Lambda(t)}\sum_{j=1}^{n} p_j(m), \tag{52}$$

where

$$p_j(m) = \frac{1}{m}\left\{\left(\frac{\nu_{0,j}X}{\alpha_j} - \frac{\nu_{0,j-1}X}{\alpha_{j-1}}\right)\frac{\alpha_j - \tilde{\alpha}_j e^{\{-(\alpha_j-\beta_j)(t_j-t_{j-1})\}}}{\alpha_j - \tilde{\beta}_j e^{\{-(\alpha_j-\beta_j)(t_j-t_{j-1})\}}}\right\}^m, j = 1, ..., n, \tag{53}$$

and $\nu_{0,0}X/\alpha_0 \equiv 0$. The parameters $\tilde{\alpha}_j$ and $\tilde{\beta}_j$ are defined recursively by

$$\tilde{\alpha}_n = \alpha_n, \tag{54}$$

$$\tilde{\alpha}_j = \alpha_j - \frac{(\alpha_j - \beta_j)}{(\alpha_{j+1} - \beta_{j+1})}(\alpha_{j+1} - \tilde{\alpha}_{j+1}e^{\{-(\alpha_{j+1}-\beta_{j+1})(t_{j+1}-t_j)\}}), \tag{55}$$

$$\tilde{\beta}_n = \beta_n, \tag{56}$$

$$\tilde{\beta}_j = \tilde{\alpha}_j - (\alpha_j - \beta_j)\prod_{l=j+1}^{n} e^{\{-(\alpha_l-\beta_l)(t_l-t_{l-1})\}}, \tag{57}$$

$$j = n-1, ..., 1. \tag{58}$$

6.1.1 Distribution of Total Number of Premalignant Cells

Let $Y(t)$ denote the total number of premalignant cells (in all lesions present) at time t. Let $P_m(t)$ denote the probability that $Y(t) = m$. Then [15]:

$$P_0(t) = \exp(-\Lambda(t)), \tag{59}$$

$$P_m(t) = \sum_{i=0}^{m-1} \frac{m-i}{m} P_i(t) p_{m-i}(t), \tag{60}$$

where

$$p_k(t) = \int_0^t \nu(u)X(u)\frac{1}{G(u,t)}\frac{g(u,t)}{G(u,t)+g(u,t)}\left(\frac{G(u,t)}{G(u,t)+g(u,t)}\right)^k du, \tag{61}$$

and $\Lambda(t)$ denotes the expected number of nonextinct premalignant lesions at time t.

The mean and variance of $Y(t)$ are

$$E[Y(t)] = \int_0^t \nu(u)X(u)\exp\left[\int_u^t (\alpha(u,s)-\beta(u,s))ds\right]du, \tag{62}$$

and

$$Var[Y(t)] = \int_0^t \nu(u)X(u)\frac{2G(u,t)+g(u,t)}{g^2(u,t)}du. \tag{63}$$

Constant Parameters.

When the parameters X, ν, α, β, and μ are constant, we can evaluate explicitly the distribution of $Y(t)$:

$$\Pr[Y(t)=n] = \frac{\Gamma(\nu X/\alpha+n)}{\Gamma(n+1)\Gamma(\nu X/\alpha)}(1-\alpha\zeta)^{\nu X/\alpha}(\alpha\zeta)^n, \tag{64}$$

where

$$\zeta = \frac{e^{(\alpha-\beta)t}-1}{\alpha e^{(\alpha-\beta)t}-\beta}. \tag{65}$$

Distribution of total number of premalignant cells conditional on no prior malignancies.

To model screening for premalignant lesions in asymptomatic individuals with no prior history of cancer, Jeon et al. derived mathematical expressions for the size distribution of premalignant lesions, conditional on no prior malignancy in the tissue of interest [34].

Let $Z(t)$ be the indicator variable for clinical detection of cancer. For $n \geq 0$, assuming constant parameters $X, \nu, \alpha, \beta, \mu$, the size distribution of the total number of premalignant cells conditioned on no malignancy satisfies [34]

$$P^*[Y(t)=n] \equiv \Pr[Y(t)=n|Z(t)=0, Y(0)=0] \tag{66}$$

$$= \frac{\Gamma(\nu X/\alpha+n)}{\Gamma(n+1)\Gamma(\nu X/\alpha)}(1-\alpha\zeta^*)^{\nu X/\alpha}(\alpha\zeta^*)^n, \tag{67}$$

where

$$\zeta^* = \frac{e^{-pt}-e^{-qt}}{(q+\alpha)e^{-pt}-(p+\alpha)e^{-qt}} \tag{68}$$

and

$$p = \frac{1}{2}\left[-(\alpha-\beta-\mu)-\sqrt{(\alpha-\beta-\mu)^2+4\alpha\mu}\right], \tag{69}$$

$$q = \frac{1}{2}\left[-(\alpha-\beta-\mu)+\sqrt{(\alpha-\beta-\mu)^2+4\alpha\mu}\right]. \tag{70}$$

6.1.2 Number and Size when Premalignant Growth is Modeled as Stochastic Gompertz Growth

Tumors do not grow without bound and in particular several types of tumors have been shown to be consistent with Gompertzian growth [70]. Hence, it is natural to calculate the number and size distribution of non-extinct premalignant clones when they follow a stochastic Gompertz growth, as defined by Tan [70].

Suppose that the net cell proliferation as a function of the age of the clone is given by

$$\alpha(t-s) - \beta(t-s) = b \exp[-a(t-s)], \tag{71}$$

where $b > 0$ and a is any real number. Then from (37) we obtain that the mean growth of a clone as a function of its age, $t - s$ is

$$E(Y(t-s)) = \exp\left[\frac{b}{a}(1 - e^{-a(t-s)})\right], \tag{72}$$

which is a deterministic Gompertz growth. Suppose that $\gamma \equiv \frac{\beta(t-s)}{\alpha(t-s)}$ is a constant. Then under these assumptions, the number and size distributions of non-extinct clones reduce to [48]:

$$\Lambda(t) = \int_0^t X(s)\nu(s)\frac{1-\gamma}{1-\gamma g(t-s)}ds \tag{73}$$

and

$$p_m(t) = \frac{(1-\gamma)^2}{\Lambda(t)} \int_0^t X(s)\nu(s) \left[\frac{g(t-s)}{(1-\gamma g(t-s))^2}\right] \left[\frac{1-g(t-s)}{1-\gamma g(t-s)}\right]^{m-1} ds.$$

It is worth mentioning that we can also obtain the number and size distributions under general assumptions of the parameters using a generalization of the Luria–Delbrück fluctuation models proposed by Dewanji *et al.* [15].

6.1.3 Other Developments

In addition to the formulas presented here, expressions for the joint analysis of premalignant and malignant lesions under the TSCE model have been derived previously [16, 11]. Furthermore, Dewanji et al. derived expressions for the analysis of longitudinal data on the number and size of premalignant clones [14]. In particular, the joint distribution of the number of premalignant clones observed at different points in time in the same subject was derived. In addition, De Gunst et al. developed methods to estimate the number and size distribution of premalignant lesions assuming that transectional data is available (stereological problem) [11, 12]. We refer the reader to the literature for further details.

7 Some Generalizations of the TSCE

There is evidence that some cancers require more than two-steps for malignant transformation. For example, it is now accepted that more than two successive genetic transformations are required for the onset of colorectal adenocarcinoma [49]. In particular, it has been shown that the inactivation of both alleles of the tumor-suppressor gene APC gives rise to premalignant polyps and that the inactivation of another tumor-suppressor gene, probably TP53, is required for the transition from premalignant to malignant lesions [49]. For this reason, several generalizations of the TSCE model have been proposed to accommodate the recent advances in cancer biology [71, 41, 49, 42, 44].

7.1 Multiple Clonal Expansions

Multistage carcinogenesis models with several clonal-expansion stages have been analyzed by Tan [71], Little [41] and Little *et al.* [42]. Figure 3 shows a schematic representation of a multistage model that assumes $k-1$ clonal expansion stages before malignancy. The first transition event is modeled as a non-homogeneous Poisson process and the dynamics of all subsequent stages are modeled as birth-death and mutation processes. This implies that cells in stage i $(i = 1, ..., k-1)$ can divide into two identical cells with birth rate $\alpha_i(t)$, die or differentiate with rate $\beta_i(t)$ and divide into one type i and one type $i+1$ cell with rate $\mu_{k-1}(t)$. The last stage I_k corresponds to malignancy. In this discussion we concentrate only on the occurrence of the first malignant cell. However, if we are interested in modeling the dynamics of malignant cells as well, we can associate the $k-1$-th stage with malignancy and consider the last transition event as clinical diagnosis. In this case, the probability of diagnosis of a malignant clone would be proportional to its size.

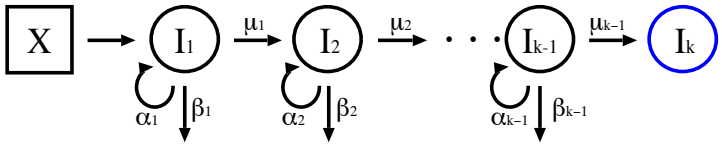

Figure 3: Multistage model with several clonal expansions.

7.1.1 Forward Kolmogorov Differential Equation

Let

$$P_{j_1,\ldots,j_k} \equiv P[I_1(t) = j_1, ..., I_k(t) = j_k | I_1(0) = 0, ..., I_k(0) = 0] \tag{74}$$

and let

$$\Psi(y_1, ..., y_k, t) = \sum_{j_1,\ldots,j_k} P_{j_1,\ldots,j_k}(t) y_1^{j_1} \cdots y_k^{j_k} \tag{75}$$

be the probability generating function of the process. This model is again a Markov process, hence $\Psi(y_1, ..., y_k, t)$ satisfies the forward Kolmogorov differential equation:

$$\begin{aligned} \frac{\partial \Psi(y_1, ..., y_k, t)}{\partial t} &= (y-1)X(t)\nu(t)\Psi(y_1, ..., y_k, t) \\ &+ \sum_{i=1}^{k-1} \{[\alpha_i(t)y_i - \alpha_i(t) - \beta_i(t) - \mu_i(t)(1 - y_{i+1})]y_i + \beta_i(t)\} \frac{\partial \Psi(y_1, ..., y_k, t)}{\partial y_i}. \end{aligned} \tag{76}$$

As mentioned before, the PGF contains all the probabilistic information of the process, so equation (76) can be used to obtain expressions for the hazard, the survival function and the distributions of the number of cells in each compartment. In particular, the survival function of the time of appearance of the first malignant cell is given by

$$S(t) = \Psi(1, ..., 1, 0, t), \tag{77}$$

which we can obtain by solving the characteristic equations of (76). It is important to mention that it is very difficult to solve (76) analytically when several clonal expansion stages are assumed, so numerical methods are usually needed. To obtain the hazard, we must use numerical approximations to $\Psi(1, ..., 1, 0, t)$ to approximate the logarithmic derivative of $S(t)$. This increases the number of computations required to calculate the hazard and makes it more difficult to control the error of the numerical approximations. A better alternative is to use the Kolmogorov backward differential equations of the process as shown below.

7.1.2 Backward Kolmogorov Differential Equations

The model shown in figure 3 can also be seen as a multi-type branching process [24]. This implies that for each fixed t, the following set of conditional probability generating functions characterize the whole process [24]:

$$\begin{aligned} \Psi(y_1, ..., y_k, \tau; t) &= E[y_1^{I_1(t)}, ..., y_k^{I_k(t)} | I_1(\tau) = 0, ..., I_k(\tau) = 0], \\ \Phi_1(y_1, ..., y_k, \tau; t) &= E[y_1^{I_1(t)}, ..., y_k^{I_k(t)} | I_1(\tau) = 1, I_2(\tau) = 0, ..., I_k(\tau) = 0], \\ &\vdots \\ \Phi_{k-1}(y_1, ..., y_k, \tau; t) &= E[y_1^{I_1(t)}, ..., y_k^{I_k(t)} | I_1(\tau) = 0, ..., I_{k-1}(\tau) = 1, I_k(\tau) = 0], \\ \Phi_k(y_1, ..., y_k, \tau, t) &= E[y_1^{I_1(t)}, ..., y_k^{I_k(t)} | I_1(\tau) = 0, ..., I_{k-1}(\tau) = 0, I_k(\tau) = 1] = y_k, \end{aligned} \tag{78}$$

where $0 \leq \tau \leq t$. Notice that the last equation expresses the fact that we are interested only in the first transition to the malignant stage I_k.

It is easy to show that $\Psi(y_1,...,y_k,\tau;t), \Phi_1(y_1,...,y_k,\tau;t)$,...,$\Phi_{k-1}(y_1,...,y_k,\tau;t)$ satisfy the backward Kolmogorov differential equations [24]:

$$\begin{aligned}
\frac{\partial\Psi(\tau;t)}{\partial\tau} &= -\nu(\tau)X(\tau)\Psi(\tau;t)(\Phi_1(\tau;t)-1),\\
\frac{\partial\Phi_1(\tau;t)}{\partial\tau} &= -\beta_1(\tau)+(\alpha_1(\tau)+\beta_1(\tau)+\mu_1(\tau))\Phi_1(\tau;t)\\
&\quad -\mu_1(\tau)\Phi_1(\tau;t)\Phi_2(\tau;t)-\alpha_1(\tau)\Phi_1^2(\tau;t),\\
&\vdots\\
\frac{\partial\Phi_{k-1}(\tau;t)}{\partial\tau} &= -\beta_{k-1}(\tau)+(\alpha_{k-1}(\tau)+\beta_{k-1}(\tau)+\mu_{k-1}(\tau))\Phi_{k-1}(\tau;t)\\
&\quad -\mu_1(\tau)\Phi_{k-1}(\tau;t)\Phi_k(\tau;t)-\alpha_{k-1}\Phi_{k-1}^2(\tau;t),
\end{aligned} \tag{79}$$

where the dependency on $(y_1,...,y_k)$ has been omitted for convenience. Let $y_1=1,...,y_{k-1}=1,y_k=0$ and $s\equiv t-\tau$. Following the derivations of Little *et al.* [42] and Crump [9], we obtain the differential equations

$$\begin{aligned}
\frac{\partial\Psi(s;t)}{\partial s} &= \nu(t-s)X(t-s)\Psi(s;t)(\Phi_1(s;t)-1),\\
\frac{\partial\Psi'(s;t)}{\partial s} &= \nu(t-s)X(t-s)[\Psi'(s;t)(\Phi_1(s;t)-1)+\Psi(s;t)\Phi_1'(s;t)],\\
\frac{\partial\Phi_1(s;t)}{\partial s} &= \beta_1(t-s)-(\alpha_1(t-s)+\beta_1(t-s)+\mu_1(t-s))\Phi_1(s;t)\\
&\quad +\mu_1(t-s)\Phi_1(s;t)\Phi_2(s;t)+\alpha_1(t-s)\Phi_1^2(s;t),\\
\frac{\partial\Phi_1'(s;t)}{\partial s} &= -(\alpha_1(t-s)+\beta_1(t-s)+\mu_1(t-s))\Phi_1'(s;t)\\
&\quad +\mu_1(t-s)[\Phi_1'(s;t)\Phi_2(s;t)+\Phi_1(s;t)\Phi_2'(s;t)]\\
&\quad +2\alpha_1(t-s)\Phi_1(s;t)\Phi_1'(s;t),\\
&\vdots\\
\frac{\partial\Phi_{k-1}(s;t)}{\partial s} &= \beta_{k-1}(t-s)-(\alpha_{k-1}(t-s)+\beta_{k-1}(t-s)\\
&\quad +\mu_{k-1}(t-s))\Phi_{k-1}(s;t)+\alpha_{k-1}(t-s)\Phi_{k-1}^2(s;t),\\
\frac{\partial\Phi_{k-1}'(s;t)}{\partial s} &= -(\alpha_{k-1}(t-s)+\beta_{k-1}(t-s)+\mu_{k-1}(t-s))\Phi_{k-1}'(s;t)\\
&\quad +2\alpha_{k-1}(t-s)\Phi_{k-1}(s;t)\Phi_{k-1}'(s;t),
\end{aligned} \tag{80}$$

for $\Psi(s;t), \Phi_1(s;t), ..., \Phi_{k-1}(s;t)$ and their derivatives with respect to t. Here, $'$ represents differentiation with respect to t. The initial conditions are

$$\begin{aligned}
\Psi(0;t)=1 \quad &, \quad \Psi'(0;t)=0,\\
\Phi_1(0;t)=1 \quad &, \quad \Phi_1'(0;t)=0,\\
&\vdots\\
\Phi_{k-2}(0;t) &= 1, \Phi_{k-2}'(0;t)=0,\\
\Phi_{k-1}(0;t) &= 1, \Phi_{k-1}'(0;t)=-\mu_{k-1}.
\end{aligned} \tag{81}$$

The hazard and survival functions of the model are then given by

$$S_k(t)\equiv[\Psi(t;t)] \tag{82}$$

and

$$h_k(t) \equiv -\frac{\Psi'(t;t)}{\Psi(t;t)}. \tag{83}$$

Alternatively let $\varphi(s,t) \equiv -\ln[\Psi(s,t)]$. Then the second equation of (81) can be replaced with

$$\frac{\partial \varphi'(s,t)}{\partial s} = -\nu(t-s)X(t-s)\Phi_1'(s,t). \tag{84}$$

Clearly $h_k(t) = \varphi'(t,t)$ so we can solve simultaneously for $S_k(t)$ and $h_k(t)$ if we substitute the second equation in (81) for (84). As mentioned above, it is very difficult to obtain closed form expressions for the hazard and survival function when several clonal expansion stages are assumed even with constant parameters, so numerical methods are normally required. As pointed out by Little *et al.* [42] and Crump *et al.* [9], solving the system of backward Kolmogorov differential equations (81) & (84) to compute the hazard and survival functions of the model is more efficient and allows for better numerical error control than any other alternative methods available in the literature. Moreover, let $h_{k,i}(t)$ and $S_{k,i}(t)$ denote the hazard and survival function of a cell that starts in stage i at birth, $i = 1, ..., k-1$. Then,

$$S_{k,i}(t) = \Phi_{k-i}(t,t)$$

and

$$h_{k,i}(t) = -\frac{\Phi'_{k-i}(t,t)}{\Phi_{k-i}(t,t)}, \tag{85}$$

which is a nice property, especially if we are interested in comparing the cancer risk coming from cells that start at different stages of the process (see, for example, the chapter on gestational mutations).

7.2 Several Pre-initiation Stages with Constant Parameters

Luebeck and Moolgavkar [49] showed that a k-stage model with two rare and one fast event before initiation is consistent with the incidence of colorectal cancer. A graphical representation of a general k-stage model with more than one pre-initiation stages is shown in Figure 4. In this model, it is assumed that normal susceptible stem cells (stage 0) have to go through $k-2$ pre-initiation stages, before being able to expand clonally (initiated stage). Normal cells become pre-initiated according to a Poisson process with intensity $\mu_0 X(t)$, where $X(t)$ is the number of normal stem cells at age s. Each cell in the pre-initiation stage i can divide (with rate μ_i) into one stage i and one stage $i+1$ cell, $i = 1, ..., k-2$. Once a cell reaches the initiation stage $(k-1)$, it expands clonally via a birth and death process. Each time an initiated cell divides, it can produce two initiated cells (with birth rate α) or one initiated and one malignant cell (with rate μ_{k-1}). Initiated cells can also die or differentiate with rate β.

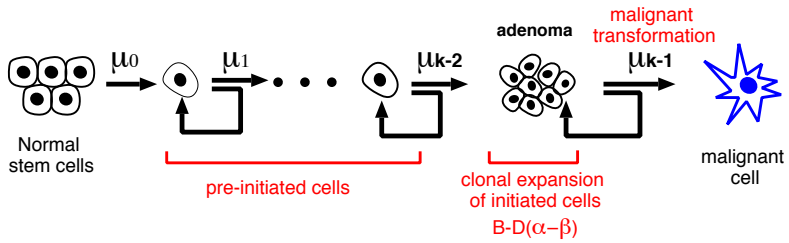

Figure 4: k-stage carcinogenesis model.

The probability generating function of the process satisfies the forward Kolmogorov differential equa-

tion [49]:

$$\begin{aligned}\frac{\partial\Psi(y_1,...,y_k;t)}{\partial t} &= (y-1)X(t)\nu(t)\Psi(y_1,...,y_k;t)\\ &\quad-\sum_{i=1}^{k-2}\{\mu_i(t)(1-y_{i+1})y_i\}\frac{\partial\Psi(y_1,...,y_k;t)}{\partial y_i}\\ &\quad+[(\alpha y_{k-1}-\alpha-\beta-\mu_{k-1}(1-y_k))y_{k-1}+\beta]\frac{\partial\Psi(y_1,...,y_k;t)}{\partial y_{k-1}},\end{aligned} \tag{86}$$

which is identical to (76) when $\alpha_i=\beta_i=0$, for $i=1,...,k-2$, i.e., when cells don't die or proliferate in any stage before $k-1$.

Using the forward Kolmogorov equation for the PGF of the process, we can obtain in a straightforward manner the survival function of the time of appearance of the first malignant cell, which is given by [49]

$$S_k(t)=\exp\{\int_0^t\mu_0X(s_1)[e^{\int_{s_1}^t\mu_1[...[e^{\int_{s_{k-2}}^t\mu_{k-3}[S_2(t-s_{k-2})-1]ds_{k-2}}-1]...-1]ds_2}-1]ds_1\}, \tag{87}$$

where $S_2(u)$ is the survival function of the TSCE model [62, 56]:

$$S_2(u)=\left(\frac{q-p}{qe^{-pu}-pe^{-qu}}\right)^{\frac{\mu_{k-2}}{\alpha}}, \tag{88}$$

and

$$p=\frac{1}{2}\left[(-\alpha+\beta+\mu_{k-1})-\sqrt{(\alpha-\beta-\mu_{k-1})^2+4\alpha\mu_{k-1}}\right], \tag{89}$$

$$q=\frac{1}{2}\left[(-\alpha+\beta+\mu_{k-1})+\sqrt{(\alpha-\beta-\mu_{k-1})^2+4\alpha\mu_{k-1}}\right]. \tag{90}$$

As an example, the 4-stage survival is given by

$$S_4(t)=\exp\left\{\int_0^t\mu_0X(s_1)\left[e^{\int_{s_1}^t\mu_1[S_2(t-s_2)-1]ds_2}-1\right]ds_1\right\}.$$

The corresponding hazard is easily calculated using definition (2). In particular, the three and four-stage hazards when the number of stem cells is assumed to be constant are [49]

$$h_3(t)=\mu_0X\left[1-\left(\frac{q-p}{qe^{-pt}-pe^{-qt}}\right)^{\mu_1/\alpha}\right] \tag{91}$$

and

$$h_4(t)=\mu_0X\left(1-\exp\left\{\int_0^t\mu_1\left[\left(\frac{q-p}{qe^{-p(t-u)}-pe^{-q(t-u)}}\right)^{\mu_2/\alpha}-1\right]du\right\}\right). \tag{92}$$

7.3 Concluding Remarks

We reviewed some of the most important multistage carcinogenesis models that have been developed. In particular, the Armitage–Doll and the TSCE model were discussed in detail. Both models have been used extensively to analyze cancer epidemiological and experimental data [1, 62, 56, 63, 54, 60, 59, 47, 29, 39, 6, 23, 28, 10, 30, 25, 27, 46, 26, 68, 52, 53, 13, 22]. Recent evidence suggests that more than two-stages are required for malignant transformation in some tissues [19, 49, 53, 45]. Some generalizations of the TSCE model that accommodate these findings were also discussed. Models with several clonal expansions have been used previously by Little [43] and Little *et al.* [41, 42, 40, 53] to analyze epidemiological data of leukemia, lung and colorectal cancer. Luebeck and Moolgavkar [49] and Meza et al. [53] used models

with several pre-initiation stages to analyze the incidence of colorectal cancer in the SEER database. Further generalizations that incorporate alternative pathways to malignancy have been analyzed by Tan [71] and Little *et al.* [44]. In particular, Little *et al.* [44, 40] used multiple-pathway models to investigate the relevance of genetic instability in colorectal carcinogenesis. From a mathematical perspective, it is relatively straightforward to extend the models discussed here to models with multiple pathways. Hence, we refer the reader to the literature for additional details.

8 Modeling Dose-Response in the TSCE and MSCE Models

The effects of carcinogenic agents on cancer risk can be evaluated by estimating the dose-response of TSCE or MSCE model parameters. In general, it is not convenient to assume apriori a specific mechanism of action of an agent. Therefore, in any analysis we should allow in principle all identifiable parameters in a model to be dose-dependent, and then use statistical analysis to identify the relevant effects. In the following we describe a standard approach to modeling the dose-response in biologically-based models.

Single agent. Let $d(s)$ denote the exposure of an individual to a specific agent at age s. We can then assume that each of the identifiable parameters of the model has a dose-response given by

$$\theta_{car}(d(s)) = \theta(1 + \theta_c d(s)^{\theta_e}), \tag{93}$$

where θ is the background parameter, and θ_c and θ_e are the dose-response coefficients. This functional form of the dose-response (power law) is quite general and covers a wide range of possible behaviors, and should suffice in most cases. For example, previous analyses using the TSCE model of the relationship between smoking (cigarettes per day) and lung cancer incidence or mortality rates suggested that power laws are good models for smoking dose-response [28, 27, 52]. However other functional forms can be also implemented if supported by additional information.

Multiple agents. When we have information about the exposure levels of several agents, we can assume that the model parameters are function of all the exposures simultaneously. In particular, let's assume that we have information about the exposure of an individual to m agents. Let $d_j(s)$ denote the exposure to the $j-th$ agent at age s, $j = 1, ..., m$. The following dose-response function has been used in the past to evaluate the effects of several agents combined [28, 26]

$$\theta_{car}(d_1(s), d_2(s), ..., d_m(s)) = \theta(1 + \theta_{c1} d_1(s)^{\theta_{e1}} + \theta_{c2} d_2(s)^{\theta_{e2}} + ... + \theta_{cm} d_m(s)^{\theta_{em}}), \tag{94}$$

where θ is the background parameter, and θ_{cj} and θ_{ej} are the dose-response coefficients corresponding to the $j-th$ agent, $j = 1, ..., m$. This functional form of the dose-response assumes that all the agents act independently from each other at the cellular level . However, this doesn't necessarily imply that the risks due to exposure to multiple agents predicted by the TSCE or MSCE models would behave in an additive way.

Effects on cell division and cell death rates. The cell division and cell death rates in the TSCE and MSCE model cannot be estimated independently when using incidence data alone, however, the approximate net cell proliferation rate, $-(p+q) = (\alpha - \beta - \mu_{k-1})$ is an identifiable parameter [31]. The effects of non-genotoxic agents on cancer risk can be evaluated by estimating the effect of such agents on the cell proliferation (or promotion) rate in the TSCE or MSCE model.

9 Analysis of Epidemiological Data

9.0.1 Cancer Registry Data

Secular trends in cancer data, like cancer registry data, can be analyzed using traditional Age-Period-Cohort (APC) models. However, these models are limited by a well-known non-identifiability problem by which any of the estimated trends can be transformed linearly on the log scale [7, 8, 33]. Replacing the age effects in an APC model with the hazard function of a biologically based model solves, at least in principle, the non-identifiability issue [49]. In addition, if exposure data is available, the use of a

biologically-based model may allow the estimation of the effects of agents exposure on cancer risk. Here we describe how to fit a biologically-based model to tabular data, while adjusting for calendar year (i.e., period) and birth cohort effects.

Likelihood Function (Poisson Regression).

Assume we have cancer incidence (or mortality) data in tabular form for j=1,...,J calendar years, covering i=1,...,I age groups. For each age-group, the number of cancer cases diagnosed during calendar year j can be assumed to follow a Poisson distribution with mean:

$$\Lambda_{i,j} = b_{i,j} c_j PY_{i,j} h(a_i), \tag{95}$$

where a_i is the mean age of the $i-th$ age group, $b_{i,j}$ and c_j are coefficients that adjust for birth cohort and calendar year effects, respectively, $PY_{i,j}$ is the person years at risk, and $h(a_i)$ represents the hazard function of a biologically-based model evaluated at age a_i.

The overall likelihood $\mathcal{L}$ for the observed incidence in all age-calendar year groups is given by

$$\mathcal{L} = \prod_{i,j} \frac{\Lambda_{i,j}^{o_{i,j}} e^{-\Lambda_{i,j}}}{o_{i,j}!}, \tag{96}$$

where $o_{i,j}$ is the number of cases in the $i-th$ age group during calendar year j.

The likelihood can be used to estimate not only the secular terms (birth cohort and period effects), but also the parameters of the biologically based model. In addition, if the data can be broken further by carcinogen(s) exposure (dose and duration) by age and period, one can also estimate in principle dose-response parameters.

9.0.2 Individual Data

In this section we briefly describe how to compute the likelihood function in case we have cancer incidence (or mortality) data with detailed individual exposure information.

Likelihood Function.

For prospective cohort data, the likelihood function is the product of individual likelihoods over all the subjects in the cohort. Assuming that participants are cancer free at their age of entry the study (ae_i), and that we censor individuals in case of death from other causes or in case they survive cancer until the end of follow-up, the individual likelihoods are given by

$$L_i(al_i, ae_i; \bar{\theta}(d_i)) = \begin{cases} -\frac{S'(al_i;\bar{\theta}(d_i))}{S(ae_i;\bar{\theta}(d_i))} = \frac{h(al_i;\bar{\theta}(d_i))S(al_i;\bar{\theta}(d_i))}{S(ae_i;\bar{\theta}(d_i))} & \text{for cancer cases,} \\ \frac{S(al_i;\bar{\theta}(d_i))}{S(ae_i;\bar{\theta}(d_i))} & \text{otherwise,} \end{cases} \tag{97}$$

where al_i is the individual's censoring or failure (cancer diagnosis) age, $h(t;\bar{\theta}(d_i))$ and $S(t;\bar{\theta}(d_i))$ denote the hazard and survival function at age t of an individual with exposure history d_i, respectively, and $\bar{\theta}(d_i)$ denotes the vector of identifiable model parameters given the exposure history d_i (see equations (93) and (94)). Note: the prime denotes derivative with respect to t.

The overall likelihood is then

$$\mathcal{L} = \prod_i L_i(al_i, ae_i; \bar{\theta}(d_i)), \tag{98}$$

where the product is taken over all the subjects in the cohort(s).

10 Examples

References

[1] P. ARMITAGE AND R. DOLL, *The age distribution of cancer and a multi-stage theory of carcinogenesis*, Br. J. Cancer, 8 (1954), pp. 1–12.

[2] ———, *A two-stage theory of carcinogenesis in relation to the age distribution of human cancer*, Br. J. Cancer, 11 (1957), pp. 161–169.

[3] M. T. BARRETT, C. A. SANCHEZ, L. J. PREVO, D. J. WONG, P. C. GALIPEAU, T. G. PAULSON, P. S. RABINOVITCH, AND B. J. REID, *Evolution of neoplastic cell lineages in Barrett oesophagus*, Nat. Genet., 22 (1999), pp. 106–109.

[4] I. BERENBLUM AND P. SHUBIK, *A new, quantitative, approach to the study of the stages of chemical carcinogenesis in the mouse's skin*, Br. J. Cancer, 1 (1947), pp. 383–391.

[5] H. BRENNER, J. CHANG-CLAUDE, C. M. SEILER, T. STURMER, AND M. HOFFMEISTER, *Does a negative screening colonoscopy ever need to be repeated?*, Gut, (2006).

[6] O. CASTREN, *Implications of a two-stage clonal expansion model to indoor radon risk assessment*, Health Phys., 76 (1999), pp. 393–397.

[7] D. CLAYTON AND E. SCHIFFLERS, *Models for temporal variation in cancer rates. I: Age-period and age-cohort models*, Stat Med, 6 (1987), pp. 449–467. Comparative Study.

[8] ———, *Models for temporal variation in cancer rates. II: Age-period-cohort models*, Stat Med, 6 (1987), pp. 469–481. Comparative Study.

[9] K. S. CRUMP, R. P. SUBRAMANIAM, AND C. B. VAN LANDINGHAM, *A Numerical Solution to the Nonhomogeneous Two-Stage MVK Model of Cancer*, Risk Anal., 25 (2005), pp. 921–926.

[10] S. B. CURTIS, E. G. LUEBECK, W. D. HAZELTON, AND S. H. MOOLGAVKAR, *A new perspective of carcinogenesis from protracted high-LET radiation arises from the two-stage clonal expansion model*, Adv. Space Res., 30 (2002), pp. 937–944.

[11] M. C. DE GUNST AND E. G. LUEBECK, *Quantitative analysis of two-dimensional observations of premalignant clones in the presence or absence of malignant tumors*, Math. Biosci., 119 (1994), pp. 5–34.

[12] M. C. DE GUNST AND E. G. LUEBECK, *A method for parametric estimation of the number and size distribution of cell clusters from observations in a section plane*, Biometrics, 54 (1998), pp. 100–112.

[13] L. DENG, M. KIMMEL, M. FOY, M. SPITZ, Q. WEI, AND O. GORLOVA, *Estimation of the effects of smoking and DNA repair capacity on coefficients of a carcinogenesis model for lung cancer*, Int. J. Cancer, 124 (2009), pp. 2152–2158.

[14] A. DEWANJI, M. J. GODDARD, D. KREWSKI, AND S. H. MOOLGAVKAR, *Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies*, Math. Biosci., 155 (1999), pp. 1–12.

[15] A. DEWANJI, E. G. LUEBECK, AND S. H. MOOLGAVKAR, *A generalized Luria-Delbruck model*, Math. Biosci., 197 (2005), pp. 140–152.

[16] A. DEWANJI, S. H. MOOLGAVKAR, AND E. G. LUEBECK, *Two-mutation model for carcinogenesis: joint analysis of premalignant and malignant lesions*, Math. Biosci., 104 (1991), pp. 97–109.

[17] A. DEWANJI, D. J. VENZON, AND S. H. MOOLGAVKAR, *A stochastic two-stage model for cancer risk assessment. II. The number and size of premalignant clones*, Risk Anal., 9 (1989), pp. 179–187.

[18] V. P. DORIA-ROSE, T. R. LEVIN, J. V. SELBY, P. A. NEWCOMB, K. E. RICHERT-BOE, AND N. S. WEISS, *The incidence of colorectal cancer following a negative screening sigmoidoscopy: implications for screening interval*, Gastroenterology, 127 (2004), pp. 714–722.

[19] E. R. FEARON AND B. VOGELSTEIN, *A genetic model for colorectal tumorigenesis*, Cell, 61 (1990), pp. 759–767.

[20] J. C. FISHER, *Multiple-mutation theory of carcinogenesis*, Nature, 181 (1958), pp. 651–652.

[21] S. H. FRIEND, R. BERNARDS, S. ROGELJ, R. A. WEINBERG, J. M. RAPAPORT, D. M. ALBERT, AND T. P. DRYJA, *A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma*, Nature, 323 (1986), pp. 643–646.

[22] D. L. GOLDWASSER AND M. KIMMEL, *Modeling excess lung cancer risk among screened arm participants in the Mayo Lung Project*, Cancer, 116 (2010), pp. 122–131.

[23] B. GRASL-KRAUPP, G. LUEBECK, A. WAGNER, A. LOW-BASELLI, M. DE GUNST, T. WALDHOR, S. MOOLGAVKAR, AND R. SCHULTE-HERMANN, *Quantitative analysis of tumor initiation in rat liver: role of cell replication and cell death (apoptosis)*, Carcinogenesis, 21 (2000), pp. 1411–1421.

[24] T. E. HARRIS, *The Theory of Branching Processes*, Dover Phoenix Editions, Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin].

[25] R. G. E. HAYLOCK AND C. R. MUIRHEAD, *Fitting the two-stage model of carcinogenesis to nested case-control data on the Colorado Plateau uranium miners: dependence on data assumptions*, Radiat. Environ. Biophys., 42 (2004), pp. 257–263.

[26] W. HAZELTON, S. MOOLGAVKAR, S. CURTIS, J. ZIELINSKI, J. ASHMORE, AND D. KREWSKI, *Biologically based analysis of lung cancer incidence in a large Canadian occupational cohort with low-dose ionizing radiation exposure, and comparison with Japanese atomic bomb survivors*, J. Toxicol. Environ. Health Part A, 69 (2006), pp. 1013–1038.

[27] W. D. HAZELTON, M. S. CLEMENTS, AND S. H. MOOLGAVKAR, *Multistage carcinogenesis and lung cancer mortality in three cohorts*, Cancer Epidemiol. Biomarkers Prev., 14 (2005), pp. 1171–1181.

[28] W. D. HAZELTON, E. G. LUEBECK, W. F. HEIDENREICH, AND S. H. MOOLGAVKAR, *Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model*, Radiat. Res., 156 (2001), pp. 78–94.

[29] W. F. HEIDENREICH, P. JACOB, AND H. G. PARETZKE, *Exact solutions of the clonal expansion model and their application to the incidence of solid tumors of atomic bomb survivors*, Radiat. Environ. Biophys., 36 (1997), pp. 45–58.

[30] W. F. HEIDENREICH, E. G. LUEBECK, W. D. HAZELTON, H. G. PARETZKE, AND S. H. MOOLGAVKAR, *Multistage models and the incidence of cancer in the cohort of atomic bomb survivors*, Radiat. Res., 158 (2002), pp. 607–614.

[31] W. F. HEIDENREICH, E. G. LUEBECK, AND S. H. MOOLGAVKAR, *Some properties of the hazard function of the two-mutation clonal expansion model*, Risk Anal., 17 (1997), pp. 391–399.

[32] W. F. HEIDENREICH, J. WELLMANN, P. JACOB, AND H. E. WICHMANN, *Mechanistic modelling in large case-control studies of lung cancer risk from smoking*, Stat. Med., 21 (2002), pp. 3055–3070.

[33] T. R. HOLFORD, *Understanding the effects of age, period, and cohort on incidence and mortality rates*, Annu Rev Public Health, 12 (1991), pp. 425–457.

[34] J. JEON, R. MEZA, S. MOOLGAVKAR, AND E. LUEBECK, *Evaluation of screening strategies for pre-malignant lesions using a biomathematical approach*, Math Biosci, 213 (2008), pp. 56–70.

[35] M. KAI, E. G. LUEBECK, AND S. H. MOOLGAVKAR, *Analysis of the incidence of solid cancer among atomic bomb survivors using a two-stage model of carcinogenesis*, Radiat. Res., 148 (1997), pp. 348–358.

[36] D. G. KENDALL, *Birth-and-death processes, and the theory of carcinogenesis*, Biometrika, 47 (1960), pp. 13–21.

[37] A. G. KNUDSON, *Mutation and cancer: statistical study of retinoblastoma*, Proc. Natl. Acad. Sci. U S A, 68 (1971), pp. 820–823.

[38] A. KOPP-SCHNEIDER, *Birth-death processes with piecewise constant rates*, Statist. Probab. Lett., 13 (1992), pp. 121–127.

[39] ——, *Carcinogenesis models for risk assessment*, Stat. Methods Med. Res., 6 (1997), pp. 317–340.

[40] M. LITTLE AND G. LI, *Stochastic modelling of colon cancer: is there a role for genomic instability?*, Carcinogenesis, 28 (2007), pp. 479–487.

[41] M. P. LITTLE, *Are two mutations sufficient to cause cancer? Some generalizations of the two-mutation model of carcinogenesis of Moolgavkar, Venzon, and Knudson, and of the multistage model of Armitage and Doll*, Biometrics, 51 (1995), pp. 1278–1291.

[42] M. P. LITTLE, R. G. E. HAYLOCK, AND C. R. MUIRHEAD, *Modelling lung tumour risk in radon-exposed uranium miners using generalizations of the two-mutation model of Moolgavkar, Venzon and Knudson*, Int. J. Radiat. Biol., 78 (2002), pp. 49–68.

[43] M. P. LITTLE, C. R. MUIRHEAD, AND C. A. STILLER, *Modelling lymphocytic leukaemia incidence in England and Wales using generalizations of the two-mutation model of carcinogenesis of Moolgavkar, Venzon and Knudson*, Stat. Med., 15 (1996), pp. 1003–1022.

[44] M. P. LITTLE AND E. G. WRIGHT, *A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data*, Math. Biosci., 183 (2003), pp. 111–134.

[45] E. LUEBECK, S. MOOLGAVKAR, A. LIU, A. BOYNTON, AND C. ULRICH, *Does folic acid supplementation prevent or promote colorectal cancer? Results from model-based predictions*, Cancer Epidemiol. Biomarkers Prev., 17 (2008), pp. 1360–1367.

[46] E. G. LUEBECK, A. BUCHMANN, D. SCHNEIDER, S. H. MOOLGAVKAR, AND M. SCHWARZ, *Modulation of liver tumorigenesis in Connexin32-deficient mouse*, Mutat. Res., 570 (2005), pp. 33–47.

[47] E. G. LUEBECK, S. B. CURTIS, F. T. CROSS, AND S. H. MOOLGAVKAR, *Two-stage model of radon-induced malignant lung tumors in rats: effects of cell killing*, Radiat. Res., 145 (1996), pp. 163–173.

[48] E. G. LUEBECK AND S. H. MOOLGAVKAR, *Stochastic analysis of intermediate lesions in carcinogenesis experiments*, Risk Anal., 11 (1991), pp. 149–157.

[49] ——, *Multistage carcinogenesis and the incidence of colorectal cancer*, Proc. Natl. Acad. Sci. U S A, 99 (2002), pp. 15095–15100.

[50] M. MACARI, E. J. BINI, S. L. JACOBS, S. NAIK, Y. W. LUI, A. MILANO, R. RAJAPAKSA, A. J. MEGIBOW, AND J. BABB, *Colorectal polyps and cancers in asymptomatic average-risk patients: evaluation with CT colonography*, Radiology, 230 (2004), pp. 629–636. Clinical Trial.

[51] R. MEZA, *Some Extensions And Applications of Multistage Carcinogenesis Models*, PhD in Applied Mathematics, Universiy of Washington, 2006.

[52] R. MEZA, W. HAZELTON, G. COLDITZ, AND S. MOOLGAVKAR, *Analysis of lung cancer incidence in the nurses' health and the hea lth professionals' follow-up studies using a multistage carcinogenesis model*, Cancer Causes Control, 19 (2008), pp. 317–328.

[53] R. MEZA, J. JEON, S. MOOLGAVKAR, AND E. LUEBECK, *Age-specific incidence of cancer: Phases, transitions, and biological implications*, Proc. Natl. Acad. Sci. U.S.A., 105 (2008), pp. 16284–16289.

[54] S. H. MOOLGAVKAR, F. T. CROSS, G. LUEBECK, AND G. E. DAGLE, *A two-mutation model for radon-induced lung tumors in rats*, Radiat. Res., 121 (1990), pp. 28–37.

[55] S. H. MOOLGAVKAR, A. DEWANJI, AND D. J. VENZON, *A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor*, Risk Anal., 8 (1988), pp. 383–392.

[56] S. H. MOOLGAVKAR AND A. G. KNUDSON, *Mutation and cancer: a model for human carcinogenesis*, J. Natl. Cancer Inst., 66 (1981), pp. 1037–1052.

[57] S. H. MOOLGAVKAR, D. KREWSKI, AND M. SCHWARZ, *Mechanisms of carcinogenesis and biologically based models for estimation and prediction of risk*, IARC Sci. Publ., 131 (1999), pp. 179–237.

[58] S. H. MOOLGAVKAR AND E. G. LUEBECK, *Multistage carcinogenesis: population-based model for colon cancer*, J. Natl. Cancer Inst., 84 (1992), pp. 610–618.

[59] ——, *Multistage carcinogenesis: population-based model for colon cancer*, J. Natl. Cancer Inst., 84 (1992), pp. 610–618.

[60] S. H. MOOLGAVKAR, E. G. LUEBECK, M. DE GUNST, R. E. PORT, AND M. SCHWARZ, *Quantitative analysis of enzyme-altered foci in rat hepatocarcinogenesis experiments–I. Single agent regimen*, Carcinogenesis, 11 (1990), pp. 1271–1278.

[61] S. H. MOOLGAVKAR AND G. LUEBECK, *Two-event model for carcinogenesis: biological, mathematical, and statistical considerations*, Risk Anal., 10 (1990), pp. 323–341.

[62] S. H. MOOLGAVKAR AND D. J. VENZON, *Two-event models for carcinogenesis: incidence curves for childhood and adult tumors*, Math. Biosci., 47 (1979), pp. 55–77.

[63] ——, *Confidence regions in curved exponential families: application to matched case-control and survival studies with general relative risk function*, Ann. Statist., 15 (1987), pp. 346–359.

[64] H. J. MULLER, *Radiation damage to the genetic material.*, Science in Progress, 7 (1951), pp. 93–165 & 481–493.

[65] J. NEYMAN AND E. SCOTT, *Statistical aspects of the problem of carcinogenesis*, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability., University of California Press, 1967, pp. 745–776.

[66] C. O. NORDLING, *A new theory on cancer-inducing mechanism*, Br. J. Cancer, 7 (1953), pp. 68–72.

[67] L. A. G. RIES, M. P. EISNER, C. L. KOSARY, B. F. HANKEY, B. A. MILLER, L. CLEGG, A. MARIOTTO, M. P. FAY, E. J. FEUER, AND B. K. EDWARDS, eds., *SEER Cancer Statistics Review, 1975-2000*, National Cancer Institute, 2003.

[68] H. SCHOLLNBERGER, M. MANUGUERRA, H. BIJWAARD, H. BOSHUIZEN, H. P. ALTENBURG, S. M. RISPENS, M. J. P. BRUGMANS, AND P. VINEIS, *Analysis of epidemiological cohort data on smoking effects and lung cancer with a multi-stage cancer model*, Carcinogenesis, 27 (2006), pp. 1432–1444.

[69] D. L. SNYDER, *Random Point Processes*, Wiley-Interscience [John Wiley & Sons], New York, 1975.

[70] W. Y. TAN, *A stochastic Gompertz birth-death process*, Statist. Probab. Lett., 4 (1986), pp. 25–28.

[71] ——, *Stochastic Models of Carcinogenesis*, Marcel Dekker, New York, first ed., 1991.

[72] A. WHITTEMORE AND J. B. KELLER, *Quantitative theories of carcinogenesis*, SIAM Rev., 20 (1978), pp. 1–30.