Toxicology and Industrial Health 2003; **19:** 125–145
www.tihjournal.com

The multistage model of cancer development: some implications

Gordon Ritter[a], Richard Wilson[a], Francesco Pompei[a,b] and Dimitriy Burmistrov[a,c]

[a]*Harvard University Department of Physics, Cambridge, USA*
[b]*Exergen Corp., Watertown, MA, USA*
[c]*Menzie-Cura and Associates, Winchester, MA, USA*

The multistage model, introduced by Armitage and Doll, was very successful at describing many features of cancer development. Doll and Peto noted a significant departure below the prediction of the model and suggested that this could be due to undercounting of cases at older ages, or to the 'biology of extreme old age.' Moolgavkar pointed out that it could also be due to the approximation used. The recent observation that cancer incidence falls rapidly above age 80 has stimulated new modelling investigations, such as the Pompei–Wilson beta model (which does reproduce the rapid fall). In the present paper, we argue that Moolgavkar's criticisms, while mathematically correct, do not affect the conclusions, particularly the constancy of the number of stages across different cancer registries (Cook, Doll and Fellingham. 1969: A mathematical model for the age distribution of cancer in man. *International Journal of Cancer* 4, 93–112). We discuss several exact solutions, compare them with the most recent data, and prove rigorously that the standard Armitage–Doll multistage model can never reproduce the sharp turnaround in cancer incidence at old age seen in the data. We discuss in detail multistage processes which have a property observed in many laboratory studies, namely that some stages progress much faster than the others. We verify mathematically the intuition that sufficiently fast stages do not appreciably affect the incidence rate of cancer, and discuss implications of this fact for cancer treatment strategies. We also show that the simplest possible modification of the Armitage–Doll model to incorporate cellular senescence just leads to the Pompei–Wilson beta model. *Toxicology and Industrial Health* 2003; **19:** 125–145.

Key words: *beta model; cancer incidence; carcinogenesis; cellular senescence; multistage model; old age*

Introduction

The idea that cancer proceeds by a number of stages was developed by Armitage and Doll (1954) soon after cancer registries became available in many countries. They noted that cancer incidence was not very reliable whereas mortality is much more objective. Therefore they plotted cancer mortality against age rather than incidence. However, even mortality was not often well described above age 70, since many death certificates had the nebulous entry 'old age'. They therefore only used the multistage model to describe cancer mortality up to age 70. Cancer incidence $I(t)$ is defined by epidemiologists as the rate of diagnosing new cases divided by the number of persons at risk. Similarly normalized cancer mortality is the rate of cancer deaths divided by the number of persons at risk. Armitage and Doll found that cancer mortality fitted a function:

Address all correspondence to: Gordon Ritter, Harvard University Department of Physics, 17 Oxford Street, Cambridge, MA 02138, USA
E-mail: ritter@fas.harvard.edu

10.1191/0748233703th195oa

$$I = a\,t^{k-1} \tag{1}$$

where k was interpreted as the number of stages in the progression of cancer, and a is proportional to the product of the probabilities at each stage. This model successfully described several important known features of cancer.

1. There is usually a latency period between the insult which appears to cause the cancer and the expression of the cancer.
2. If cancer-causing agents act at different stages of the cancer progression, and are present at high doses, then a multiplicative synergy is predicted as seen, for example, between smoking and asbestos exposure.
3. Cook *et al*. (1969) found that while k varied for different tumor sites, it is constant between countries. The constant a varied between countries, as might be expected from the different environmental situations.

These successes led to widespread use of the model. It was soon noted that above age 60, the age-specific mortality rate flattened and fell below the curve predicted by Armitage and Doll. Two major reasons for such a flattening have been suggested; the first is a variation in susceptibility (variation in a) between different members of the same group being considered, and the second is a failure of the mathematical approximations used by Doll and Armitage.

Cook, Doll and Fellingham considered variations of susceptibility in the population. They plotted incidence versus age on a logarithmic scale, for varying fractions C of sensitive persons (Cook *et al*., 2004, Figure 4, reproduced here as Figure 1). They concluded that if only 1% of the population is susceptible to cancer, then at older ages the normalized incidence curves must fall when all the 1% have developed cancer.

Even if the number of susceptibles is 10% then a flattening occurs. Figure 1, reproduced from Cook *et al*. (1969), indicates that by suitable choice of the parameter C, one might be able to describe a rapid fall off in incidence, a suggestion we discuss later. Cook, Doll and Fellingham ultimately rejected this possibility because data showed a constant age for the peak of cancer incidence, while their particular assumption of insensitive individuals

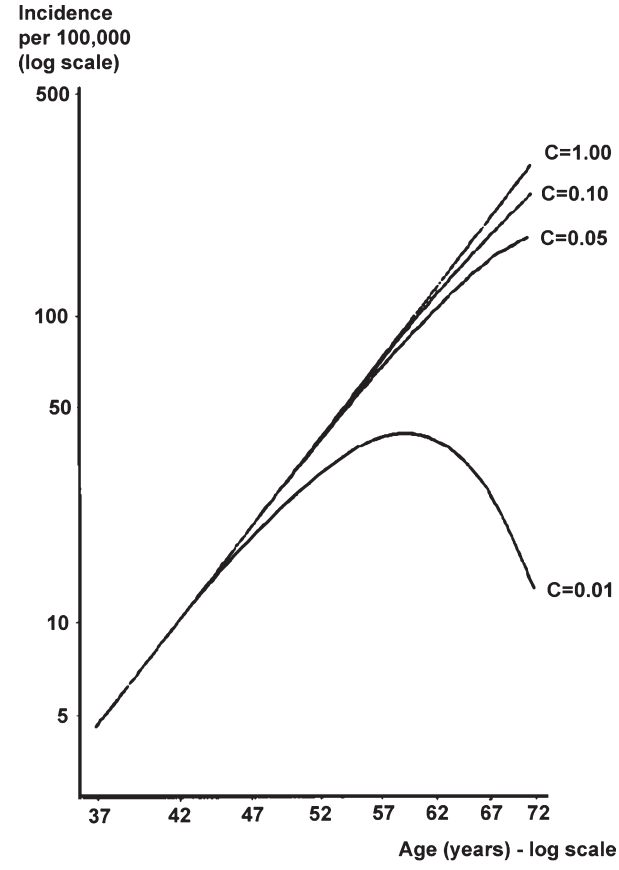


Figure 1. Cook, Doll and Fellingham (1969) considered variable susceptibility in the population. This plot shows incidence versus age on a logarithmic scale, for varying fractions C of sensitive persons. Modified from Cook *et al*. (1969), Figure 4.

indicates that incidence peaks at a younger age for rarer cancers.

The variations in cure rates for cancer in the last 50 years suggest that it is now more important to discuss incidence rather than mortality, and in addition, that incidence may now be much better determined than 50 years ago. In this paper, therefore, we reopen the discussion with an emphasis on cancer incidence rather than mortality.

Let $f(t)$ denote the event rate for the k^{th} change at time t, in a single cell. According to our definition, cancer incidence is related to $f(t)$ by $I(t) = Nf(t)$, where N is the number of affected cells. The theoretical model calculates $f(t)$; we will perform this calculation exactly for a variety of interesting situations.

In the case of the Armitage–Doll model, $I_{\text{AD}}(t) = h(t)$ is an infinitely growing hazard function, and satisfies the criterion that the cumulative probability of cancer, given by

$$F(t) \equiv 1 - \exp\left(-\int_0^t h(\tau)d\tau\right), \tag{2}$$

approaches 1 in the limit as $t \to \infty$. According to the definition of the hazard function, we consider the differential equation

$$I_{AD}(t) = at^{k-1} = \frac{F'(t)}{1 - F(t)}\,. \tag{3}$$

Solving this equation with the initial condition $F(0) = \varepsilon$ yields

$$F(t) = 1 + (\varepsilon - 1)\exp\left(\frac{-at^k}{k}\right).$$

The condition $\varepsilon \neq 0$ corresponds to fetal cancer, which, while a real condition, is ignored in this paper. Hence we are concerned with the limit $\varepsilon \to 0$,

$$F(t) = 1 - \exp(-at^k/k) \tag{4}$$

This satisfies $\lim_{t\to\infty} F(t) = 1$, which describes the fact that, according to this model, everyone will develop cancer if they do not die of something else. It is obvious from these considerations that the Armitage–Doll approximation (Equation (1)) increases without bound, and given sufficient time, any susceptible cell eventually becomes malignant.

As noted above, the Armitage–Doll model (Equation (1)) must be viewed as an approximation. The correct equations were written down by Moolgavkar (1991) and Armitage (1985), and it was seen that (Equation (1)) is only valid for small values of t. It was seen also that (Equation (1)) overestimates incidence. Moolgavkar (1991) has published a simple example supporting the idea that Armitage–Doll approximation can be inadequate if the transition rates for cancer stages are not small enough. His example presupposes a smaller number of cells $N = 10^9$ in the tissue of concern than is usually taken. In contrast to the Armitage–Doll approximation, the hazard function in the exact model has a finite asymptote $\lim_{t\to\infty} h(t) = N\mu_{\min}$, where $\mu_{\min}$ is the minimal transition rate. The suggestion that the failure of the Armitage–Doll approximation may be responsible for the observed flattening of incidence for ages above 60 follows from this.

Pompei and Wilson (2001) reopened the issue. They noted that since 1955, data on cancer incidence has improved and data on both incidence and mortality above age 70 has improved. For discussions of cancer development, incidence may now be more reliable (for many tumor sites) than mortality, due to improvements in cancer treatment. Making the assumption that the data, and particularly the SEER data in the United States, are reliable, Pompei and Wilson plotted the normalized cancer incidence data against age, finding a rapid drop in normalized incidence above age 80, which is not explained by simple parameter adjustments of the Armitage–Doll model. This is illustrated in Figure 2 for a common cancer (colon), an

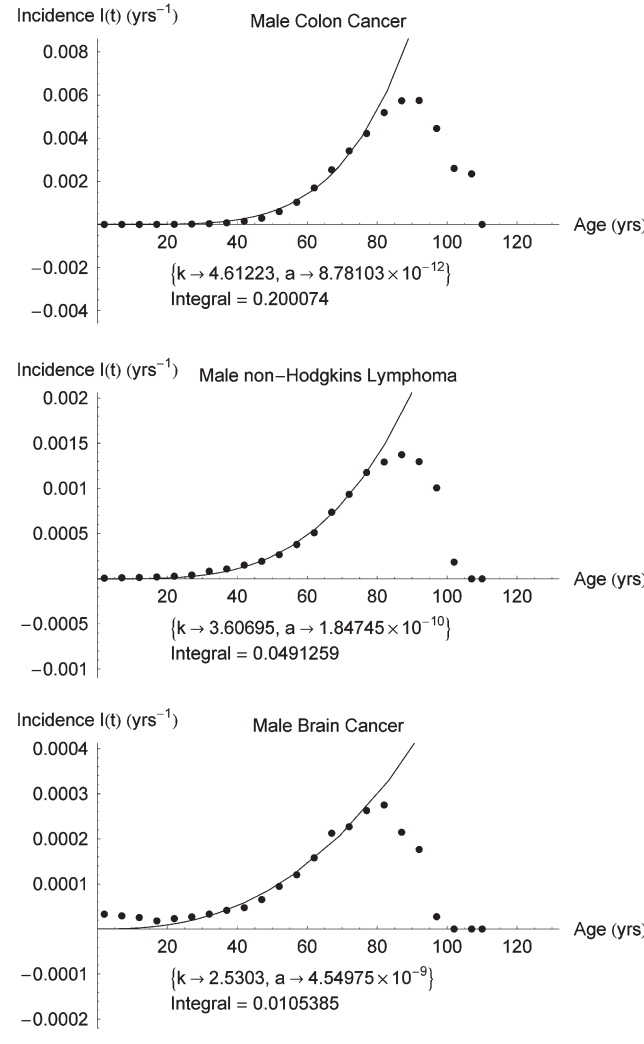


Figure 2. SEER data from Pompei *et al*. (2004) and SEER (2002) shown together with the best possible fit of the Armitage–Doll model.

intermediate-incidence cancer (non-Hodgkins lymphoma), and a rare cancer (brain).

Pompei and Wilson added an extra term to Equation (1), leading to a beta function $I = at^{k-1}(1-\beta t)$ which fits the data for most cancers remarkably well, and speculated on possible biological explanations for the extra term. Pompei, Polkanov and Wilson (Pompei *et al*., 2001) examined the limited data on laboratory mice that have been followed for their natural life of 1000 days or more rather than being sacrificed at 750 days. The age of cancer incidence is not well determined for mice, but cancer mortality suffers neither from complications of improved treatment nor from misreporting. In the absence of these systematic errors, it was observed that the cancer mortality fell to zero before the end of life. This is shown in Figure 3.

In the present paper we re-examine the criticisms of Moolgavkar, and improve upon the speculations of Pompei and Wilson. This is done by exact calculations in the multistage model. In the process we derive several closed-form solutions to the model. Following previous studies (Pompei and Wilson, 2001), we continue to make the assumption that the SEER data in the United States is reliable above age 80. We prove that under this model, as one would expect, the probability that a single cell becomes cancerous given infinite time is 1.

Crucial parameters for the multistage model are the transition rates, which we denote μ_i. We give an exact solution for a model with a number n_s of slow stages with rate μ_s and one fast stage with rate μ_f, and show that if $\mu_f \gg \mu_s$, the fast stage only slightly perturbs the final solution and may be neglected. We show that, if the multistage model is assumed to calculate the probability of cancer for a single cell, then this model can only be consistent with the rapid fall-off observed in the data if some unusual biological assumptions are made.

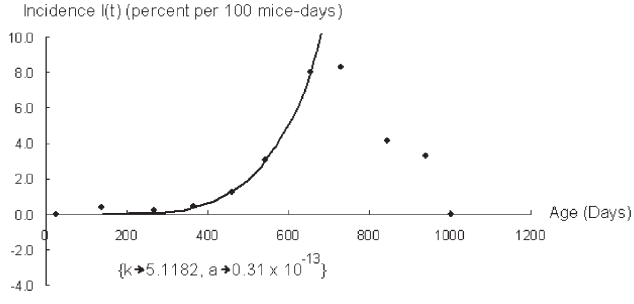


Figure 3. Mice data from Pompei *et al*. (2001) shown together with the best possible fit of the Armitage–Doll model.

Methods

Assumptions of the multistage model

In this section we repeat, starting with the fundamental assumptions, the derivation of the exact multistage model, and suggest various extensions thereof.

In the Armitage–Doll approximation, cancer incidence has a simple power law growth. This is an approximation to the exact multistage model (Moolgavkar, 1978; Moolgavkar *et al*., 1999; Pompei and Wilson, 2001) described by the Bateman equations (Bateman, 1910). The multistage model makes several basic assumptions:

1. Malignant tumors arise from a series of modifications of a single progenitor cell.
2. The process of developing a malignancy is equally likely for all cells in the same tissue.
3. The process of developing a malignancy in one cell is independent of the process in any other cell.
4. After a malignancy has developed in a cell, proliferation to an overt cancer is rapid and involves many cells in the same tissue, and may even involve metastasis to another tissue.

Under these assumptions, cancer is the last of a series of k sudden and irreversible changes which must take place in a specific order. For a cell which has experienced $(i-1)$ changes, the event rate for the next change is μ_i. We choose the convention that a single transition from a state p_0 to a state p_1,

$$p_0 \xrightarrow{\mu_1} p_1 \tag{5}$$

corresponds to one 'stage', and has $k=1$. This implies that the number of transition rate constants $\mu_1, \ldots, \mu_k$ always equals the number of stages.

Cancer in a cell

Cancer is the last of a series of k sudden and irreversible changes which must take place in a cell in a specific order. As is standard for Poisson

random processes, we neglect the probability of two or more events taking place in $(t, t+dt)$ as $dt \to 0$. This means that if a cell is in state p_i at time t, the probability of transformation to state p_{i+1} in a small time interval Δt is given by $\mu_{i+1}\Delta t + o(\Delta t)$, where $o(\Delta t)$ is very small compared to Δt, and $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$. Also, the probability of transformation $i \to i+j$ with $j>1$ in time Δt is assumed to be $o(\Delta t)$. This implies that $1/\mu_{i+1}$ is the average time required for the cell to go from state i to state $i+1$.

The probability to find a cell in the i^{th} stage by the end of time interval $(t, t+dt)$ is then given by

$$p_i(t+dt) = (1-\mu_{i+1}dt)\cdot p_i(t) + \mu_{i+1}dt\cdot 0 + \mu_i dt\cdot p_{i-1}(t) \tag{6}$$

Taking the limit $dt \to 0$, Equation (6) becomes

$$\begin{aligned} p_0'(t) &= \mu_1 p_0(t) \\ p_i'(t) &= -\mu_{i+1}p_i(t) + \mu_i p_{i-1}(t) \\ p_k'(t) &= \mu_k p_{k-1}(t)\,. \end{aligned} \tag{7}$$

The first and the last equations in this system are different because there are no stages before the stage $i=0$ (normal cell), and no stages after the k^{th} stage, which corresponds to cancer. The system (Equation (7)) should be completed with appropriate initial conditions:

$$p_0(0) = 1,\ \ p_i(0) = 0 \qquad (i = 1, 2, \ldots, k) \tag{8}$$

which mean that the cell was normal at time $t=0$. The last equation in Equation (7) also gives the probability to find the cell at its last cancerous stage. Following Armitage and Doll, we reserve the notation $f(t)$ for this quantity:

$$f(t) = p_k'(t) = \mu_k p_{k-1}(t) \tag{9}$$

Thus $f(t)$ denotes the event rate for the k^{th} change (corresponding to cancer) at time t in a single cell.

Cancer in a tissue and in humans

The event rate $f(t)$ is a probability distribution function (PDF) for appearance of cancerous cells, which means that $\int_0^\infty f(t')dt' = 1$, which means that given enough time, each cell will become cancerous. Even though the PDF is convenient for mathematical modelling, it is not directly observable. Only developed cancer in a tissue, rather than a single cancer cell, could be recognized as a cause of death, or even diagnosed at all. In practical studies cancer incidence is commonly observed. Following the notation of Armitage and Doll, cancer incidence $I(t)$ (which is the same quantity as the *hazard function* used in biostatistics and denoted there as $h(t)$) is defined by epidemiologists as the rate of diagnosing new cases, divided by the number of persons at risk.

Let us define a cumulative distribution function (CDF) for cancer per cell as $F(t) = \int_0^t f(t')dt'$, that is the probability for a cell to become a cancer by moment t. Then the probability of at least one cancer cell in a tissue of N cells by the moment t is $G(t) = 1-[1-F(t)]^N$. This definition assumes the independence of cells. Cancer incidence can be defined in terms of $G(t)$ as

$$I(t) = G'(t)/(1-G(t)) = NF'(t)/(1-F(t))$$

Cancer is a rare event even on the tissue level and therefore $F(t) \ll 1$, and hence the approximation

$$I(t) \approx Nf(t) \tag{10}$$

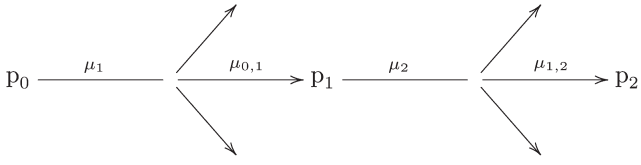
is quite good since the number of cells in a tissue, N, is greater than 10^9. Equation (10) allows fitting of the theoretically derived cancer event rate per cell $f(t)$ and the observed incidence of cases in population $I(t)$. This result was derived by Armitage and by Moolgavkar.

In the above, the function $I(t) \approx Nf(t)$ is incidence of cancer in a tissue, or equivalently, in a person. When multiplied by the size of a population, this becomes the true number of cancers in that population at risk, or the incidence. However, the number of cells N may not be large enough to ensure that the Armitage–Doll approximation is accurate, so it is important to discuss this parameter. There are about 10^{14} cells in a human body, and in any given tissue, they would be proportional to weight. Thus, as an order of magnitude approximation, the colon has roughly 10^{12} cells. The postulate that only cells on the surface are important may reduce this to 10^{11}. Therefore in what follows we take the number used by Moolgavkar (10^9 cells) as a lower limit.

The Bateman equations

The exact solution to the multistage model can be derived, by analogy, from Bateman's solution of successive radioactive decays (Bateman, 1910).

The simplest form of Bateman's solution describes a linear decay chain, leading ultimately to a stable nucleus. More complex forms have also been used to describe, for example, the decay of radium. The isotope radium C decays into radium C' and radium C'', which both decay into lead, so the diagram contains a loop. It is highly likely that these more complex forms are also relevant to cancer incidence, although in the present work we restrict attention to linear decay chains. A k-stage radioactive decay process, depicted schematically by the diagram



is described by the *Bateman equations*, a set of coupled linear first-order differential equations, in which $p_i(t)$ denotes the quantity of objects in state i at time t:

$$\frac{dp_0}{dt} = -\mu_{0,1}p_0$$

$$\frac{dp_j}{dt} = \mu_{j-1,j}p_{j-1} - \mu_{j+1}p_j \qquad (1 < i < k) \qquad (11)$$

$$\frac{dp_k}{dt} = \mu_{k-1,k}p_{k-1}$$

where $\mu_{j,j+1}$ is the partial rate constant for the transition from state j to state $j+1$, and μ_j is the total decay rate[1] of state $j-1$. The partial decay rates are related to the total decay rates by constants $b_{j,j+1}$, called *branching ratios*. This relationship takes the form

$$\mu_{j,j+1} = b_{j,j+1}\mu_{j+1}$$

In the case of a linear decay chain, the branching ratios equal unity, and the partial rate constants are given by

$$\mu_{j,j+1} = \mu_{j+1}.$$

In this simple case, Equation (11) is equivalent to the equations derived earlier in our discussion of cancer in a cell. Most discussions of the model make this assumption. However, we note that not all 'initiated' cells proceed to cancer, but the vast majority are either repaired or excreted. Some may undergo clonal expansion and multiply. This could be modelled by taking values of the branching ratio $b_{j,j+1}$ less than unity (in the first case) or greater than 1 for clonal expansion. However this is a very simplistic view of clonal expansion which treats the expansion as a definitive process rather than a stochastic one. We are attempting to understand this further. The realization that the multistage model only takes into account the slow, late limiting steps in the cancer process allows a theory in which clonal expansion occurs rapidly, but is not noticed as a separate stage.

However, if these branching ratios are the same for each cell, they may be subsumed in an overall constant factor in the formula for incidence, and the fitted values of μ adjusted accordingly. Then the algebra is the same as if each and every cell eventually becomes cancerous.

We will work with linear decay chains in what follows. In that case, the diagram of the process simplifies to

$$p_0 \overset{\mu_1}{\to} p_1 \overset{\mu_2}{\to} \cdots \overset{\mu_k}{\to} p_k \qquad (12)$$

In the earlier section, 'Cancer in a cell', it was shown that the same differential equations describe the multistage model for the formation of malignant cells. Every cell eventually transforms given sufficient time, so in a realistic model,

$$p_k(\infty) = \int_0^\infty f(t)dt = 1 \qquad (13)$$

In Lemma 3 we show that Equation (13) is a true mathematical consequence of the Bateman equations with appropriate initial conditions.

[1]In a model describing radioactivity, $\mu_{i+1} = \ln(2)/\tau_{1/2}$ where $\tau_{1/2}$ is the half-life of that species. Our notation differs slightly from that of Moolgavkar who identifies the rate constant with the state being left rather than with the state to which the transition is being made.

The most general initial conditions for the system Equation (11) are to allow $p_i(0)$ to be arbitrary nonzero constants for all i. The solution to the system Equation (11) with general initial conditions is

$$p_{m-1}(t)=\sum_{i=0}^{m-1}\left[\left(\prod_{j=0}^{m-2}\mu_{j,j+1}\right)\sum_{j=1}^{m}\frac{p_i(0)e^{-\mu_j t}}{\prod_{\substack{\ell=(i+1)\ldots m\\ \ell\neq j}}(\mu_\ell-\mu_j)}\right] \quad (14)$$

for $m=1,\ldots,k$.

A simplifying assumption, consistent with the use of Equation (11) to analyse the probability of a single cell becoming cancerous (or the radioactive decay of a single nuclide), is that the concentrations of the later stages (or daughter nuclides) are all initially zero. This assumption is equivalent to the initial conditions

$$p_0(0)=\alpha,$$
$$p_i(0)=0 \text{ (for all } i>1). \quad (15)$$

With these initial conditions, the solution (Equation (14)) reduces immediately to the simpler form, valid for $m\in\{1,\ldots,k\}$,

$$_{m-1}(t)=c_m\sum_{j=1}^{m}\chi_{j,m}\,e^{-\mu_j t}, \quad (16)$$

where $c_m=\alpha\prod_{j=1}^{m-1}\mu_{j-1,j}$

Here, c_1 is an empty product which is 1 by convention, $c_2=\mu_{1,2}$ etc., and we have defined time-independent constants

$$\chi_{j,m}=\prod_{\substack{\ell=1\ldots m\\ \ell\neq j}}(\mu_\ell-\mu_j)^{-1}. \quad (17)$$

Equations (16) and (17) were shown, for the linear chain, by Moolgavkar (1991) [(A.1), p. 217] in his work on the multistage theory of carcinogenesis. When evaluating the sum in Equation (14) the user should be warned that the terms tend to be large and of opposite sign. Indeed when the μ are equal, another formula has to be used. The cancer incidence in a tissue with N cells (which is the approximation for a person), is $Nf(t)$ to a very high degree of accuracy, and

$$f(t)=p_k'(t)=\mu_k p_{k-1}(t).$$

We state our most important mathematical assertions as a series of Lemmas, for which we provide proofs in an appendix.

Lemma 1 (Solution of the Bateman equations) *Assume that $\mu_i\neq\mu_j$ for all $i\neq j$. Then the functions $p_m(t)$ defined by Equation (16) and Equation (17) solve the Bateman equations Equation (11) with initial conditions Equation (15).*

Generally, we take the initial condition $\alpha=1$. It is easy to see that the first nonvanishing term in a Taylor series of $p_{m-1}(t)$ is of the Armitage–Doll form, where m is an index for an intermediate stage. Indeed, $p_{m-1}(t)$ is the constant c_m, times a linear combination of exponentials $\Sigma_{j=1}^{m}\chi_{j,m}\,e^{-\mu_j t}$; the n^{th} term in the Taylor series of this linear combination is clearly

$$\sum_{j=1}^{m}\chi_{j,m}\frac{(-\mu_j t)^n}{n!}. \quad (18)$$

This vanishes by algebraic identity if $n<m-1$, and for $n=m-1$, one may check that $\Sigma_{j=1}^{m}\chi_{j,m}\mu_j^{m-1}=1$, so to lowest nonvanishing order, $p_{m-1}(t)\approx c_m t^{m-1}/(m-1)!$.

The next nonvanishing term ($n=m$) also has a simple form. Equation (18) with $n=m$ becomes

$$-\frac{t^m}{m!}\sum_{j=1}^{m}\mu_j=-\frac{t^m}{(m-1)!}\overline{\mu}$$

where $\overline{\mu}=\frac{1}{k}\Sigma_{i=1}^{k}\mu_i$ denotes the average μ. We may then write:

$$p_{m-1}(t)=c_m\frac{t^{m-1}}{(m-1)!}(1-\overline{\mu}\,t+\ldots)$$

Substituting $m=k$ for the last stage, we find

$$I(t)=at^{k-1}(1-\overline{\mu}\,t+\cdots)$$

where

$$a=\frac{\prod_{j=1}^{k}\mu_j}{(k-1)!}. \quad (19)$$

This is the Taylor expansion discussed by Moolgavkar. It was referred to (Moolgavkar, 1978) as the *MacLaurin series*, but that is a special case of the more general Taylor series. We use the more general term in this paper.

As one might expect, the solution (Equation (16)) has a simple series expansion which can be expressed in closed form. This is given in the following Lemma.

Lemma 2 (Taylor expansion)
Let $\{\mu_i : i = 1 \ldots k\}$ *be any set of nonzero constants with* $\mu_i \neq \mu_j$ *if* $i \neq j$. *Then*

$$\sum_{i=1}^{m} \chi_{i,m} e^{-\mu_i t} = \sum_{j \geq m-1} \alpha_j t^j$$

with $\alpha_{m-1} = 1/(m-1)!$, *and*

$$\alpha_{j-1} = \frac{(-1)^{j-m}}{j!} \sum_{i_1 \leq \cdots \leq i_{j-m}} \mu_{i_1} \mu_{i_2} \ldots \mu_{i_{j-m}} \qquad (20)$$

if $j > m$. *This implies that*

$$p_{m-1}(t) = c_m \left(\frac{t^{m-1}}{(m-1)!} - \frac{1}{m!} \left(\sum_{i=1}^{m} \mu_i \right) t^m + O(t^{m+1}) \right)$$

Remark 1. In various applications, it is most convenient to use the lowest nonvanishing order (the Armitage–Doll approximation) in this series expansion, and it's necessary to have conditions which guarantee that this truncation is valid. We compare the lowest nonvanishing order to the (assumably better) approximation obtained by considering higher-order terms. In particular, let us compare the lowest nonvanishing order ($j = m-1$ in the above formulae) to the following term. In this paper, as with the papers of Moolgavkar, we are concerned with the last stage only and this corresponds to $m = k$ in the above equations. Let Δ_n denote the n^{th} nonzero term in the Taylor expansion, which has coefficient α_{k+n-2}.

By a simple calculation,

$$\frac{\Delta_2}{\Delta_1} = -t\,\overline{\mu}. \qquad (21)$$

This formula shows that Δ_2/Δ_1 can be arbitrarily large by inclusion of a fast stage or stages, each with large μ. When attempting, as we are doing, to find a model to fit the data from ages 20 to 70 years, the constant a in the first term of Equation (19) is heavily constrained by the data. If k is not varied, the constant a is proportional to the product $\mu_1 \cdots \mu_k$. One can only make $\overline{\mu}$ arbitrarily large if this product is fixed to the correct value. This is discussed further in the section on determination of k.

We will show later that if n_f of the stages are much faster than the remaining $n_s = k - n_f$ stages, then the exact solution for the k-stage model is approximately equal to the exact solution for the remaining n_s stages, which we call *slow* stages. Equation (21) indicates that, the Taylor series for k stages converges more slowly than the Taylor series for n_s stages because inclusion of fast stages increases the average transition constant. In the presence of fast stages, a large number of terms must be included before the Taylor series becomes accurate, and consideration of only the first and second terms in the expansion is incorrect. However, the exact solution for the model without the fast stages has a different Taylor series which converges quite rapidly. This realization leads to important practical conclusions that are discussed in more detail later. Specifically, if inclusion of a few fast stages set $\overline{\mu} \approx 10^{-3}$, then Equation (21) shows that Δ_2 is 10 percent of Δ_1 at $t = 100$. If the stages are fast enough to set $\overline{\mu} \approx 10^{-2}$ per year, then Δ_2/Δ_1 is of order one at 100 years. It is clear that, by including fast stages, Δ_2/Δ_1 can be made arbitrarily large.

Moolgavkar deliberately chose a set of parameters which included some slow, and some fast (but not very fast) stages (see for example, Moolgavkar, 1978, Table 1). Although the above estimate for Δ_2/Δ_1 does hold for Moolgavkar's example, a discussion of his chosen parameters needs more careful numerical analysis. Consider a tissue of 10^9 cells affected by a hypothetical six-stage[2] malignant tumor, with transition rates (per year) given by

[2]Moolgavkar (1991) refers to this as a seven-stage process, but only gives values for six transition constants. We assume here that six stages are meant.

$$10^{-4},\ 2\times10^{-4},\ 34\times10^{-4},\ 7\times10^{-3},$$
$$8\times10^{-3},\ 9\times10^{-3}.$$

In this example,

$$\overline{\mu} = 0.0046 \text{ per year}$$

So our formula Equation (21) predicts $\Delta_2 = 0.46\Delta_1$ at $t=100$ years. This is precisely what is observed in Figure 4, which shows the exact solution, together with its first five Taylor approximants. The numerical factor relating the second approximation, which lies below the exact curve, to the first approximation, lying above the curve, at $t=100$ years is seen to be about 1.46.

As noted in the introduction, 10^9 cells may be too few. Bearing in mind that the model is used to fit incidence data, each value of the six constants μ would have to be reduced by a factor of 2 if N were increased to 64×10^9, and the exact formula at age $t=100$ years would then be only 1.28 times the Armitage–Doll approximation.

The implications for determination of the value of k using Moolgavkar's parameters is discussed later. More generally,

$$\frac{\Delta_{n+1}}{\Delta_n} = \left(\frac{\alpha_{k+n-1}}{\alpha_{k+n-2}}\right)t = -\frac{t}{k+n}\frac{\mathcal{P}_{n,k}}{\mathcal{P}_{n-1,k}}$$

where $\mathcal{P}_{n,k}$ is the symmetric degree n polynomial in k variables, given by $\Sigma_{i_1\le\cdots\le i_n}\mu_{i_1}\cdots\mu_{i_n}$. These polynomials satisfy

$$\mathcal{P}_{n,k} = \mu_k\mathcal{P}_{n-1,k} + (\text{terms involving } \mu_1,\ldots,\mu_{k-1})$$

It follows that for $n\ge2$,

$$\frac{\Delta_{n+1}}{\Delta_n} = -\frac{t}{k+n}\left[\mu_k + \begin{pmatrix}\text{ratio approaching}\\ \text{zero as } \mu_k\to\infty\end{pmatrix}\right] \quad (22)$$

The surprising conclusion of Equation (22) is that, if we add one *very* fast ($\mu_m\sim10^4$) stage at the end, it may be necessary to consider hundreds or thousands of terms in the Taylor expansion in order to attain accurate results! The rate of convergence of this expansion is affected strongly by the appearance of fast stages, while the exact solution is not. This is a very important paradoxical result, the solution for which may be found in the earlier comments. If there are n_s slow stages and n_f very fast stages, the exact solution is close to the Armitage–Doll approximation for the n_s slow stages alone; the Armitage–Doll approximation for the full exact model with n_s+n_f stages is not accurate. For equal transition probabilities, given by the single transition constant μ, the above implies $\Delta_2/\Delta_1 = \mu t$.

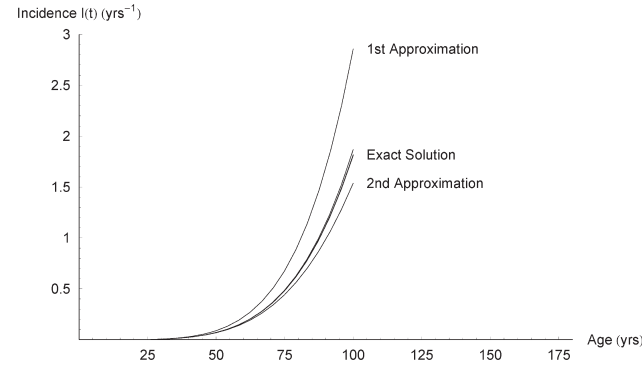


Figure 4. Exact solution for Moolgavkar's example, shown with its first five Taylor approximants.

Lemma 3 (Normalization)
Let $p_0(t),\ldots,p_k(t)$ satisfy the k-stage Bateman equations (Equation (11)) with any collection $(\mu_1,\ldots,\mu_k)$ of transition constants. Then

$$\int_0^\infty p_k'(t)dt = \left(\prod_{i=1}^k \mu_i\right)\sum_{j=1}^k \frac{\chi_{j,k}}{\mu_j} = 1 \quad (23)$$

This equation shows that given sufficient time, the cancer process will eventually reach the last stage.

General properties of the solutions

The general solution, for reasonable values of k, has far too many constants to be determined by epidemiological data alone. However, there are various cases in which the solution of the Bateman equations Equations (16) and (17) can be simplified enough to depend on only a few independent constants, even if the number of stages is large.

Equation (21) implies that the ratio of the second term to the first in the Taylor series for $I(t)$,

$$\frac{\Delta_2}{\Delta_1}$$

is proportional to the average $\overline{\mu}$. But the product of the μ's is constrained by the value of the observed incidence. With this constraint, the average $\overline{\mu}$ will

be a minimum when all the values of μ are equal. For this important special case Δ_2/Δ_1 is much smaller than, for example, in Moolgavkar's case.

Another interesting result is that *permuting the stages makes no difference to the final result* $p_k'(t)$ or, more generally, to $\mu_m p_{m-1}(t)$ for any $m \le k$. This is self-evident physically for radioactive decay, which is also described by the Bateman equations. This result, used implicitly in the Armitage–Doll approximation, is a fundamental property of the exact solution. It may be proved as follows. It suffices to consider a transposition of the ith stage and the jth stage, as any permutation may be represented as a product of transpositions. Now $\chi_{i,m}$ is a product of $(\mu_\ell - \mu_i)^{-1}$, where $\ell \neq i$. Applying the transposition, these terms become $(\mu_\ell - \mu_j)^{-1}$ except for the $\ell = j$ term, which becomes $(\mu_i - \mu_j)^{-1}$. The result is then $\prod_{\ell \neq j}(\mu_\ell - \mu_j)^{-1} \equiv \chi_{j,m}$.

A third general result is that if there are, say, 4 stages which proceed slowly (small, equal, values of μ) and many stages which act much faster (values of μ that are 10 times larger), then the final result depends only on the 4 slow stages. This is self-evident physically and was probably implicit in the thinking of earlier users of the multistage model. This is proven rigorously in Lemma 6, and discussed with a specific example in the section on measurement of k.

We pay special attention to a simple case (one not covered by Lemma 1) in which all stages have equal transition probabilities. This simple case is described by Lemma 4.

Equal transition rates

Lemma 4 (Equal transition rates)
Consider the Bateman equations $p_j'(t) = \mu(p_{j-1} - p_j)$ *arising from equal transition rates,* $\mu_1 = \mu_2 = \ldots = \mu_k \equiv \mu$. *For any natural number* $j \le k-1$, *the solution is*

$$p_j(t) = \frac{1}{j!}\, e^{-\mu t}(\mu t)^j \tag{24}$$

In particular, for the penultimate stage,

$$p_{k-1}(t) = \frac{1}{(k-1)!}\, e^{-\mu t}(\mu t)^{k-1}\,.$$

The global maximum of $p_j(t)$ *occurs at the age* $t_{peak} = j/\mu$. *Further, as before we have*

$$\int_0^\infty p_k'(t)dt = \int_0^\infty \mu p_{k-1}(t)dt = 1\,.$$

Remark 2. Since $f(t)$ applies to a single cell, μ is very small (on the order of 10^{-3}) and hence for any nonzero number of stages, $t_{\text{peak}} = j/\mu$ always occurs at a much greater age than those contained in a human lifespan. This important point is discussed in detail in the earlier section, 'Can the multistage model ever fit the data'?

The algebraic simplicity of equation Equation (24) makes it attractive as a starting point for our discussions. It has two adjustable constants, k and μ. A useful numerical example is the following. Suppose we would like to know how the exact solution Equation (24) differs from the first term in its series expansion

$$p_j(t) = \sum_{n=0}^{\infty} \frac{(\mu t)^{j+n}}{j!n!}\,. \tag{25}$$

In general, the first term in the series expansion of $e^{-x}x^n$ ($n>0$ an integer) is x^n and $|e^{-x}x^n - x^n| = x^n(1-e^{-x})$. In a numerical example with $G = 0.3$, $G' = 0.01$ per year and $k=7$ we have $\mu t \le 0.2$ for all t in a human lifespan. Then $|1 - e^{-\mu t}| \le 0.2$ and so for the first Taylor remainder R_1 of $p_j(t)$, we have

$$|R_1| \le \frac{1}{5^{k+1}} = 2.56 \times 10^{-6}\,. \tag{26}$$

Equally separated transition rates

Another case which is amenable to an exact analytical solution is when the transition constants μ_i for the various stages increase by a fixed amount. Of course, it is not necessary to assume that in the actual cancer process the transition rates occur in increasing order, thanks to invariance of the process under general permutations as remarked above, so this solution applies to the general situation when there exists *some* arbitrary reordering of the transition constants so that they increase linearly.

Lemma 5 (Equal separation)
Suppose that the transition constants $\mu = \mu_1, \mu_2, \ldots, \mu_k$ *are uniformly separated, so that* $\mu_{i+1} - \mu_i = \delta$ *for all i and for some constant* $\delta > 0$. *Then for any* $n \le k$, *the solution is given by*

$$p_{n-1}(t) = C_{\delta,\mu,n}\, e^{-\mu t}(1-e^{-t\delta})^{n-1}\,, \tag{27}$$

where $C_{\delta,\mu,n} = \dfrac{\Gamma\left(\dfrac{\mu}{\delta}+n-1\right)}{\Gamma(\mu/\delta)\Gamma(n)}$ *is a constant.*

Remark 3. To calculate the penultimate stage, $p_{k-1}(t)$, simply replace n by k in the above: $p_{k-1}(t) = C_{\delta,\mu,k}\, e^{-\mu t}(1-e^{-t\delta})^{k-1}$.

Remark 4. This model Equation (27), although an exact model with unequal transition constants, allows us to find an analytic expression for the position of the peak. Setting $p'_{k-1}(t)=0$ in Equation (27) leads to

$$t_{\text{peak}} = \delta^{-1}\ln\left(1+\frac{\delta}{\mu}(k-1)\right) \tag{28}$$

This is an exact expression. If $\delta \ll \mu$, which corresponds to the model with equal transition rates, the log can be expanded, leading to

$$t_{\text{peak}} \approx \frac{k-1}{\mu}\,. \tag{29}$$

This agrees with what was found in Lemma 4.

Remark 5. For consistency, Lemma 5 should reduce to the exact solution for equal transition rates given in Lemma 4. This is true, since for any $m \le k$, c_m clearly approaches μ^{m-1}, and

$$\begin{aligned}\frac{p_{m-1}}{c_m} &= \lim_{\delta\to 0}\left(\sum_{j=1}^{m}\chi_{j,m}e^{-\mu_j t}\right)\\ &= e^{-\mu t}\lim_{\delta\to 0}\left(\delta^{1-m}\sum_{j=1}^{m}\frac{(-1)^{j+1}}{(j-1)!(m-j)!}e^{-(j-1)\delta t}\right)\\ &= \frac{e^{-\mu t}}{\Gamma(m)}\lim_{\delta\to 0}\left(\delta^{1-m}e^{t\delta}\frac{(1-e^{-t\delta})^m}{e^{t\delta}-1}\right)\\ &= e^{-\mu t}\frac{t^{m-1}}{(m-1)!}.\end{aligned}$$

This is consistent with Lemma 4.

Remark 6. The solution obtained in Lemma 5 is an analytic expression containing no sums or products which would force the number of stages k to be an integer; therefore it is amenable to computer simulations which allow k to be a continuous variable. In particular, one can perform nonlinear regression to optimize the value of k, together with δ and μ which are the other free parameters in the model given by Equation (27).

Remark 7. The analytic solution for t_{peak} allows us to also find the height of the function at its peak:

$$p_{k-1}(t_{\text{peak}}) = C_{\delta,\mu,k}\frac{\left(\dfrac{(k-1)\,\delta}{(k-1)\,\delta+\mu}\right)^{k-1}}{\left(1+\dfrac{(k-1)\,\delta}{\mu}\right)^{\frac{\mu}{\delta}}}$$

From this, one obtains the qualitative observation that large n shifts the peak down. To see this, note that the asymptotic expansion for large k is

$$p_{k-1}(t_{\text{peak}}) \underset{\text{large } k}{\to} \text{const.}\times\frac{1}{k}+O\left(\frac{1}{k}\right)^2,$$

where the constant is $e^{-\frac{\mu}{\delta}}\left(\dfrac{\delta}{\mu}\right)^{-\frac{\mu}{\delta}}\Gamma\left(\dfrac{\mu}{\delta}\right)^{-1}$. This is important, because it explains why, when we attempt to fit real-world data, the curve-fitting routine finds unrealistically high values of k, the number of stages.

Fast and slow stages

In the discussion of Equation (21), we outlined the effect of including a stage much faster than the others on the Taylor expansion. In this section we analyse this in a more precise way (Figure 5). To illustrate the general effect which occurs when one, or a few, stages are much faster than the rest, we first consider a simple example, which entails the convergence of the two-stage solution to the single-stage solution, as the transition rate for the second stage becomes large. We exhibit this convergence in the function $p_k(t)$, which corresponds to the *last* stage, in Equations (30) and (31). If $k=1$ we have (for $I(t)=p'_k(t)$)

$$I(t) = \mu_1 e^{-\mu_1 t}\,, \tag{30}$$

and if $k=2$,

$$I(t) = \frac{\mu_1\mu_2}{\mu_2 - \mu_1}\, e^{-\mu_1 t} + \frac{\mu_1\mu_2}{\mu_1 - \mu_2}\, e^{-\mu_2 t}. \qquad (31)$$

One may now observe directly that Equation (31) converges to Equation (30) in the limit that $\mu_2 \gg \mu_1$. In fact, this phenomenon is quite general, as follows from Lemma 6.

Lemma 6 (Slow and fast stages)
Suppose that the first n_s stages have transition constant μ_s, followed by one stage with transition constant μ_f. Then the exact solution is given by

$$p_{n_s+1}(t) = \frac{(\mu_s)^{n_s+1}}{n_s!}\, e^{\mu_f t} \int_0^t e^{(\mu_f-\mu_s)x} x^{n_s} dx$$
$$= e^{\mu_f t}\left(\frac{\mu_s}{\mu_s - \mu_f}\right)^{n_s+1}\left[1 - \frac{\Gamma(1+n_s,\ t(\mu_s-\mu_f))}{n_s!}\right] \qquad (32)$$

Further, we assert that

$$\lim_{\mu_f\to\infty}(\mu_f p_{n_s+1}(t)) = \mu_s p_{n_s}(t). \qquad (33)$$

The statement of Lemma 6 places the fast stage at the end. However, previously we have shown that permuting the stages has no effect on the final solution, i.e., on the incidence function. Therefore Lemma 6, and in particular Equation (33), applies to a multistage progression in which any one of the stages is very fast compared to the others. By induction, the same is then true of the case with multiple fast stages.

This is an exact solution; no approximations have been made. Our intuition is that if the fast stage is fast enough, then it will have no effect on the properly normalized solution, and Equation (32) will reduce to $p_{n_s}(t)$. This intuition is rigorously expressed as Equation (33). We note that both sides of Equation (33) integrate to one. This generalizes the simple observation following Equation (31). We illustrate this general phenomenon with a specific example. Consider the Bateman equations with n_s stages at rate v, and the remaining $k-n_s$ stages at rate μ. Let $g(t,\mu,v)$ denote the solution $p_k'(t) = \mu p_{k-1}(t)$, and let

$$S(t,v) = \frac{v}{n_s!}\, e^{-tv}(tv)^{n_s}$$

denote the exact solution for n_s slow stages of rate v. To illustrate that the effect of the fast stages becomes negligible for $\mu \gg v$, we plot $g(t,nv,v)$ and $S(t,v)$ on the same graph for increasing values of n, with $k=5$ total stages and $n_s=3$ slow stages. For the slow transition constant, we take a reasonable value of 5×10^{-3}, which is comparable to the values chosen by Moolgavkar (1991). The relevant Bateman equations are

$$p_0'(t) = -\mu p_0(t),$$
$$p_1'(t) = \mu p_0 - v p_1,$$
$$p_2'(t) = v p_1 - v p_2, \ldots, p_5'(t) = v p_4$$

We conclude in this numerical example that for $\mu = 50v$, the effect of the faster stages is in the order of a few per cent. In the limit as $\mu/v \to \infty$, we observe that $g(t,\mu,v)$ converges to $S(t,v)$ for all t.

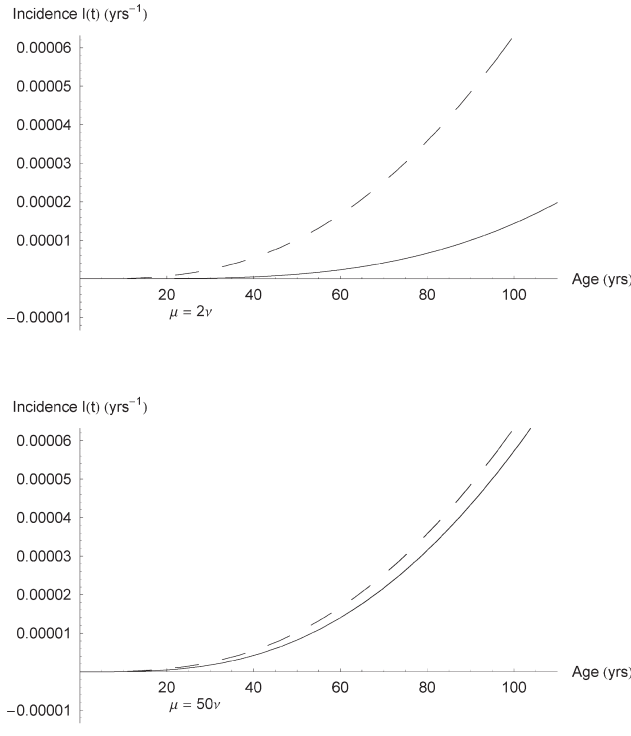


Figure 5. Exact solutions with three slow stages, having rates $v = 5\times10^{-3}$, followed by two 'fast' stages at rate μ. Graphs are shown for $\mu = 2v$ and $\mu = 50v$. The solution $S(t,v)$ with $n_s=3$ slow stages and no fast stages is shown as a dashed line for comparison.

Results and discussion

On the number of stages

In the introduction we pointed out that a fundamental success of the multistage model (1) was by

showing that whereas normalization constant a in the incidence formula varied from country to country, number of stages k stayed constant. This they did by plotting $\ln I$ against $\ln t$. The slope was $(k-1)$.

How does this change when we go to the exact model? The theorem that one very fast stage makes no difference to the result suggests that Armitage and Doll were measuring the number of slow stages. However, Moolgavkar showed a plausible set of parameters where the Armitage–Doll approximation was 1.46 times the exact model at age $t=100$. We show that a straight-line fit in log–log scale is still good approximation in this example, and the value of k that Armitage and Doll would have derived in such a case is only slightly less than the true number of stages. We investigate this as follows.

We simulate data from the exact model using the Moolgavkar parameters (Moolgavkar, 1978). Then we analyse these simulated data in the Armitage–Doll manner to derive k by fitting these data using a least squares fit.

In Figure 6, line 1 we show the simulated data using Moolgavkar's parameters (squares) and the best linear fit of $\ln I$ versus $\ln t$. From the slope of this line, 4.8068, we see that Armitage and Doll would have derived 5.8 stages instead of the correct

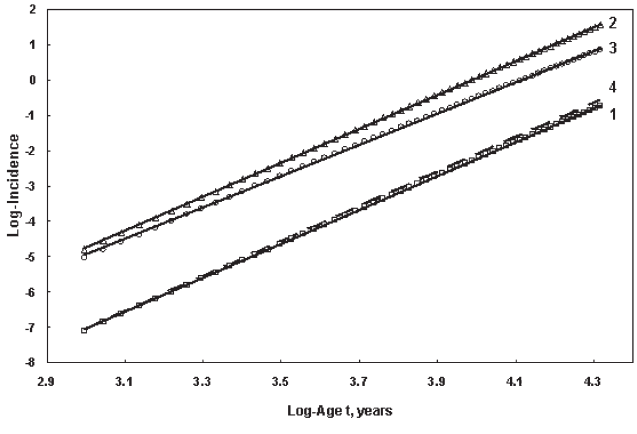


Figure 6. Armitage–Doll approximation (line fit in log–log scale) of the exact multistage model in the example proposed by Moolgavkar (1991). **1**. Squares: exact model, $N=1e9$, $k=6$, $\mu_1=0.0001$, $\mu_2=0.0002$, $\mu_3=0.0034$, $\mu_4=0.007$, $\mu_5=0.008$, $\mu_6=0.009$; Line (solid): least square fit, $\ln I=4.8068\ln t-21.46$; **2**. Triangles: exact model, $\mu_1=0.001$ (increased 10 times), Line: least square fit $\ln I=4.8001\ln t-19.139$; **3**. Circles: exact model, $\mu_6=0.09$ (increased 10 times), Line: least square fit $\ln I=4.4395\ln t-18.261$; **4**. Dashed line: exact model $\mu_1=0.00005$, $\mu_2=0.00001$, $\mu_3=0.0017$, $\mu_4=0.0035$, $\mu_5=0.004$, $\mu_6=0.0045$ (all transition rates decreased 2 times), $N=6.4e10$ (increased 64 times).

6.0. This is greater than the small number, 2, of slowest stages but less than the total number of stages. If the number of cells is 64 times greater than Moolgavkar assumed, then the values of μ must all be reduced a factor of 2 to give similar incidence (Figure 6, line 4), and the derived number of stages, 5.9, is even closer.

We can go further and ask whether such a derived value of k is stable under changing an individual transition rate μ, corresponding to a difference between cancer registries in different countries. Line 2 of Figure 6 shows a similar plot with one of the stages, μ_1, increased tenfold to 0.001. The value of k that Armitage and Doll would have derived is the same, 5.8. Increasing the transition rate tenfold for the fastest stage μ_6 resulted in $k=5.4$, which deviates from the exact k by less than 10%. We conclude that, although using an approximation can reduce the derived number of stages, this reduction is not dramatic, and it is stable across registries so that one of the principal successes of the multistage model is unaltered.

That many more mutations are present in cancers than the 4–8 slow stages predicted from the epidemiological data is well supported. For example a test for mutated DNA for colon cancer includes 21 specific mutations (Tagore *et al*., 2003). Far larger numbers of mutations are known to exist (Lengauer *et al*., 1998; Duensing and Munger, 2002) as a consequence of genetic instabilities caused by early stage alterations, and 11 000 are reported by Stoler *et al*. (1999) for colon cancers. Clearly the vast majority of these alterations must occur very rapidly and thus do not affect the age distribution of cancer, which is determined by the much slower (and therefore rarer) rate limiting stages.

As a strategy for reducing cancer incidence, it appears that it is much more productive to develop environmental, diet or behavioral strategies which would further slow (making them less probable) the slow stages to reduce cancers, rather than strategies which make fast stages less probable, which would not reduce cancers appreciably. It is a challenge to cancer biologists to classify the stages as fast or slow in order to identify the most effective prevention strategies.

There are arguments that numerous stages are unnecessary for modelling, which entail a two-stage clonal expansion (Moolgavkar and Luebeck, 2003). However, it is important to recognize that the additional stages are part of the carcinogenesis process, and it is clearly of interest to possess models which reflect, in as much detail as possible, the correct structure of the carcinogenesis process.

Can the multistage model ever fit the data?

A salient feature of the data is that cancer incidence appears to fall off fast above age 80, to nearly zero at age 100 (Pompei *et al*., 2004). Consider an exact model for $I(t)=Nf(t)$ with equal transition rates,

$$N\frac{\mu(t\mu)^k}{e^{t\mu}k!}$$

which, as discussed above, has its peak at $t_{\text{peak}}=k/\mu$. Since k is the number of stages, it is at least of order one, and on the other hand, if the model is applied to a single cell, then epidemiological data requires that, even for common cancers, the transition rate μ must be $O(10^{-3})$. Therefore $t_{\text{peak}}\approx O(10^3)$ and the turnaround can never occur within a human lifespan.

We now give a second argument that the turnaround is not relevant within a human lifespan. Note that $p_k'(t)$ is a linear combination of exponentials, so for any set of positive, real transition constants, it increases from zero to a peak, turns over, and asymptotically approaches zero as $t\to\infty$. Qualitatively, this has the same feature as the data, and a more careful analysis is necessary to see the problem.

As noted by Heidenreich *et al*. (2002), the incidence function

$$I(t)=\frac{Np_k'(t)}{1-p_k(t)} \tag{34}$$

increases to a finite asymptote as $t\to\infty$. With stages and transition constants based on fitting real-world data, $p_{\text{k}}(t)$ is around 10^{-9} or smaller for $0<t<100$ years, so in this range, $Np_k'(t)$ is well approximated by Equation (34), which is monotone increasing. However, the time scale at which $p_k'(t)$ starts to turn over is precisely the scale at which $p_k(t)$ becomes of order one, and the approximation $p_k(t)\ll 1$ is no longer valid. In conclusion, we have the following assertion.

Lemma 7 (Turnover)
The turnover in $p_k'(t)$ can never explain the turnover in the data.

It is, of course possible to fit an exact solution with $N_{\text{cells}}=10^{10}$ to the (monotone increasing) part of the data set between 20 years and 70 years. We consider both equal transition constants (Figure 7) and equally separated transition constants (Figure 8). In the region shown in Figure 7, which corresponds to a human lifespan, the lowest-order Taylor approximation to the exact solution with equal transition rates is an extremely good approximation (see Remark 1 above) and thus the exact curve is indistinguishable to the Armitage–Doll approximation $I=at^{k-1}$.

Interestingly, even though we only fitted the equal separation model to the data up to age 70, one can see in Figure 8 that the equal separation model correctly *predicted* the next few data points.

Susceptibility differences

An early attempted explanation of the decrease in incidence after age 80 is that the cancer-susceptible people are being depleted. Thus the incidence levels off and drops to zero when all the senistive people have developed cancer. This was described by Cook *et al*. (1969) who used the Armitage–Doll approximation and made the simple assumption that only a fixed fraction of population is susceptible (sensitive), e.g., able to develop cancer. We derive a

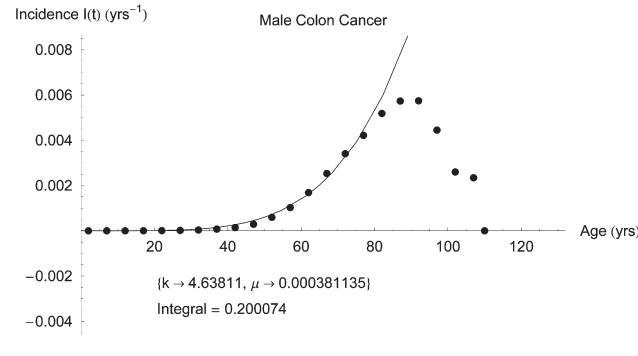


Figure 7. Fit of exact solution with equal transition constants to male colon cancer data, with $N_{\text{cells}}=10^{10}$.

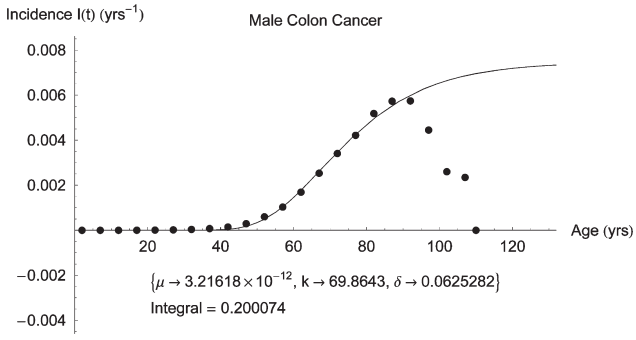


Figure 8. Fit of exact equal separation model to male colon cancer data, with $N_{\text{cells}} = 10^{10}$. Although the fit was done using data up to age 70, the model remains accurate to age 90.

more general expression for the age-specific incidence. Let $S(t)$ be the number of sensitive persons at time t. Then

$$\frac{d}{dt}S(t) = -S(t)I_1(t)$$

where $I_1(t)$ is the incidence per sensitive person. Let P be the total number of persons at $t=0$. This leads to

$$S(t) = PC\exp\left(-\int_0^t I_1(t')dt'\right)$$

where C is the fraction of sensitive persons. The incidence in the population, which includes both sensitive and insensitive persons, is

$$\begin{aligned} I(t) &= \frac{\text{new cases per unit time}}{\text{persons at risk at time } t} \\ &= \frac{-dS(t)/dt}{(1-C)P + CP\exp\left(-\int_0^t I_1(t')dt'\right)} \\ &= \frac{CPe^{-\int_0^t I_1(t')dt'}I_1(t)}{(1-C)P + CPe^{-\int_0^t I_1(t')dt'}} \\ &= \frac{I_1(t)}{1 + \left(\frac{1-C}{C}\right)\exp\int_0^t I_1(t')dt'} \end{aligned} \tag{35}$$

The choice $C=1$ corresponds to the assumption that all people are sensitive, and implies $I(t)=I_1(t)$. With $C=0$, which corresponds to a completely insensitive population, we have $I(t)=0$. Further, for any $0<C<1$, we have $I(t) \to 0$ as $t \to \infty$. Thus if there are any immune persons in the population, $I(t)$ will eventually turn over and tend to zero which might enable such an equation to fit the data. The challenge is whether with reasonable assumptions the incidence will approach zero for all cancers within a human lifetime, as SEER data suggests. If we define $G = \left(\frac{1-C}{C}\right)\exp\int_0^t I_1(t')dt'$ then Equation (35) implies that the equation determining the peak, $I'(t)=0$, takes the form

$$\frac{I_1'}{I_1^2} = \frac{G}{G+1}$$

In the general case, one would not expect to solve this except numerically, but for the Armitage–Doll case, $I_1(t)=at^{k-1}$ which implies $I_1'/(I_1)^2 = \frac{k-1}{a}t^{-k}$ and we have

$$\frac{k-1}{a} = \frac{t^k(1-C)}{Ce^{-at^k/k} + 1 - C}$$

Taylor expanding the right hand side gives the approximate solution

$$t_{\text{peak}} \approx \left(\frac{k-1}{ak!(1-C)}\right)^{1/k}. \tag{36}$$

For typical laboratory numbers, $k=6$, $a=10^{-13}$, $C=0.1$, the approximation Equation (36) gives $t_{\text{peak}} \approx 65$. However, this depends sensitively on k. With $I_1(t)$ given by the exact model with equal transition rates (ETR), in a tissue with N cells we find

$$I_{\text{ETR}}(t) = \frac{\frac{N}{\Gamma(k)}e^{-\mu t}(\mu t)^{k-1}}{1 + \frac{1-C}{C}\exp\left(\frac{N}{\mu}\frac{\Gamma(k) - \Gamma(k,\ \mu t)}{\Gamma(k)}\right)} \tag{37}$$

The fact that the function $I_{\text{ETR}}(t)$ presented in Equation (37) is extremely sensitive to its fundamental parameters presents a problem from a modelling perspective. With typical biological parameters of $N \approx 10^{13}$ and $\mu \approx 10^{-3}$, the factor N/μ appearing under the exponential in Equation (37) is of order 10^{16}. To obtain a reasonable function, the

factor $\frac{\Gamma(k) - \Gamma(k,\ \mu t)}{\Gamma(k)}$ must cancel this divergence, which implies fine-tuning of k and μ.

Cook *et al*. (1969) rejected this approach because it appeared that the peak shifted with tumor probability. In contrast we find that we can directly fit the SEER data for our three example cancers by adjusting the parameter values a, k, and C, and obtain the results in Figure 9. Also shown for comparison is the fit obtained with the senescence model of (Pompei and Wilson (2001, 2004) $I(t) = at^{k-1}(1-\beta t)$. The fitted values of a and k are similar for both the susceptibility and senescence fits. Figure 10 shows the three cancers plotted on the same axes, each with a corresponding susceptibility model fit.

Both these fits to the data use the A–D approximation, but we believe that use of the exact model will not make a change in the general conclusions. The senescence model modifies the A–D assumption of the inevitability of cancer at the cellular level by including the idea that, due to senescence, every cell has an increasing probability with age that it never reaches the cancerous final stage.

The susceptibility and senescence fits are very similar except for age $t > 105$ years. The susceptibility fit has a small tail and incidence never reaches zero, while incidence does reach zero in the senescence fit. The SEER data (and the ED01 mouse data) actually reach zero incidence, but the number of men alive at this age ($\sim$700) is small and age reporting is unverified, thus there is some uncertainty that the true incidence number is zero. The susceptibility model requires that the fraction of susceptible persons C be very different for each cancer to fit the data (0.185, 0.046, 0.0095 for the cancers shown). Although it is widely believed that there are differences up to a factor of 3 or so in individual susceptibility, there is no indication in the study of chemical produced cancers that there is such a large difference in susceptibility.

The senescence conjecture has important possible implications on the relationship between cancer and longevity, as well as on the age distribution of cancer, as discussed by Pompei and Wilson (2004). The basic idea is that as each cell ages it tends to progress to a senescent stage, a state characterized by normal functioning but inability to divide and repair genomic damage. There is evidence, although not conclusive, that as tissue ages a larger fraction of cells senesce, and thus tissue is slow to repair. The mathematical modelling makes the simple assumption that at the end of a human lifespan ($\sim$110 years) all cells are senescent, and lack of repair capacity is related to the end of life. It is the same lack of repair capacity by cell division that stops the carcinogenesis process, and thus produces the sharp turnover in cancer incidence when a significant fraction of cells are unable to reproduce.

Senescence and the beta model

Interestingly, the beta model introduced by Pompei and Wilson (2001) may also be derived from a simplified model of senescence, which we call *deterministic senescence*. For this derivation, we assume that cancer cells are mortal in the sense of Hayflick (1965), which means that a finite length is imposed on each chain of cell divisions. At each point in the process, the number of divisions remaining is called the *residual doubling potential*. Assume that senescence does not affect the stage

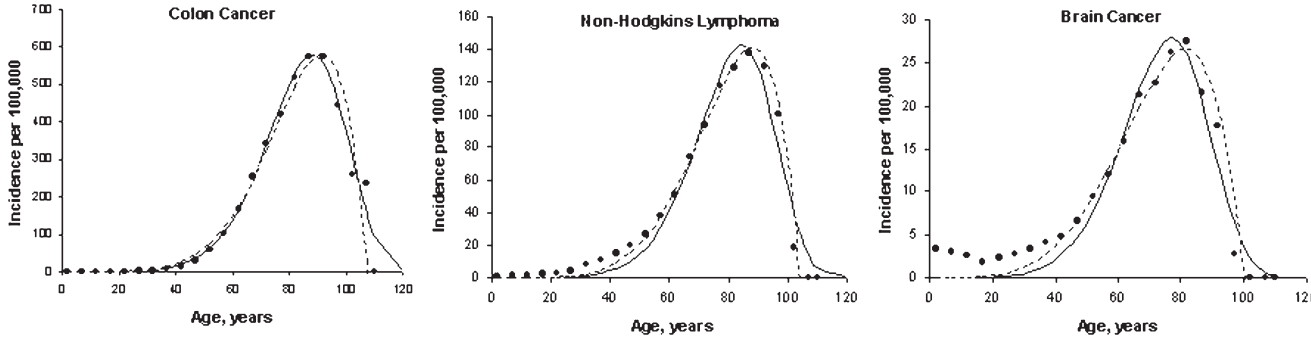


Figure 9. Comparison between SEER data (●), susceptibility model fit (—, $a = 6.3\times10^{-13}$, 2.2×10^{-13}, 9.6×10^{-12}, $k = 6.68$, 6.98, 6.22; $C =$ 0.185, 0.046, 0.0095) and senescence model fit (— -, $a = 1.90\times10^{-13}$, 1.36×10^{-13}, 4.7×10^{-12}, $k = 6.77$, 5.44; $\beta = 0.0093$, 0.0096, 0.01).

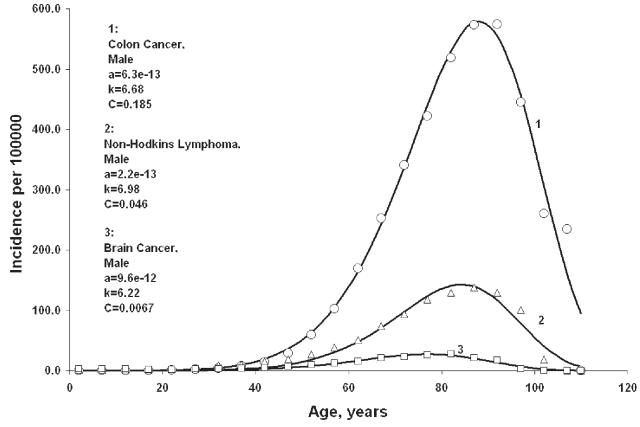


Figure 10. Comparison between SEER data for male colon cancer (○), non-Hodgkins lymphoma (△) and brain cancer (□), each with susceptibility model fit.

transitions or transition rates in the Armitage–Doll process. Assume a constant (deterministic) cell division rate; then the residual doubling potential for a cell decreases linearly with age.

If a malignant cell is completely senescent, so that the residual number of doublings is zero, then this cell does not produce observable cancer. More generally, if the residual doubling potential is small, then the malignant cell is limited to a small number of descendants, and this produces an unobservable tumor, or one which is effectively destroyed by the immune system. A larger tumor has a proportionally greater chance of survival.

Armitage–Doll theory assumes that each malignant cell leads to a tumor, but naturally this cell must undergo many divisions to produce an observable tumor. Accordingly, the Armitage–Doll model $I=at^{k-1}$ must be modified to reflect the possibility that the cell, as it progresses to cancer, may be simultaneously approaching mortality in the sense of Hayflick. The appropriate modification which takes account of this effect is to introduce a linear multiplicative factor, which must be of the form $(1-\beta t)$, and which now represents the conditional probability for an observable tumor, given one cell which is malignant (or in the k^{th} stage of the multistage process) at time t. The incidence function now takes the form of the Pompei–Wilson beta model. As before, the parameter β may be determined through curve-fitting. As a caveat to this approach, it must be noted that the existence of *in vivo* senescence in the human body has not been established.

If the explanation is accepted that cellular senescence is responsible for the rapid fall off in cancers above age 80, there are several consequences, some of which are discussed in Pompei and Wilson (2004). For example, it is not sufficient for a drug or environmental agent (or lack thereof) to be shown to reduce cancer. If the action of the drug or agent is to increase senescence to reduce cancer, then it will be accompanied by the serious side effect of reduction in longevity. Alterations in the p53 gene have been shown to do this.

Melatonin, a known antioxidant that reduces DNA damage, has been shown in mice experiments to both increase cancers and increase longevity, suggesting that antioxidants might require more careful consideration. If a drug or environmental agent could be targeted to one or more specific stages in the multistage cancer interpretation, then perhaps reduction in cancer might be accomplished without reduction in longevity. This is seems to be the case when known strong carcinogens are prevented from acting on cells, such as stopping smoking, or stopping β-naphthylamine inhalation.

Another consequence is that drugs such as cortisone, which are carcinogenic and dangerous for use on the young, may well be relatively safe for anyone over age 80, for treatment, for example, of arthritis. Further study of this point is very important. If modeling and tests bear this out, it would be welcome news for the aged.

At the very least, anyone testing or regulating an anticancer drug must be aware of the possibility that the drug might be acting as an accelerator of senescence, which would reduce life expectancy, and demand the appropriate test regimen.

Conclusions

This study set out to investigate whether it is possible to describe cancer incidence above age 80 using the exact multistage model but no other assumptions. We prove rigorously that it is not possible. In the process we have proved several theorems about the multistage model that may be useful in further investigations. For example, we conclude that the multistage model considers only the slow stages, and that the incidence function is invariant under permutation of the rate constants. Common cancers are thought to have 4–8 slow

stages; our conclusion is that there could be many more fast stages, without an appreciable effect on the age distribution of cancer. The possible number of these 'fast' stages depends on how much faster they are. If fast stages are present, they may dramatically affect the convergence of the Taylor expansion.

We found that with additional biological assumptions, it is possible to fit the data and we presented two possibilities. One is the addition of cellular senescence; the other is an assumption that most people are not susceptible to cancer, and in fact for rare cancers, very few people are susceptible.

The incidence data alone cannot decide between the model that there are large differences in susceptibility, and the model that cellular senescence is important, or whether either one is likely. We merely present the data fits, and the problems in their understanding for others to consider.

Acknowledgements

The authors would like to thank Suresh Moolgavkar for helpful discussions.

References

Armitage, P. 1985: The multistage models of carcinogenesis. *Environmental Health Perspectives* 53, 195–201.

Armitage, P. and Doll, R. 1954: The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* VIII, 1–12.

Bateman, H. 1910: Solutions of a system of differential equations occurring in the theory of radioactive transformation. *Proceedings of the Cambridge Philosophical Society* 15, 423–27.

Cook, P.J., Doll, R. and Fellingham, S.A. 1969: A mathematical model for the age distribution of cancer in man. *International Journal of Cancer* 4, 93–112.

Duensing, S. and Munger, K. 2002: Human papillomaviruses and centrosome duplication errors: modeling the origins of genomic instability. *Oncogene* 21, 6241–48.

Hayflick, L. 1965: The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research* 37, 614; 36.

Heidenreich, W.F., Luebeck, E.G. and Hazelton, W.D. 2002: Multistage models and the incidence of cancer in the cohort of A-bomb survivors. *Radiation Research* 158, 607–14.

Lengauer, C., Kinzler, K.W. and Vogelstein, B. 1998: Genetic instabilities in human cancers. *Nature* 396, 643–49.

Moolgavkar, S.H. 1978: The multistage theory of carcinogenesis and the age distribution of cancer in man. *Journal of the National Cancer Institute* 61, 49–52.

Moolgavkar, S.H. 1991: Carcinogenesis models: an overview. *Basic Life Science* 58, 387–96.

Moolgavkar, S.H., Krewski, D. and Schwarz, M. 1999: Mechanisms of carcinogenesis and biologically-based models for quantitative estimation and prediction of cancer risk. In Moolgavkar, S.H., Krewski, D. and Zeise, L. *et al*., editors, *Quantitative Estimation and Prediction of Cancer Risk*. *IARC Scientific Publications*. Lyon, France: Oxford University Press, 131.

Moolgavkar, S.H. and Luebeck E.G. 2003: Multistage carcinogenesis and the incidence of human cancer. *Genes, Chromosomes and Cancer* 38, 302–306.

Pompei, F. and Wilson, R. 2001: The age distribution of cancer: the turnover at old age. *Health and Environmental Risk Assessment* 7, 1619–50.

Pompei, F. and Wilson, R. 2004: A quantitative model of cellular senescence influence on cancer and longevity. *Toxicology and Industrial Health* 18, 365–76.

Pompei, F., Polkanov, M. and Wilson, R. 2001: Age distribution of cancer in mice: the incidence turnover at old age. *Toxicology and Industrial Health* 17, 1–16.

Pompei, F., Lee, E.E. and Wilson, R. 2004: Cancer turnover (Letter). http://www.nature.com/nrc (accessed January 2004).

Pompei, F., Lee, E.E. and Wilson, R. Cancer incidence turnover to age 110+. Gerontological Society of America Annual Meeting, Nov. 2004, Washington, DC.

SEER Public-Use 1995–1999 when Using SEER*Stat: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence-SEER 11 Regs, Nov 2001 Sub (1973–2000), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2002, based on the November 2001 submission.

Stoler, D.L., Chen, N. and Basik, M. *et al*. 1999: The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proceedings of the National Academy of Science* 96, 15121–26.

Tagore, K.S., Lawson, M.J. and Yucaitis, J.A. *et al*. 2003: Sensitivity and specificity of a stool DNA multitarget assay panel for the detection of advanced colorectal neoplasia. *Clinical Colorectal Cancer* 3, 47–53.

Appendix A

Proofs of assertions

Proof of Lemma 1.

$$c_{m+1} = \mu_{m-1,m} c_m \text{ and } \chi_{j,m+1} = (\mu_{m+1} - \mu_j)^{-1} \chi_{j,m}$$

for all $j \neq m+1$.

$$\begin{aligned}
\frac{d}{dt} p_m &= c_{m+1} \sum_{j=1}^{m+1} \chi_{j,m+1} (-\mu_j) e^{-\mu_j t} \\
&= c_{m+1} \sum_{j=1}^{m+1} \chi_{j,m+1} (\mu_{m+1} - \mu_j) e^{-\mu_j t} \\
&\quad - c_{m+1} \sum_{j=1}^{m+1} \mu_{m+1} \chi_{j,m+1} e^{-\mu_j t} \\
&= \mu_{m-1,m} c_m \sum_{j=1}^{m+1} \chi_{j,m} e^{-\mu_j t} - \mu_{m+1} p_m \\
&= \mu_{m-1,m} p_{m-1} - \mu_{m+1} p_m
\end{aligned}$$

as desired.

Proof of Lemma 2. Equation (20) is proved by double induction on m and j. The series $\Sigma \alpha_j t^j$ converges absolutely for all $t \in \mathbb{C}$, since Equation (16) is a linear combination of exponentials.

Proof of Lemma 3. $p_k'(t) = \mu_k p_{k-1}(t)$, so the lemma is equivalent to the statement that for $0 < i < k$,

$$\int_0^\infty p_i(t) dt = 1/\mu_{i+1}.$$

Generally speaking, $p_i(t)$ represents the number of objects in state i of the process (normalized by the initial condition). Considering the process for one object forces the initial condition to be $p_0(0) = 1$, $p_i(0) = 0$, $i > 1$, and it follows that $p_i(t) = 1$ if the object is in state i, and zero otherwise. It is then obvious that $\int_0^\infty p_i(t) dt$ does not depend on the rates to get into state i, and is inversely proportional to the rate to get out of state i. Therefore

$$\int_0^\infty p_i(t) dt = 1/\mu_{i+1}$$

(this can also be proved by a tedious direct computation). Setting $i = k-1$ gives the result that

$$\int_0^\infty \mu_k p_{k-1}(t) dt = 1,$$

as desired.

Proof of Lemma 4.

$$\begin{aligned}
p_j'(t) &= \mu \left[\frac{1}{(j-1)!} e^{-\mu t} (\mu t)^{j-1} - \frac{e^{-\mu t} (\mu t)^j}{j!} \right] \\
&= \mu (p_{j-1} - p_j).
\end{aligned}$$

We infer that the global maximum occurs at the solution of $p_j = p_{j-1}$, which immediately gives $t = j/\mu$. The value of the integral follows by direct computation, as well as by Lemma 3 in the limit of equal transition rates.

Proof of Lemma 5. $\mu_j = \mu + (j-1)\delta$, $j \geq 1$, so that $\mu_p - \mu_j = (p-j)\delta$. This implies that

$$\begin{aligned}
\chi_{j,n} &= \delta^{1-n} \prod_{\substack{p=1\ldots n \\ p \neq j}} \frac{1}{p-j} = \delta^{1-n} \prod_{i=1}^{j-1} \frac{1}{(-i)} \prod_{i=1}^{j-1} \frac{1}{i} \\
&= \frac{\delta^{1-n} (-1)^{j+1}}{(j-1)!(n-j)!}
\end{aligned}$$

$$\begin{aligned}
c_n &= \prod_{j=1}^{n-1} \mu_j = \prod_{i=0}^{n-2} (\mu + i\delta) = \delta^{n-1} \left(\frac{\mu}{\delta} \right)_{n-1} \\
&= \delta^{n-1} \frac{\Gamma\left(\frac{\mu}{\delta} + n - 1 \right)}{\Gamma(\mu/\delta)}
\end{aligned}$$

where in the last line we have made use of the Pochhammer symbol or 'forward factorial' and its identity in terms of gamma functions

$$(a)_n = a(a+1)\ldots(a+n-1) = \Gamma(a+n)/\Gamma(a)$$

We find by direct calculation that

$$\sum_{j=1}^{n} \frac{(-1)^{j+1}}{(j-1)!(n-j)!} e^{-(j-1)\delta t} = \frac{e^{t\delta} (1 - e^{-t\delta})^n}{(e^{t\delta} - 1)\Gamma(n)}$$

hence

$$p_{n-1}(t) = c_n \sum_{j=1}^{n} \chi_{j,n} e^{-\mu_j t}$$

$$= \frac{\Gamma\left(\frac{\mu}{\delta} + n - 1\right)}{\delta^{n-1}\Gamma(\mu/\delta)} \sum_{j=1}^{n} \frac{(-1)^{j+1} e^{-(\mu+(j-1)\delta)t}}{(j-1)!(n-j)!}$$

$$= \frac{\Gamma\left(\frac{\mu}{\delta} + n - 1\right)}{\Gamma(\mu/\delta)} e^{-\mu t} \frac{e^{t\delta}(1 - e^{-t\delta})^n}{(e^{t\delta} - 1)\Gamma(n)}$$

$$= \left(\frac{\Gamma\left(\frac{\mu}{\delta} + n - 1\right)}{\Gamma\left(\frac{\mu}{\delta}\right)\Gamma(n)}\right) e^{-\mu t}(1 - e^{-t\delta})^{n-1}$$

This completes the proof.

Proof of Lemma 6. Let A be any constant. In general, the solution of the differential equation

$$f'(t) = Af(t) + g(t), \qquad f(0) = 0$$

where g is an arbitrary function and f is to be solved for, is given by

$$f(t) = e^{At} \int_0^t e^{-Ax} g(x) dx \,.$$

Suppose that stages zero through n_s are described by the transition constant μ_s, and the stages n_s through k are described by transition constant μ_f. Later we will consider the limit $\mu_f \to \infty$ while keeping μ_s fixed. This means that

$$\frac{d}{dt} p_{n_s+1} = \mu_s p_{n_s} - \mu_f p_{n_s+1} \,.$$

By the above reasoning with $A = -\mu_f$, $g = \mu_s p_{n_s}$, this implies that

$$p_{n_s+1}(t) = e^{-\mu_f t} \int_0^t e^{\mu_f x} \mu_s p_{n_s}(x) dx \,.$$

However, from our solution of the model with equal transition rates, we know that

$$p_{n_s}(x) = \frac{1}{n_s e^{-\mu_s x} (\mu_s x)^{n_s}}$$

so we have

$$p_{n_s+1}(t) = \frac{(\mu_s)^{n_s+1}}{n_s!} e^{-\mu_f t} \int_0^t e^{(\mu_f - \mu_s)x} x^{n_s} dx.$$

This is as far as we can go without the use of special functions. To proceed, note that

$$\int_0^t e^{-\alpha x} x^{\beta} dx = \frac{\beta\Gamma(\beta) - \Gamma(1 + \beta, t\alpha)}{\alpha^{\beta+1}} \,.$$

Apply this with $\alpha = \mu_s - \mu_f$, and $\beta = n_s$. This gives

$$\int_0^t e^{(\mu_f - \mu_s)x} x^{n_s} dx = \frac{n_s\Gamma(n_s) - \Gamma(1 + n_s,\ t(\mu_s - \mu_f))}{(\mu_s - \mu_f)^{n_s+1}}$$

We infer that

$$p_{n_s+1}(t) = \frac{(\mu_s)^{n_s+1}(n_s! - \Gamma(1 + n_s,\ t\mu_s - t\mu_f))}{n_s! e^{t\mu_f}(\mu_s - \mu_f)^{n_s+1}}$$

$$= e^{-\mu_f t} \left(\frac{\mu_s}{\mu_s - \mu_f}\right)^{n_s+1} \times \left[1 - \frac{\Gamma(1 + n_s,\ t(\mu_s - \mu_f))}{n_s!}\right]$$

We note that the last line is precisely Equation (32), and completes the proof of that assertion. It remains to examine the limit as $\mu_f \to +\infty$. Elementary methods of asymptotic analysis show that for fixed a and $|z| \to \infty$,

$$\Gamma(a, z) = z^{a-1} e^{-z} (1 + O(z^{-1})) \,.$$

Thus, neglecting terms proportional to inverse powers of μ_f, we find

$$\begin{aligned}\mu_f p_{n_s+1}(t) &\sim \mu_f e^{\mu_f t}\left(\frac{\mu_s}{\mu_s-\mu_f}\right)^{n_s+1}\\ &\quad\times\left[1-\frac{(\mu_s-\mu_f)^{n_s}e^{-t(\mu_s-\mu_f)}t^{n_s}}{n_s!}\right]\\ &\sim -e^{-\mu_s t}\mu_f\left(\frac{\mu_s}{\mu_s-\mu_f}\right)\mu_s^{n_s}\frac{t^{n_s}}{n_s!}\\ &= -e^{-\mu_s t}\left(\frac{\mu_s}{\mu_s/\mu_f-1}\right)\frac{(\mu_s t)^{n_s}}{n_s!}\\ &\to \mu_s\frac{1}{n_s!}e^{\mu_s t}(\mu_s t)^{n_s}=\mu_s p_{n_s}(t)\end{aligned}$$

This completes the proof. We remark in passing that

$$\begin{aligned}p_{n_s+1}(t) = \left(\frac{\mu_s}{\mu_s-\mu_f}\right)^{n_s+1} e^{-t\mu_f}\gamma(n_s\\ +1, t(\mu_s-\mu_f))/n_s!\end{aligned}$$

where γ is the lower incomplete gamma function.