DE-FG02-08ER64605

LOW DOSE RADIATION CANCER RISKS: EPIDEMILOGY AND TOXICOLOGY MODELS

July 1, 2008 – December 31, 2009

David G. Hoel, PhD

The basic purpose of this one year research grant was to extend the two stage clonal expansion

model (TSCE) of carcinogenesis to exposures other than the usual single acute exposure.

The two-stage clonal expansion model of carcinogenesis incorporates the biological process of carcinogenesis, which involves two mutations and the clonal proliferation of the intermediate cells, in a stochastic, mathematical way. The current TSCE model serves a general purpose of acute exposure models but requires numerical computation of both the survival and hazard functions. The primary objective of this research project was to develop the analytical expressions for the survival function and the hazard function of the occurrence of the first cancer cell for acute, continuous and multiple exposure cases within the framework of the piece-wise constant parameter two-stage clonal expansion model of carcinogenesis. For acute exposure and multiple exposures of acute series, it is either only allowed to have the first mutation rate vary with the dose, or to have all the parameters be dose dependent; for multiple exposures of continuous exposures, all the parameters are allowed to vary with the dose. With these analytical functions, it becomes easy to evaluate the risks of cancer and allows one to deal with the various exposure patterns in cancer risk assessment.

A second objective was to apply the TSCE model with varing continuous exposures from the cancer studies of inhaled plutonium in beagle dogs. Using step functions to estimate the retention functions of the pulmonary exposure of plutonium the multiple exposure versions of the TSCE model was to be used to estimate the beagle dog lung cancer risks.

The mathematical equations of the multiple exposure versions of the TSCE model were developed. A draft manuscript which is attached provides the results of this mathematical work. The application work using the beagle dog data from plutonium exposure has not been completed due to the fact that the research project did not continue beyond its first year.

Two Stage Clonal Expansion Models of Carcinogenesis for Acute, Continuous, and Multiple Exposure

Boshao Zhang*, David G. Hoel**

Department of Biostatistics, Bioinformatics, and Epidemiology

Medical University of South Carolina

135 Cannon Street, Suite 303 , Charleston, SC 29425, USA

Telephone: (843) 876-1100, Facsimile: (843) 876-1126

Email addresses: * zhang@musc.edu ** hoel@musc.edu

Abstract

The two-stage clonal expansion model of carcinogenesis incorporates the biological process of carcinogenesis, which involves two mutations and the clonal proliferation of the intermediate cells, in a stochastic, mathematical way. The current model serves general purpose and requires numerical computation of the survival and hazard functions. We derived the analytical expressions for the survival function and the hazard function of the occurrence of the first cancer cell for acute, continuous and multiple exposure cases within the framework of the piece-wise constant parameter two-stage clonal expansion model of carcinogenesis. For acute exposure and multiple exposures of acute series, it is either only allowed to have the first mutation rate vary with the dose, or to have all the parameters be dose dependent; for multiple exposures of continuous series, all the parameters are allowed to vary with the dose. With these analytical functions, it becomes easier to evaluate the risks of cancer and allows one to deal with the various exposure patterns in cancer risk assessment.

1. Introduction

From 1970s through 1990s Moolgavkar, Venzon and Knudson [1-3] further developed the two-hit theory of cancer proposed by Nordling [4] and the multi-stage model proposed by Armitage and Doll [5], by adding the clonal expansion of intermediate cells to the carcinogenesis process, and translated the biological mechanism into a stochastic, mathematical model --- the two-stage clonal expansion model (TSCE) for carcinogenesis. Since then, a large number of theoretical and applied studies [6-33] have been published. Among them, the studies by Little *et al* [30-33] generalized the TSCE model to multistage cases with theoretical results and applications.

The TSCE model [3] assumes that within a pool of *N(t)* normal cells each cell has a probability to be initiated and become an intermediate abnormal cell with a rate of $\mu_1(t)$, or with an overall rate of $\nu(t) = \mu_1(t)N(t)$ for the pool of cells; an intermediate cell may divide into two daughter cells with a rate of $\alpha(t)$, or die or differentiate with a rate of $\beta(t)$, or further be transformed into a cancer cell with a rate of $\mu(t)$ (in this paper "first mutation rate" refers to $\mu_1(t)$, "second mutation rate" refers to $\mu(t)$. Here, we assume that *N(t)* is a constant.).

(Figure 1)

With the assumption of a non-homogeneous Poisson process for all the transitions mentioned above (see Figure 1), Moolgavkar *et al.* [3] developed the probability generating function for the number of intermediate and malignant cells at time t. They then applied the Kolmogorov forward differential equation to the probability generating

function, and solved the characteristic equations associated with the Kolmogorov equation and derived the survival function and hazard function for the occurrence of the first malignant cell at time *t*:

$$S(t) = \exp \int_0^t [y(u,t) - 1]\nu(u)du \tag{1}$$

$$h(t) = -\int_0^t y_t(u,t)\nu(u)du \tag{2}$$

where $y(u,t)$ is the solution to the Riccati equation

$$\frac{dy}{du} = -R(y,u) = -\{\alpha y^2 - (\alpha + \beta + \mu)y + \beta\} \tag{3}$$

with $y_t(u,t) = \frac{\partial}{\partial t} y(u,t)$

For the piece-wise constant parameter two-stage model, $y(u,t)$ can be solved explicitly in piece-wise form [3]. In an application of radon-induced lung tumor in rats, Moolgavkar *et al.* [20] applied numerical methods to evaluate the survival function and hazard function. Since then researchers had been attempting to derive analytical forms for these functions. For the time-independent parameter model, Kopp-Schneider, Portier and Sherman [34] derived the closed form solution of the incidence function, and Heidenriech, Jacob and Paretzke [8, 10] derived the hazard function. The hazard function derived from Kopp-Schneider's incidence function appears different from Heidenriech's [10] formula, but it can be easily shown that the two are equivalent. For the piece-wise constant parameter model Heidenreich, Luebeck and Moolgavkar [9] derived the survival function and hazard function in recursive form, avoiding numerical differentiation and integration.

We consider two basic types of exposure in terms of duration: acute exposure and continuous exposure. We define acute exposure as the one occurs instantaneously at a time point and a continuous exposure as one persists for a fixed period of time. Multiple exposures can be composed of a series of acute exposure, or a series of continuous exposure, or some mixture of these two basic types. As such, acute exposure and continuous exposure are just special cases of multiple exposures. For the modeling for acute exposure, Jacob and Jacob [35] approximated an instantaneous effect on the first mutation rate using a short time interval of 1 minute, and allowed a delayed effect on the proliferation rate for a week in their study on atomic bomb survivors. Heidenreich, Jacob and Paretzke [9, 10] proposed an approach allowing the first mutation rate to vary instantaneously while treating the other three parameters as constants, resulting in an analytical formula for the time dependent parameter model for acute exposure. For continuous case, there are no analytical formulae available.

The purpose of our paper is to derive the analytical forms of the survival function and hazard function for the acute, continuous and multiple exposure patterns within the framework of the piece-wise constant parameter model. Once the analytical formulae for the acute and continuous exposure are obtained, they are used as the building blocks for the multiple exposure cases. With these analytical formulae we can deal with a variety of exposure patterns.

2. Acute exposure

In describing the effect of an acute exposure, mathematically we take the limit of the effect, say, hazard, with the length of the exposure duration approaching zero.

Heidenreich, Jacob and Paretzke [10] derived the hazard function for acute exposure allowing the first mutation rate to jump (equivalent to a jump for ν) and then fall back to the background level. We call this simplified acute exposure model as the simple acute exposure model. For the purpose of generalization we will use several equivalent expressions that Kopp-Schneider, Portier and Sherman [34] derived for the constant parameter model and Heidenrich , Jacob and Paretzke [10] derived for the simple acute exposure model as follows.

Let *A, B* denote the smaller and greater roots of the quadratic equation associated with the Riccati equation (3):

$$\alpha y^2 - (\alpha + \beta + \mu) y + \beta = 0 \tag{4}$$

and $g = \alpha(B - A)$ (5)

then, for any time *t* the survival function for the control group can be obtained by completing the integration in equation (1):

$$S_0(t) = \exp\left\{ v_0 \left[(B_0 - 1)t + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 t}} \right] \right\} \tag{6}$$

where subscript 0 indicates the parameters for the control group. Expression (6) can be proved equivalent to Kopp-Schneider's incidence formula [34].

The hazard function can be obtained by taking the negative derivative of the logarithm of the survival function (6):

$$h_0(t) = \nu_0(1-A_0)\left[1 - \frac{B_0 - A_0}{(1-A_0) - (1-B_0)e^{g_0 t}}\right] \tag{7}$$

Let t_1 be the time when the acute exposure occurs. For $t \geq t_1$, with the assumption that only ν is affected by the exposure, completing the integration in (1) at t_1, using the integration of the impulse function (Dirac delta function) [10] [36], gives the survival function of the occurrence of the first cancer cell:

$$S(t) = \exp\left\{\ln(S_0(t)) + \left[\frac{B_0(1-A_0) + A_0(B_0-1)e^{g_0(t-t_1)}}{(1-A_0) + (B_0-1)e^{g_0(t-t_1)}} - 1\right]\nu_0 rd\right\} \tag{8}$$

with the jump of ν described by the linear additive model:

$$\nu = \nu_0(1+rd) \tag{9}$$

Here *r* is the unit excess relative increase in ν induced by the exposure of dose *d*. Thus, we obtain a slightly different representation of the hazard function as derived by Heidenriech, Jacob and Paretzke [10]:

$$h(t) = h_0(t) + \frac{(1-A_0)(B_0-1)(B_0-A_0)g_0 e^{g_0(t-t_1)}}{\left((1-A_0) + (B_0-1)e^{g_0(t-t_1)}\right)^2}\nu_0 rd \tag{10}$$

In (8-10), the term *rd* can be replaced by functional form *f(d)* with *f(0) = 0*. This applies to the dose response of ν in all the formulas for acute exposure and multiple acute exposure cases in following sections.

2.1 A model with more than one time-dependent parameter

Here we consider a more complete model by treating the effect of acute exposure on the first mutation rate as an instant jump, and also allowing other parameters to vary for a limited period of time after the exposure. After that, all the parameters return to their background levels.

We use parameter subscripts to indicate whether the parameters are affected by the exposure. Thus, α_0, β_0, ν_0, μ_0, A_0 , B_0 , g_0 are for the intervals before the exposure and the time after the effect of exposure terminates, while α_d, β_d, ν_d, μ_d, A_d , B_d , g_d are for the interval in which the effect of the exposure on the parameters other than the first mutation rate is active.

Let t_q be the time when the effect on the parameters other than the first mutation rate terminates. For $t \geq t_q$, integrating the exponential part of (1) at t_1 over the three intervals, i.e., $[0, t_1)$, $[t_1, t_q)$, $[t_q, t]$, gives the logarithm of the survival function for the exposed:

$$\log(S_d(t)) = \nu_0 \left[(B_0 - 1)t_1 + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} \right] + \nu_0 rd(y_1 - 1)$$

$$+ \nu_0 \left[(B_d - 1)(t_q - t_1) + \frac{1}{\alpha_d} \log \frac{B_d - A_d}{(y_2 - A_d) - (y_2 - B_d)e^{g_d (t_q - t_1)}} \right]$$

$$+ \nu_0 \left[(B_0 - 1)(t - t_q) + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_q)}} \right] \qquad (11)$$

where $y_1 = y(t_1, t)$, $y_2 = y(t_q, t)$ are values of the function $y(u,t)$ calculated backward at right ends of the first two intervals, with $y_3 = y(t,t) = 1$, the value of $y(u,t)$ at t given time t, defined for the third interval. Moolgavkar and Luebeck [3] provided the details for

the calculation of the values of the function $y(u,t)$ for the piece-wise constant parameter case. It follows from (11) that

$$h_d(t) = \frac{\nu_0 y_{1t}}{\alpha_0} \frac{1-e^{g_0 t_1}}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} + \nu_0 r d y_{1t}$$

$$+ \frac{\nu_0 y_{2t}}{\alpha_d} \frac{1-e^{g_d (t_q - t_1)}}{(y_2 - A_d) - (y_2 - B_d)e^{g_d (t_q - t_1)}} + \nu_0 (1 - A_0) \left[1 - \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_q)}} \right] \quad (12)$$

where $$y_{1t} = \frac{(B_d - A_d)^2 e^{g_d (t_q - t_1)}}{[(y_2 - A_d) - (y_2 - B_d)e^{g_d (t_q - t_1)}]^2} y_{2t} \quad (13)$$

$$y_{2t} = y_t(t_q, t) = \frac{\partial}{\partial t} y(t_q, t) = \frac{(1 - A_0)(1 - B_0)(B_0 - A_0) g_0 e^{g_0 (t - t_q)}}{\left[(1 - A_0) - (1 - B_0)e^{g_0 (t - t_q)} \right]^2} \quad (14)$$

In (13) and (14) the subscript *t* is used to indicate the partial derivatives of the functions with respect to *t*.

For $t_1 \le t \le t_q$, the intervals of concern are [*0, t1*), [*t1, t*]. Accordingly, the logarithm of the survival function and the hazard function reduce to:

$$\log(S_d(t)) = \nu_0 \left[(B_0 - 1)t_1 + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} \right] + \nu_0 r d (y_1 - 1)$$

$$+ \quad \nu_0 \left[(B_d - 1)(t - t_1) + \frac{1}{\alpha_d} \log \frac{B_d - A_d}{(1 - A_d) - (1 - B_d)e^{g_d (t - t_1)}} \right] \quad (15)$$

$$h_d(t) = \frac{\nu_0 y_{1t}}{\alpha_0} \frac{1-e^{g_0 t_1}}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} + \nu_0 r d y_{1t} + \nu_0 (1 - A_d) \left[1 - \frac{B_d - A_d}{(1 - A_d) - (1 - B_d)e^{g_d (t - t_1)}} \right]$$

(16)

with $y_2 = y(t,t) = 1$, and y_{1t} obtained by replacing the parameters in (14) by A_d, B_d, g_d and t_q by t_1.

For the case $t \le t_1$, the survival function and hazard function are the same as the ones for the control group. Thus we have the three-piece survival function and hazard function for the exposed group for all *t*.

In later sections, $y_i = y(t_i, t)$, and $y_{it} = \frac{\partial}{\partial t} y(t_i, t)$ are used to denote the right endpoint values of the function $y(u,t)$ for multiple intervals and their partial derivatives. As we observed in equations from (11) to (16), the parameters for the intervals determine the endpoint values, and the values of the survival function and hazard function given *t*. Thus extra attention should be paid to the last interval for the case where the subject dies between the time when the exposure starts and the time when the effect of exposure terminates. For simplicity, survival functions and hazard functions will be given only for those subjects who live beyond the time when the possible effect of exposure terminates.

It can be shown that (11) and (12) reduce to (8) and (10) through rearrangement respectively, if other three parameters than the first mutation rate are set at their background levels. Mathematically the dose responses of the three parameters other than the first mutation rate can take any functional form.

3. Continuous exposure

In this section the three interval case which Moolgavkar described in the appendix in his application [20] will be discussed. The intervals are: before exposure, during exposure, and after exposure. In the intervals of before and after exposure, the parameters of the model---first and second mutation rates, division rate, and death and differentiation rate of intermediate cell--- are all at the background level. In the exposure interval the

exposure has an effect on the four parameters. Here we assume that the effect of the exposure on parameters disappears right after the exposure terminates.

Let t_s be the start time of the exposure and t_q be the stop time of the exposure. For the case $t \geq t_q$, the logarithm of the survival function and hazard function can be derived from (1), (2):

$$\log(S_d(t)) = \nu_0 \left[(B_0 - 1)t_s + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_s}} \right]$$

$$+ \nu_d \left[(B_d - 1)(t_q - t_s) + \frac{1}{\alpha_d} \log \frac{B_d - A_d}{(y_2 - A_d) - (y_2 - B_d)e^{g_d (t_q - t_s)}} \right]$$

$$+ \nu_0 \left[(B_0 - 1)(t - t_q) + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_q)}} \right] \qquad (17)$$

$$h_d(t) = \frac{\nu_0 y_{1t}}{\alpha_0} \frac{1 - e^{g_0 t_s}}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_s}} + \frac{\nu_d y_{2t}}{\alpha_d} \frac{1 - e^{g_d (t_q - t_s)}}{(y_2 - A_d) - (y_2 - B_d)e^{g_d (t_q - t_s)}}$$

$$+ \nu_0 (1 - A_0) \left[1 - \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_q)}} \right] \qquad (18)$$

If we need to add a post exposure effect period after t_q for some parameters, the functions above can be derived accordingly. For the general cases, the exposure period can be divided into any number of intervals, in which parameters are assumed constant.

4. Multiple exposures

The models in section 2 and 3 can be generalized to multiple exposure cases to suit different situations.

4.1 Multiple exposures of acute series

Suppose n multiple doses $d_1, d_2, \ldots, d_n$ are acutely applied to a subject at $t_1, t_2, \ldots, t_n$. For simplicity here we assume that the first mutation rate takes jumps at n exposures with the linear response model (9) and the effects on A, B, α will be effective between any two exposures and in a period of x after the last exposure, then fall back to the background levels. Let $t_q = t_n + x$. For $t \geq t_q$, there are $n+2$ intervals involved in the computation of the survival function and hazard function. The parameters are the same as the ones for the control group for the first and last interval, and are changed accordingly for n intervals in the middle. Let $\alpha_i, \beta_i, \nu_i, \mu_i, A_i, B_i, g_i$ denote the parameters and derived parameters for the i^{th} interval, the logarithm of the survival function of cancer can be derived:

$$\log(S_d(t)) = \nu_0 \sum_{i=1}^{n+2} \left[(B_i - 1)(t_i - t_{i-1}) + \frac{1}{\alpha_i} \log \frac{B_i - A_i}{(y_i - A_i) - (y_i - B_i) e^{g_i (t_i - t_{i-1})}} \right] + r\nu_0 \sum_{i=1}^{n} (y_i - 1) d_i \tag{19}$$

with $t_0 = 0$, $t_{n+1} = t_q$, $t_{n+2} = t$ and $y_i = y(t_i, t)$ is the value of $y(u,t)$ for the right end of the interval (t_{i-1}, t_i), calculated backward iteratively for $i = 1$ to $n+1$, with $y_{n+2} = 1$.

And the hazard function is:

$$h_d(t) = \nu_0 \sum_{i=1}^{n+1} \frac{y_{it}}{\alpha_i} \frac{1 - e^{g_i (t_i - t_{i-1})}}{(y_i - A_i) - (y_i - B_i) e^{g_i (t_i - t_{i-1})}} + \nu_0 (1 - A_0) \left[1 - \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0) e^{g_0 (t - t_q)}} \right]$$

$$- r\nu_0 \sum_{i=1}^{n} d_i y_{it} \tag{20}$$

For the simple case of evenly spaced exposure times, i.e., $t_i - t_{i-1} = c$, and equal doses $d_i = d$ we have:

$$\log(S_d(t)) = \nu_0 \left[(B_0 - 1)t_1 + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} \right]$$

$$+ \nu_0 \left[(B_d - 1)(t_{n+1} - t_1) + \frac{1}{\alpha_d} \log \frac{B_d - A_d}{(y_n - A_d) - (y_n - B_d)e^{g_d (t_{n+1} - t_1)}} \right]$$

$$+ \nu_0 \left[(B_0 - 1)(t - t_{n+1}) + \frac{1}{\alpha_0} \log \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_{n+1})}} \right] + \nu_0 rd \sum_{i=1}^{n} (y_i - 1) \quad (21)$$

$$h_d(t) = \frac{\nu_0 y_{1t}}{\alpha_0} \frac{1 - e^{g_0 t_1}}{(y_1 - A_0) - (y_1 - B_0)e^{g_0 t_1}} + \frac{\nu_0}{\alpha_d} y_{nt} \frac{1 - e^{g_d (t_{n+1} - t_1)}}{(y_n - A_d) - (y_n - B_d)e^{g_d (t_{n+1} - t_1)}}$$

$$+ \nu_0 (1 - A_0) \left[1 - \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0 (t - t_{n+1})}} \right] + \nu_0 rd \sum_{i=1}^{n} y_{it} \quad (22)$$

with $B_d = B_i, A_d = A_i, \alpha_d = \alpha_i$ for i from 2 to n+1.

Further, for the multiple acute exposure cases where only the first mutation rate jumps *n* times but other parameters remain unaffected, we have for $t \geq t_n$,

$$S(t) = \exp\left\{ \log(S_0(t)) + \sum_{i=1}^{n} \left[\frac{B_0(1 - A_0) + A_0(B_0 - 1)e^{g_0 (t - t_i)}}{(1 - A_0) + (B_0 - 1)e^{g_0 (t - t_i)}} - 1 \right] \nu_0 rd_i \right\} \quad (23)$$

$$h(t) = h_0(t) + \sum_{i=1}^{n} \frac{(1 - A_0)(B_0 - 1)(B_0 - A_0) g_0 e^{g_0 (t - t_i)}}{\left((1 - A_0) + (B_0 - 1)e^{g_0 (t - t_i)} \right)^2} \nu_0 rd_i \quad (24)$$

4.2 Multiple exposures of continuous series

Similarly the formulae (15) and (16) can be generalized to multiple exposures of the continuous model.

Suppose that the *n* exposures start at at $T_1, T_2, \ldots, T_n$ and the durations of exposures are $l_1, l_2, \ldots, l_n$ respectively. No post exposure effect period is assumed for all exposures.

Let $t_0=0$, $t_1 = T_1$, $t_2 = T_1 + l_1$, …, $t_{2n-1} = T_n$, $t_{2n} = T_n + l_n$, , $t_{2n+1} = t$, and $t_i \geq t_{i-1}$ for i from *1* to *2n+1*. For $t \geq t_{2n}$, the survival function and hazard function are:

$$S_d(t) = \exp\left\{\sum_{i=1}^{2n+1} \nu_i \left[(B_i - 1)(t_i - t_{i-1}) + \frac{1}{\alpha_i} \log \frac{B_i - A_i}{(y_i - A_i) - (y_i - B_i)e^{g_i(t_i - t_{i-1})}}\right]\right\} \quad (25)$$

$$h_d(t) = \sum_{i=1}^{2n} \frac{\nu_i y_{it}}{\alpha_i} \frac{1 - e^{g_i(t_i - t_{i-1})}}{(y_i - A_i) - (y_i - B_i)e^{g_i(t_i - t_{i-1})}} + \nu_0(1 - A_0)\left[1 - \frac{B_0 - A_0}{(1 - A_0) - (1 - B_0)e^{g_0(t - t_q)}}\right] \quad (26)$$

The following table lists the numbers of the equations representing the survival and hazard functions for the models discussed in the text.

(Table 1)

For mixed exposure cases composed of exposure patterns listed above, the survival function and hazard function can be derived similarly.

5. Discussion

We have presented the analytical formulas for the hazard function and survival function for acute, continuous and multiple exposure cases. With these functions, the risks of cancer can be easily evaluated for any type of exposure within the framework of the piece-wise constant parameter model. With these functions in analytical form, some questions can also be addressed.

5.1 Some comments on constant parameter model

If we assume that the biological model parameters for the control group do not change, then the single interval model is appropriate for the control group. For simplicity, (6) and (7) may also be used to evaluate the effect of exposure. For some applications as we tried,

the constant parameter model is adequate. But for most cases, the result will be inappropriate. Actually this approach is theoretically problematic. If we apply the constant parameter model above to the exposed group and compare it with the control group using relative risk, then the relative risk of the exposed group before exposure will be above 1. Also the limit of the hazard function will be different from the one for a time dependent parameter model. Thus the time dependent parameter models are needed for the exposed groups, with the exception where the subjects' parents were exposed and the objective of the study is to evaluate the effect of the exposure on these subjects.

5.2 Long term effect of exposure

Researchers want to know how the exposure affects the hazard of cancer for the subject through creation and promotion of the initiated cells. Here we take the continuous exposure as an example. Compared with the *h(t)* of the control group, the last term in (18) differs only at the power of the exponential sub-term. If t_s, t_q are set to zero, that is, no exposure happens, then it is exactly the same as the one for the control group. When the subject is exposed, the effect of the exposure on the hazard will last for a considerable time, then, the risk will converge to the background risk. Actually the summation of the first two terms in (18) can be thought of as a record of the effect on the hazard as a function of time. Notice that this summation depends on time t through y_1, y_2, $y_{1t,}$, y_{2t} and y_{1t} and y_{2t} tend toward the limit zero. In the time course after the exposure takes place, this summation will approach zero. This dependence of the hazard on previous exposure well incorporates the fact that after the exposure stops, a considerable excess number of

intermediate cells beyond background will be gradually reduced, bringing the hazard to the background level. This confirms Heidenriech's analysis [9] (page 394).

5.3 Instant excess risk at exposure

For acute exposure, we are interested in the instant excess risk at the exposure. From (10) this quantity can be calculated by setting t at t_1:

$$h_d(t_1) - h_0(t_1) = \alpha_0(1 - A_0)(B_0 - 1)\nu_0 rd \tag{27}$$

With the appropriate assumptions and approximations, it reduces approximately to:

$$\nu_0 \mu_0 rd \tag{28}$$

This identity makes sense in that the instant excess risk at the exposure is proportional to the product of two mutation rates, and dose, but has little relation with the proliferation rate of the intermediate cells.

5.4 Age at exposure effect

Another issue is the age-at-exposure effect, that is, how will the age at exposure affect the risk of cancer in later life? It is not easy to give a mathematical analysis for all the cases, but for the pulse model we can calculate the point T_m where the excess hazard reaches the maximum after exposure at t_1. For this simple case, Heidenriech [10] gave the expressions of T_m as a function of exposure time t_1 and maximum excess hazard that the exposed may have. From (10) it is easy to derive these quantities in terms of the new set of notations.

$$T_m = t_1 + \frac{1}{g_0} \log \frac{1 - A_0}{B_0 - 1} \tag{29}$$

Assuming that $t_1' > t_1$, we want to compare the excess hazard at time t induced by exposure of dose d at t_1' with the one induced by the exposure of the same dose at t_1. The curve for the former can be obtained by shifting the curve for the former to the right by $t_1' - t_1$ (Figure 2). The two curves intersect at

$$t'_{\text{int}} = \frac{t_1 + t_1'}{2} + \frac{1}{g_0} \log \frac{1 - A_0}{B_0 - 1} \tag{30}$$

(Figure 2)

Thus the excess hazard induced by the latter exposure is less than the one induced by the former exposure when the attained age of the exposed is younger than t'_{int}, but it will catch up and become greater than the one induced by the latter exposure. For $t_1' > t_1 + \frac{2}{g_0} \log \frac{1 - A_0}{B_0 - 1}$, the two curves will not intersect with each other. Thus for $t > t_1'$, the excess hazard induced by the latter is always greater than the one induced by the former.

The excess hazard in (10) and (24) do not depend on exposure time(s), but only on the time interval between t_i*s* and *t*, implying no age at exposure in terms of excess hazard since the acute or multiple acute exposures. This observation suggests some limitation of the multiple-pulse model.

5.5 Fractionation effect

Researchers are interested in fractionation effect, that is, the effect of splitting of the dose over time. It is difficult to analyze the fractionation effect mathematically for most of the models given above. But if (24) can be applied for fractionated exposure, then the following is obtained:

Let t_1 denote the exposure time for the unfractionated exposure of dose d, t the current age of the subject. Assume that the exposure is fractionated evenly among 2 exposures of dose ½ d at t_1' and t_1'', with $t_1' < t_1''$. For easy comparison, the unfractionated exposure may be treated as two fractionated exposures of dose ½ d at t_1, From the discussion above the curves of excess hazard induced by the half dose exposure at t_1' and t_1'' intersect with the one induced by the half dose exposure at t_1 at $\frac{t_1 + t_1'}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}$, and $\frac{t_1 + t_1''}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}$, respectively. Thus for $t \in (t_1'', \frac{t_1 + t_1'}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1})$, the excess hazard induced by the unfractionated exposure is greater than the one induced by the fractionated exposure; for $t \in (\frac{t_1 + t_1'}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}, \frac{t_1 + t_1''}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1})$, this relationship still holds for sometime, then flips over, but the location of the change point depends on the locations of the fractionated exposures; for $t \in (\frac{t_1 + t_1''}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}, +\infty)$, the excess hazard induced by the fractionated exposure is greater than the one induced by the unfractionated exposure. Again, if the fractionated exposures happen after $t_1 + \frac{2}{g_0}\log\frac{1-A_0}{B_0-1}$, then there are no intersections, thus in this case the fractionated

exposure always induce higher excess risk than the unfractionated exposure for attained age *t*. Unevenly splitting the dose over the fractionated exposures does not change the locations of the intersections, but does change the location of the change point of relationship. Actually, if the fractionated exposures are not separated far, the distribution of the dose over the exposure does not make much difference (for the two exposures above, the width of the interval in the middle is only $\frac{t_1'' - t_1'}{2}$). For large number of fractionations or small number of fractionations with exposures separated considerably far, the fractionation attenuates the effect on the hazard between the first and last intersection points of the curves. That is, fractionation effect is always positive because of the additive functional form in (24). This observation implies some limitations of the multiple pulse model with linear dose response on first mutation rate, suggesting that multi-pulse model with some non-linear response, such as $f(d) = \sqrt{d}$, or the model specified in section 4.1 may be appropriate for some exposures since the inverse fractionation effect has been reported by Grahn, Lombard and Carnes [37].

5.6 Applications of the models

With the survival time t_i and cancer status ($\delta_i = 0$ for a subject without cancer, $\delta_i = 1$, for a subject with cancer) for each subject under study, the log-likelihood function can be calculated by summing over subjects:

$$l = \sum_i \delta_i \log h(t_i) + \log S(t_i) \tag{31}$$

where $h(t_i)$ and $S(t_i)$ are the hazard value and survival rate at t_i, calculated with the formulas depending on the model applicable. Maximum likelihood estimation can be

applied to obtain the parameter estimates of the model, which, in turn, may be used to calculate the survival rate and hazard rate at any age. For grouped data, Poisson likelihood estimation [38], which only requires hazard function to calculate the contribution for the subjects in each stratum, may be applied for estimating the parameters of the model.

With these models, we may have many applications. One analysis is the application of the continuous exposure model to the data of animals that inhaled fine particles of plutonium dioxide at a young age and were followed through their life time. The continuous exposure model allows us to incorporate the pharmacokinetics of the plutonium dioxide in specific organs over time. The estimates of the parameters obtained from MLE are then used to predict the risk of cancer as a function of dose and age. Also the results may be used to compare the relative carcinogenic effects of different radio isotopes. Another study is the analysis of the data of the mice that exposed to external gamma or neutron radiation with the total dose given at one time, or multiple times. In this study the acute and multiple acute exposure models can be applied to investigate the fractionation effect, and the relative biological effect of neutron compared with gamma radiation. Grahn, Lombard and Carnes [37] and Heidenriech, Carnes and Paretzke [39] conducted research on these issues for several cancer sites.

Conclusion

Mainly based on Moolgavkar's and Heidenriech's work on the two stage model, we extended the model for acute exposure, continuous exposure and multiple exposure cases with analytical formulas for the survival function and the hazard function. These

functions can be readily incorporated in likelihood function for the estimation of the model parameters, and thus cancer risk estimates as a function of dose and time can be given.

Acknowledgements

This research is supported by the funds from the US Department of Energy DE-FG02-08ER64605.

References

[1] S.H. Moolgavkar, Venzon, D,J., Two-Event models for carcinogenesis:Incidence curves for childhood and adult tumors. Math. Biosci. 47 (1979).
[2] S.H. Moolgavkar, Knudson, A. G., Jr., Mutation and cancer: a model for human carcinogenesis. Journal of the National Cancer Institute 66 (1981) 1037-52.
[3] S.H. Moolgavkar, Luebeck, G., Two-event model for carcinogenesis: biological, mathematical, and statistical considerations. Risk Analysis 10 (1990) 323-41.
[4] C. Nordling, A new theory on cancer-inducing mechanism. Br J Cancer 7 (1): 68-72. PMID 13051507 (1953).
[5] P. Armitage, Doll, R. , The age distribution of cancer and a multistage theory of carcinogenesis. Br. J. Cancer 8 (1954) 1-12.
[6] L.G. Hanin, Yakovlev, A.Y., A nonidentifiability aspect of the two-stage model of carcinogenesis. Risk Analysis 16 (1996) 711-5.
[7] W.D. Hazelton, Clements, Mark S., Moolgavkar, Suresh H., Multistage carcinogenesis and lung cancer mortality in three cohorts. Cancer Epidemiology, Biomarkers & Prevention 14 (2005) 1171-81.
[8] W.F. Heidenreich, On the parameters of the clonal expansion model. Radiation & Environmental Biophysics 35 (1996) 127-9.
[9] W.F. Heidenreich, Luebeck, E. G., Moolgavkar, S. H., Some properties of the hazard function of the two-mutation clonal expansion model. Risk Analysis 17 (1997) 391-9.
[10] W.F. Heidenreich, Jacob, P., Paretzke, H. G., Exact solutions of the clonal expansion model and their application to the incidence of solid tumors of atomic bomb survivors. Radiation & Environmental Biophysics 36 (1997) 45-58.
[11] W.F. Heidenreich, Brugmans, M. J., Little, M. P., Leenhouts, H. P., Paretzke, H. G., Morin, M., and Lafuma, J., Analysis of lung tumour risk in radon-exposed rats: an intercomparison of multi-step modelling. Radiation & Environmental Biophysics 39 (2000) 253-64.

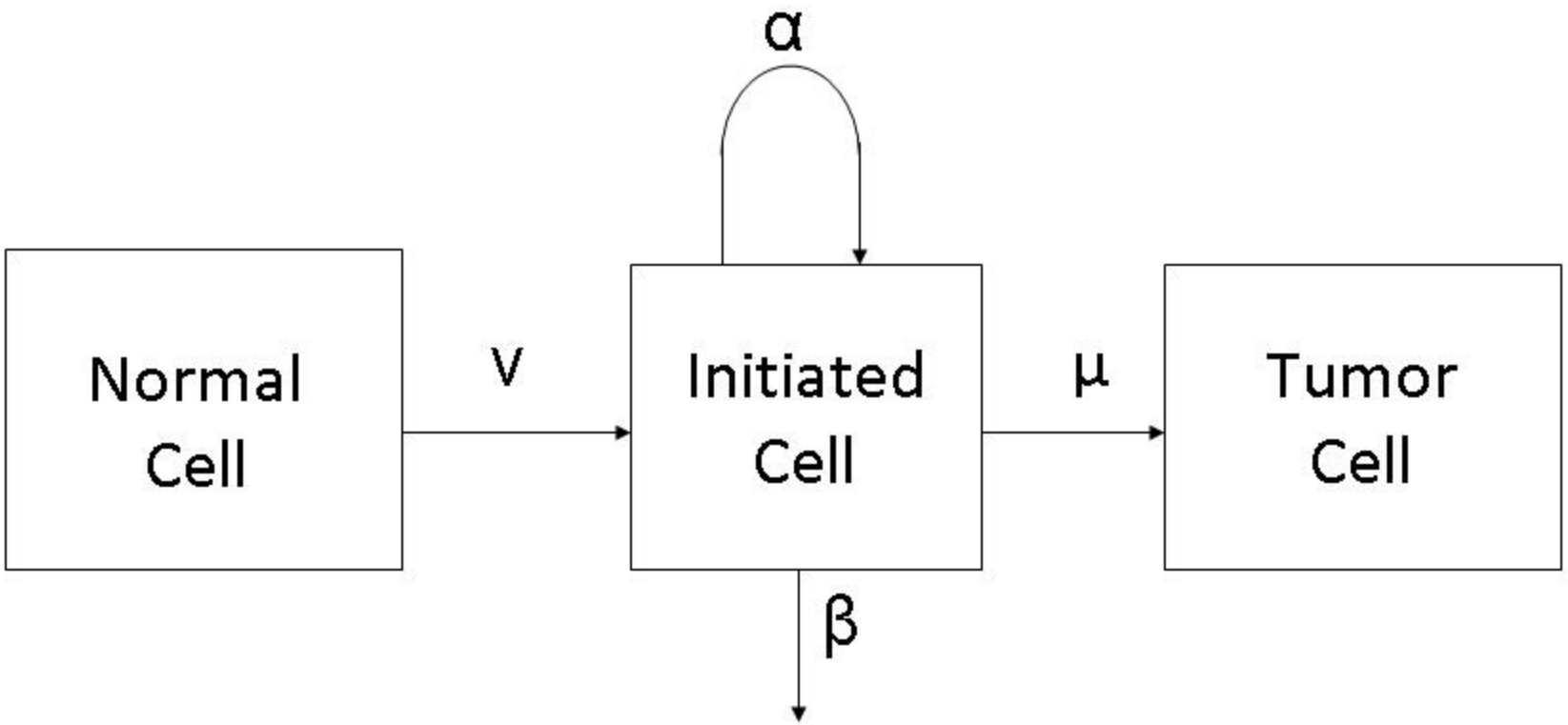
[12] W.F. Heidenreich, Luebeck, E. Georg, Hazelton, William D., Paretzke, Herwig G., and Moolgavkar, Suresh H., Multistage models and the incidence of cancer in the cohort of atomic bomb survivors.[see comment]. Radiation Research 158 (2002) 607-14.
[13] W.F. Heidenreich, Muller, W. A., Paretzke, H. G., and Rosemann, M., Bone cancer risk in mice exposed to 224Ra: protraction effects from promotion. Radiation & Environmental Biophysics 44 (2005) 61-7.
[14] H.P. Leenhouts, Radon-induced lung cancer in smokers and non-smokers: risk implications using a two-mutation carcinogenesis model. Radiation & Environmental Biophysics 38 (1999) 57-71.
[15] E.G. Luebeck, Moolgavkar, Suresh H., Multistage carcinogenesis and the incidence of colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America 99 (2002) 15095-100.
[16] S.H. Moolgavkar, Day, N. E., Stevens, R. G., Two-stage model for carcinogenesis: Epidemiology of breast cancer in females. Journal of the National Cancer Institute 65 (1980) 559-69.
[17] S.H. Moolgavkar, Dewanji, A., Venzon, D. J., A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor. Risk Analysis 8 (1988) 383-92.
[18] S.H. Moolgavkar, Biologically motivated two-stage model for cancer risk assessment. Toxicology Letters 43 (1988) 139-50.
[19] S.H. Moolgavkar, A two-stage carcinogenesis model for risk assessment. Cell Biology & Toxicology 5 (1989) 445-60.
[20] S.H. Moolgavkar, Cross, F. T., Luebeck, G., and Dagle, G. E., A two-mutation model for radon-induced lung tumors in rats. Radiation Research 121 (1990) 28-37.
[21] S.H. Moolgavkar, Cell proliferation and carcinogenesis models: general principles with illustrations from the rodent liver system. Environmental Health Perspectives 101 Suppl 5 (1993) 91-4.
[22] S.H. Moolgavkar, Biological models of carcinogenesis and quantitative cancer risk assessment. Risk Analysis 14 (1994) 879-82.
[23] S.H. Moolgavkar, Luebeck, E. G., Incorporating cell proliferation kinetics into models for cancer risk assessment. Toxicology 102 (1995) 141-7.
[24] S.H. Moolgavkar, Multistage models and the A-bomb survivor data: implications for carcinogenic mechanisms? Radiation Research 154 (2000) 728-9;discussion 730-1.
[25] S.H. Moolgavkar, Luebeck, E. Georg, Multistage carcinogenesis and the incidence of human cancer. Genes, Chromosomes & Cancer 38 (2003) 302-6.
[26] C.J. Portier, Sherman, C. D., Kohn, M., Edler, L., Kopp-Schneider, A., Maronpot, R. M., Lucier, G., Modeling the number and size of hepatic focal lesions following exposure to 2,3,7,8-TCDD. Toxicology & Applied Pharmacology 138 (1996) 20-30.
[27] Z. Qi, On the exact hazard and survival functions of the MVK stochastic carcinogenesis model. Risk Analysis 14 (1994) 1081-1084.

[28] C. Sherman, Portier, CJ., and Kopp-Schneider, A., Multistage models of carcinogenesis: an approximation for the size and number distribution of late-stage clones. Risk Analysis 14 (1994) 1039-48.
[29] C.D. Sherman, Portier, C. J., The two-stage model of carcinogenesis: overcoming the nonidentifiability dilemma. Risk Analysis 17 (1997) 367-74.
[30] M. 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).
[31] M. Little, Generalisations of the two-mutation and classical multi-stage models of carcinogenesis fitted to the Japanese atomic bomb survivor data. J Radiol Prot 16 (1996) 7-24.
[32] M.P. Little, Haylock, R. G. E., Muirhead, C. R. , Modelling lung tumour risk in radon-exposed uranium miners using generalizations of the two-mutation model of Moolgavkar, Venzon and Knudson. International journal of radiation biology 78 (2002) 49-68
[33] M.P. Little, Wright, E.G., A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data
Mathematical Biosciences 183 (2003) 111-134.
[34] A. Kopp-Schneider, Portier, CJ., and Sherman, CD, The exact formula for tumor incidence in the two-stage model. Risk Analysis 14 (1994) 1079-80.
[35] V. Jacob, Jacob, P. , Modelling of carcinogenesis and low-dose hypersensitivity: an application to lung cancer incidence among atomic bomb survivors. Radiat Environ Biophys. 42 (2004) 265-273.
[36] Y.T. Li, Wong, R. , Integral and series representations of the Dirac delta function. Communications On Pure And Applied Analysis 7 (2008) 229--247.
[37] D. Grahn, Lombard, LS, Carnes, BA, The comparative tumorigenic effects of fission neutrons and cobalt-60 gamma rays in the B6CF1 mouse. Radiation Research 129 (1992).
[38] M. Kai, Luebeck, E. G., Moolgavkar, S. H., Analysis of the incidence of solid cancer among atomic bomb survivors using a two-stage model of carcinogenesis.[see comment][erratum appears in Radiat Res 1998 Mar;149(3):309]. Radiation Research 148 (1997) 348-58.
[39] W. Heidenreich, Carnes, BA, Paretzke, HG., Lung cancer risk in mice: analysis of fractionation effects and neutron RBE with a biologically motivated model. Radiat Res. 166 (2006) 794-801.

Figure legends

Figure 1. Illustration of the two-stage clonal expansion model of carcinogenesis

Figure 2. The hazard curves predicted by the pulse model for an acute exposure at t_1 and the delayed acute exposure at t_1' and t_1''. Also $t_{\text{int}}' = \frac{t_1 + t_1'}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}$ $t_{\text{int}}'' = \frac{t_1 + t_1''}{2} + \frac{1}{g_0}\log\frac{1-A_0}{B_0-1}$ and $t_{R1} = t_1 + \frac{2}{g_0}\log\frac{1-A_0}{B_0-1}$.



ν----initiation rate
μ----transformation rate
α----division rate of initiated cells
β----death and differentiation rate

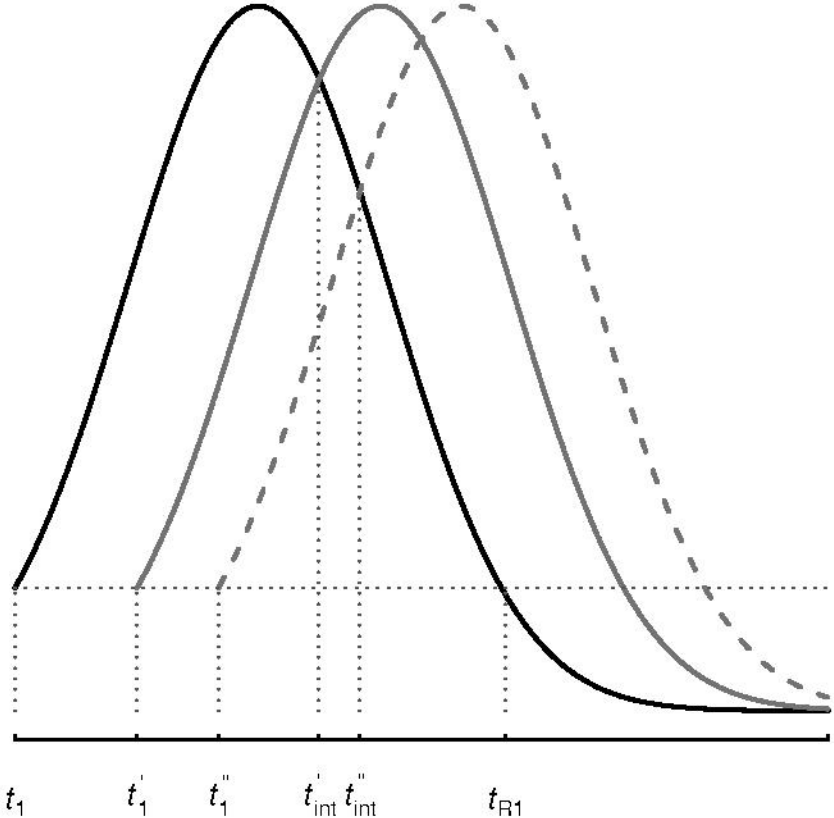

t_1
$t_1^{'}$
$t_1^{''}$
$t_{int}^{'}$
$t_{int}^{''}$
t_{R1}

Table 1 Equation numbers for the survival functions and hazard functions depending upon exposure pattern.

Exposure Pattern	S(t)	h(t)
Control group	(6)	(7)
Simple acute	(8)	(10)
Acute	(11)	(12)
Continuous	(17)	(18)
Multiple acute	(19)	(20)
Multiple simple acute	(23)	(24)
Multiple continuous	(25)	(26)