# NOTICE

## CERTAIN DATA CONTAINED IN THIS DOCUMENT MAY BE DIFFICULT TO READ IN MICROFICHE PRODUCTS.

DOE/ER/60657--T1

DOE/ER/60657--T1

DE93 010960

# PROGRESS REPORT

P.I.: Suresh H. Moolgavkar

**PROGRESS REPORT:** During the past several years, we have made considerable progress in modelling carcinogenesis in general, and in modelling radiation carcinogenesis, in particular. This work has led to the publication of numerous papers, which are listed in the biosketches of the investigators attached to this application. Rather than discussing each of these publications briefly, we present an overview of our progress in developing stochastic carcinogenesis models and applying them to experimental and epidemiologic data sets.

Traditionally, cancer models have been used for the analysis of incidence (or prevalence) data in epidemiology and time to tumor data in experimental studies. The relevant quantities for the analyses of these data are the hazard function and the probability of tumor. The derivation of these quantities is briefly described here. More recently, we began to use these models for the analysis of data on intermediate lesions on the pathway to cancer. Such data are available in experimental carcinogenesis studies, in particular in initiation and promotion studies on the mouse skin and the rat liver. Typically, such data take the form of information on the number and size distribution of intermediate lesions as functions of the dose of the chemical agents applied and time on study. Because such information is not available for radon-induced lung tumors, the relevant mathematical development is not described here. If, however, quantitative information on intermediate lesions on the pathway to lung cancer were to become available at some future date, the methods that we have developed for the analysis of initiation-promotion experiments could easily be applied to the analysis of these lesions. The mathematical derivations here are couched in terms of a particular two-mutation model of carcinogenesis. Extension to models postulating more than two mutations is not always straightforward.

**THE TWO-MUTATION MODEL** : The version of the two-mutation model discussed here has been widely used for the analysis of data. Similar models were considered by Neyman and Scott (1967) and Kendall (1960). A detailed description of the model can be found in a recent paper by Moolgavkar and Luebeck (1990), which is part of the appendix. The development here follows the development in that paper and uses the same notation. The fundamental biological assumptions are: (1) In any tissue there is a pool of cells susceptible to malignant transformation. This pool is generally identified with the stem cell pool in the tissue of interest and may change in size during life; (2) malignant tumors are clonal, i.e., they arise from a single progenitor cell that has become malignantly transformed;

(3) malignant transformation of a susceptible cell is the result of two specific, rate-limiting, hereditary (at the level of the cell) and irreversible events. For a discussion of the biological interpretation of the two events, see the above referenced paper by Moolgavkar and Luebeck (1990), and the paper by Moolgavkar and Knudson (1981). The model also provides a natural framework for interpretation of initiation and promotion, as is discussed in the two papers referenced above (see figure 1).

The following assumptions are used in the mathematical development. Let $X(s)$ represent the number of normal susceptible cells in the tissue of interest at time (age) $s$, and suppose that intermediate cells, i.e., cells that have sustained the first rate-limiting event on the pathway to malignancy, arise from normal cells as a non-homogeneous Poisson process with intensity $\nu(s)X(s)$, where $\nu(s)$ is the first event rate. Note that, although $\nu$ and $X$ are not separately identifiable, we prefer to model the two separately because information on one or the other may be available from independent sources. In time interval $(s, s+\Delta s)$, an intermediate cell divides into two intermediate cells with probability $\alpha(s)\Delta s + o(\Delta s)$; it dies or differentiates with probability $\beta(s)\Delta(s) + o(\Delta s)$ (note that death and differentiation are equivalent events for carcinogenesis because both events remove the cell from the pool of susceptible cells); it divides into one intermediate cell and one cell that has sustained the second event (malignant cell) with probability $\mu(s)\Delta s + o(\Delta s)$; the probability of more than one event is $o(\Delta s)$. In many applications, the parameters are assumed to be constant or piecewise constant. In particular this implies that the waiting times to cell division and cell death are assumed to be exponential.

Some comments on these mathematical assumptions are in order. The cell kinetics of intermediate cells are modelled in primitive fashion. There are entire tomes on the mathematical modelling of the cell cycle and it is clear that cells do not divide or die with exponential waiting times (see, e.g., Jagers, 1983). Nevertheless, in the context of carcinogenesis modelling these simplifications appear to be entirely appropriate as a first approximation. However, more realistic cell cycle models need to be investigated. Once a malignant cell is generated it is assumed to give rise to a detectable tumor after a suitable lag time. This assumption is clearly false and there is clearly a time to detection distribution. Further, malignant cells undoubtedly execute a birth-death process and as a consequence become extinct with non-zero probability. Incorporation of such considerations into the model would hope-

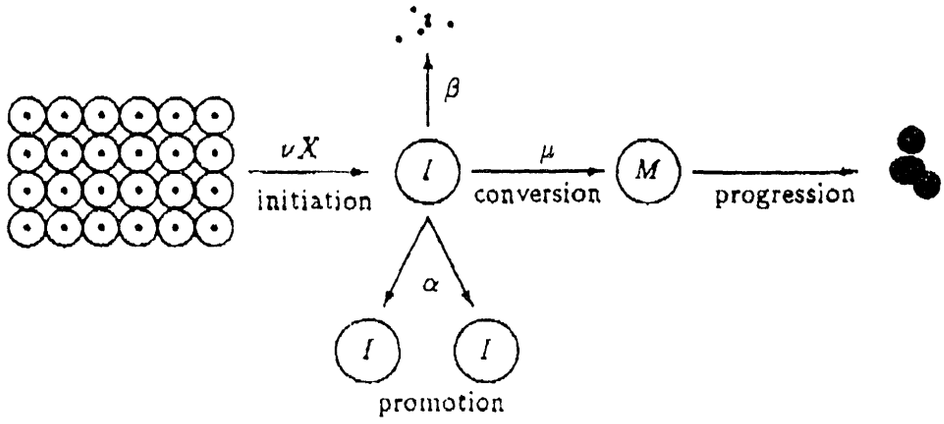


Figure 1: Pictorial representation of the two-mutation model. In the parlance of chemical carcinogenesis, the first rate-limiting event can be identified with initiation, the second rate-limiting event with malignant conversion, and the clonal expansion of intermediate cells with promotion. Thus, a promoter increases the net proliferation rate, $\alpha - \beta$, of initiated cells, either by increasing $\alpha$, the cell division rate, or decreasing $\beta$, the cell death (or differentiation) rate, or both. Increased cell division rates and decreased cell death rates have both been implicated in promotion (Schulte-Hermann et al, 1989). After malignant transformation, relatively rapid changes lead to tumor progression. These are not explicitly modelled.

lessly overparametrize it; however, if information on the growth kinetics of malignant cells is available from independent sources, the model could quite easily be extended to incorporate it. We have begun the mathematical development necessary for this (Dewanji, Moolgavkar and Luebeck, 1991).

*THE HAZARD FUNCTION* (Moolgavkar and Luebeck, 1990):

Let $Y(t), Z(t)$, represent the number of intermediate and malignant cells, respectively, at time $t$ and let

$$\Psi(y,z;t) = \sum_{j,k} P_{j,k}(t) y^j z^k$$

be the probability generating function with

$$P_{j,k}(t) = Prob\,[Y(t) = j, Z(t) = k \mid Y(0) = 0, Z(0) = 0]$$

Then $(Y(t), Z(t))$ is Markovian, and $\Psi$ satisfies the Kolmogorov forward differential equation

$$\Psi'(y,z;t) = \frac{\partial \Psi(y,z;t)}{\partial t} = (y-1)\nu(t)X(t)\Psi(y,z;t)$$
$$+ \{\mu(t)yz + \alpha(t)y^2 + \beta(t) - [\alpha(t) + \beta(t) + \mu(t)]y\}\frac{\partial \Psi}{\partial y} \qquad (1)$$

with initial condition $\Psi(y,z;0) = 1$ (Moolgavkar, Dewanji and Venzon, 1988). $\Psi(1,0;t)$ is the survival function for this model, and the hazard (incidence) function is given by

$$h(t) = -\Psi'(1,0;t)/\Psi(1,0;t) \qquad (2)$$

It follows immediately from the Kolmogorov equation that

$$\Psi'(1,0;t) = -\mu(t)\frac{\partial \Psi}{\partial y}(1,0;t)$$

and thus

$$h(t) = \mu(t)E[Y(t) \mid Z(t) = 0] \qquad (3)$$

where E denotes the expectation and we have used the relationship

$$E[Y(t) \mid Z(t) = 0] = \frac{\partial \Psi}{\partial y}(1,0;t)/\Psi(1,0;t)$$

If the probability of tumor is small enough then $E[Y(t)] \approx E[Y(t) \mid Z(t) = 0]$ and $h(t) \approx \mu(t)E[Y(t)]$. The differential equation, derived from the Kolmogorov equation, for $E[Y(t)]$ can be readily solved to yield

$$h(t) \approx \mu(t) \int_0^t \{\nu(s)X(s) \exp \int_s^t [\alpha(u) - \beta(u)]du\}ds \qquad (4)$$

This is the approximate solution that has been used for the analysis of epidemiologic data. It is reasonably accurate when tumors are rare as in epidemiologic data. However, our recent experience is that it can lead to misleading results even for epidemiologic data. Thus, the exact solution must be used whenever possible, and this is quite a bit more complicated. Note that the approximate solution is of the proportional hazards form if only the second mutation rate is a function of the covariates of interest.

Two approaches can be used to obtain the exact solution to the two-mutation model. The first approach involves solving the characteristic equations (see e.g., John, 1971, pp. 6-15) associated with the Kolmogorov equation. The second approach is somewhat more general, and is not described here, but can be found in Moolgavkar and Luebeck (1990). Specifically, the characteristic equations are

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

$$\frac{dz}{du} = 0 \quad (z \text{ is constant along characteristics})$$

$$\frac{dt}{du} = 1, \text{and} \frac{d\Psi}{du} = (y-1)\nu(u)X(u)\Psi$$

Now, the ordinary differential equation for $\Psi$ may be solved along characteristics to yield

$$\Psi(y(t),z,t) = \Psi_0 \exp \int_0^t [y(u,t) - 1]\nu(u)X(u)du \tag{5}$$

where $\Psi_0 = \Psi(y(0),z,0) = 1$ is the initial value of $\Psi$. We are interested in computing $\Psi(1,0;t)$ for any $t$, and thus we need to find the values of $\Psi$ along the characteristic through $(y(0),0,0)$ where $y(0)$ is the initial value of $y$ and $y(t) = 1$. Now, along the characteristic, $Y$ satisfies the differential equation $dy/du = -R(y,u)$ and this is just a Riccati equation which can be readily integrated in closed form if the parameters of the model are piecewise constant (see e.g., Ince, 1956, p.311). To be precise, the Riccati equation for $y$ can be solved to yield a value for $y(u)$ for any $u$, with initial condition $y(t) = 1$. Note that $y$ depends on $u$ and $t$.

Thus, the survival function

$$\Psi(1,0;t) = \exp \int_0^t [y(u,t) - 1]\nu(u)X(u)du \tag{6}$$

where the explicit dependence of $y$ on $u$ and $t$ is acknowledged. The hazard function then is given by

$$h(t) = -\Psi'(1,0;t)/\Psi(1,0;t) = -\int_0^t \nu(u)X(u)y_t(u,t)du \tag{7}$$

where $y_t$ denotes the derivative of $y$ with respect to $t$.

Suppose now that $0 = t_0 < t_1 < \cdots < t_k = t$, and suppose that the parameters $\alpha$, $\beta$, and $\mu$, are piecewise constant, i.e., on $(t_{i-1}, t_i)$ the parameters are $\alpha_i$, $\beta_i$ and $\mu_i$. Suppose further that $A_i$ and $B_i$ are the two roots of the polynomial $\alpha_i x^2 - [\alpha_i + \beta_i + \mu_i]x + \beta_i$. It can be easily shown that $0 < A_i < 1 < B_i$. A closed form expression for $y(u,t)$ can be obtained by solving the Riccati equation for $y$ successively on each subinterval starting with $(t_{k-1}, t)$ with initial condition $y(t_k, t_k) = 1$. The initial condition on the interval $(t_{k-2}, t_{k-1})$ is then the solution obtained on the interval $(t_{k-1}, t_k)$ evaluated at $t_{k-1}$, i.e., $y(t_{k-1}, t)$. Thus the solution $y(u,t)$ can be inductively built up. Explicitly, for $u \in (t_{i-1}, t_i)$, $y(u,t)$ can be defined inductively by

$$y(u,t) = \frac{B_i - A_i \frac{y(t_i,t)-B_i}{y(t_i,t)-A_i} \exp[\alpha_i(A_i - B_i)(u - t_i)]}{1 - \frac{y(t_i,t)-B_i}{y(t_i,t)-A_i} \exp[\alpha_i(A_i - B_i)(u - t_i)]} \tag{8}$$

with $y(t_k, t) = 1$. This is a generalization of the result in Moolgavkar and Venzon (1979).

The derivative $y_t(u,t)$ is now straightforward, albeit cumbersome, to compute using the chain rule repeatedly. The equations for $\Psi(1,0;t)$ and $h(t)$ can be numerically integrated using the values of $y(u,t)$ and $y_t(u,t)$ computed above. If $\nu(u)$ is piecewise constant too and if, as is often the case, $X(u)$ is taken to be constant, then, in principle, these equations can be integrated in closed form. However, a numerical procedure is simpler.

Sometimes, the time-scale of interest is not the age of the animal, or time since start of treatment, but the age of individual intermediate clones. Then, $(Y(t), Z(t))$ is not Markovian, and the Kolmogorov differential equation does not exist. The second approach, described in Moolgavkar and Luebeck (1990), must then be used.

**Applications to the analysis of data**: The hazard function derived from the two-mutation model has been used for the analysis of both experimental and epidemiologic data (for a review see Moolgavkar and Luebeck, 1990). When the approximate solution (expression 4) is used, the hazard function is of the proportional hazards form if only the second mutation rate is a function of the covariates of interest. We believe now that the exact solution should always be used, and we briefly describe the analyses of two data sets, one experimental the other epidemiologic, of particular relevance to this grant application.

Radon and lung cancer in rats: The data included in this analysis were from rat experiments conducted under carefully controlled conditions by Dr. Fred Cross at the Battelle

Pacific Northwest Laboratories at Richland. The experiments were conducted under radon-daughter exposure conditions that resulted in a dose at the cellular level of approximately 5 mGy per working level month (WLM) of exposure. When animals were found dead or were sacrificed, the lungs (and selected other organs) were removed and fixed in 10 per cent neutral buffered formalin for subsequent histopathologic examination. Data from 1797 animals exposed to radon daughters over the approximate range 320 to 10,240 WLM (1.1 to 36 $Jhm^{-3}$) were included in the analysis. The following information was available on each animal in the data set: the exact age when exposure to radon was begun, the radon-daughter exposure rate in WLM/week (WLM/w), the age at which exposure was stopped, age at death or sacrifice, and presence or absence of malignant lung tumor. All animals were followed until sacrifice or death.

The objectives of the analysis were to estimate the mutation rates and intermediate cell proliferation parameters as functions of the exposure rates of radon. This was achieved by maximizing the likelihood of the data. Let $P(t)$ be the probability of tumor by age $t$ for some particular exposure-rate regimen. Then, the survivor function $S(t) = 1 - P(t)$ and the hazard function $h(t) = -S'(t)/S(t)$. The expressions for these quantities derived from the two-mutation model are discussed above. In the opinion of the pathologist, the lung tumors were incidental, i.e., they did not cause death of the animal. Thus, the likelihood of the data was constructed as follows. Because the tumors were incidental, the contribution to the likelihood by an animal that died (or was sacrificed) at age $t$ is $P(t)$ if it had a tumor, or $S(t)$ if it was free of tumor. The full likelihood is the product of these terms over all the animals. The exact expressions and other details can be found in the relevant publication (Moolgavkar et al, 1990), which is part of the appendix to this application. The likelihood was maximized using a software program that employs three subroutines sequentially. The first suroutine is a Monte Carlo search , which helps to guard against local maxima. This is followed by a simplex algorithm and finally by a gradient algorithm.

Based on a comparison of observed and expected numbers of tumors in various exposure rate categories, we concluded that the model described the data well. The analysis indicated that radon increases the first mutation rate and the net proliferation of intermediate cells, but has little effect on the second mutation rate, suggesting that the nature of the two mutational events is different. The analysis also confirmed an inverse exposure-rate effect. Further, the

inverse exposure-rate effect could be attributed to the effect of radon on intermediate cell kinetics, i.e., on the promotional effect of radon. Why should this be so? Because with agents that increase the net proliferation rate of intermediate cells, duration of exposure is more important than exposure rate in dtermining risk. This is a mathematical consequence of the model. Why should exposure to radon increase the net proliferation rate of intermediate cells? One possible explanation is that radon causes a subtle shift away from differentiation. Thus, it has been suggested that, in epithelial tissue that is turning over, stem cells undergo asymmetric division, i.e., one of the daughter cells remains a stem cell, whereas the other daughter is committed to differentiation. Radon could act as a promoter if it caused a shift away from asymmetric division towards symmetric division, i.e., stem cell division in which both daughters are stem cells. This hypothesis, generated by analysis of these data, should be testable in the laboratory.

Reanalysis of the Colorado Plateau uranium miners' data: Much of our knowledge regarding the interaction of radon and tobacco smoke in the etiology of human lung cancer derives from studies of uranium miners. We recently undertook a reanalysis of the lung cancer mortality in the Colorado Plateau uranium miners' cohort within the framework of the two-mutation model described here. The analysis takes explicit account of the patterns of exposure to both radon and cigarette smoke experienced by individuals in the cohort. A simultaneous reanalysis of the British doctors' cohort indicated that those model parameters relating to the effect of tobacco were similar in the two data sets. No evidence of interaction between radon and cigarette smoke was found with respect to their joint effect on the first or second mutation rates, or on the proliferation of intermediate cells. However, the age-specific relative risks associated with joint exposure to radon and cigarette smoke were supra-additive but sub- multiplicative. The analysis also confirmed that fractionation of radon exposure leads to higher lung cancer risks. As in the analysis of the experimental data described above, the parameters of the model were estimated by maximizing the likelihood. However, in contrast to the rat lung malignancies, which were incidental, human lung cancers are rapidly fatal. Thus, individuals who develop lung cancer contribute the probability density function for the time to tumor to the likelihood function. Individuals who do not develop lung cancer contribute the survivor function, as in the case of the experimental data. As judged by a comparison of observed and expected numbers of lung cancers in various

categories, the model described the data well. In addition a comparison of theoretical and empirical Kaplan-Meier plots indicated that the model described the temporal pattern of failures well. Further details can be found in the preprint by Moolgavkar et al, which is part of the appendix to this application. It is interesting to note that the analyses of experimental and epidemiologic data yielded consistent results despite the different likelihoods maximized. Thus, as in the case of the experimental data, analysis of the epidemiologic data indicated that radon strongly affected the first mutation rate and the proliferation rate of intermediate cells. It was to the latter effect of radon that the inverse exposure-rate effect could be attributed. In both data sets, the second mutation rate was little affected.

Finally, we list here the publications resulting from current DOE grant support.

1. Moolgavkar SH, Cross FT, Luebeck G, Dagle GE: A two-mutation model for radon-induced lung tumors in rats. Radiation Research 121:28-37, 1990.

2. Moolgavkar SH, Luebeck G: Two-event model for carcinogenesis: biological, mathematical and statistical considerations. Risk Analysis 10:323-341, 1990.

3. Moolgavkar SH, Luebeck G, DeGunst M: Two mutation model for carcinogenesis: Relative roles of somatic mutations and cell proliferation in determining risk. In Scientific Issues in Quantitative Cancer Risk Assessment, SH Moolgavkar (ed.), Birkhauser Boston, 1990, 136-152.

4. Moolgavkar SH: Cancer Models, invited editorial. Epidemiology 1:419-420, 1990.

5. Moolgavkar SH: Cell proliferation in carcinogenesis (letter). Science 251:143, 1991.

6. Moolgavkar SH: Stochastic models of carcinogenesis in Handbook of Statistics Vol 8 CR Rao and R. Chakraborty (eds.) Elsevier, 1991, 373 - 393.

7. Moolgavkar SH. Carcinogenesis models: An overview. To appear in the Proceedings of the 29th Hanford Symposium on Health and the Environment. FT Cross, (ed.), 1991.

8. Luebeck EG, Moolgavkar SH: Stochastic description of initiation and promotion in experimental carcinogenesis. Annali dell'Istituto Superiore di Sanita 27: 575 - 580, 1991.

9. Moolgavkar SH, Luebeck EG: Multistage carcinogenesis: A population-based model for colon cancer. JNCI, 84: 610 - 618, 1992.

10. Moolgavkar SH: Cancer models. In "Biophysical Modelling of Radiation Effects", K Chadwick, G Moschini, M Varma (eds.), Adam Hilger, Bristol 1992, 239 - 252.

11. Moolgavkar SH, Luebeck EG, Krewski D, Zielinski JM: Radon, cigarette smoke and lung cancer: A reanalysis of the Colorado Plateau miners' data. Epidemiology, to appear.

12. Moolgavkar SH, Luebeck EG: A two-mutation model for radiation carcinogenesis in humans and rodents. To appear in the Proceedings of the Second International Conference on Theories of Carcinogenesis, Oslo, August 1992.

# END

DATE
FILMED
5/04/93