NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT
BILTHOVEN, THE NETHERLANDS

Report no. 610065006

A Two-Mutation Model of Carcinogenesis: Application to Lung Tumours Using Rat Experimental Data

M.D. Harris, H.P. Leenhouts and P.A.M. Uijt de Haag

January 1998

This report was commissioned by the National Institute of Public Health and the Environment (project no. 610065).

National Institute of Public Health and the Environment (RIVM), P.O. Box 1,
3720 BA Bilthoven, the Netherlands
Telephone: +31 - 30 - 274 25 15, telefax: +31 - 30 - 274 44 28

MAILING LIST

1	Directie RIVM
2	Directeur Sector Stoffen en Risico's
3	CEC, DGXII, Head of Radiation Protection Research Action Unit
4 - 8	Contract Partners ...
9	Hoofd van het Laboratorium voor Stralingsonderzoek
10	Hoofd van de afdeling Modellen en Processen van het Laboratorium voor Stralingsonderzoek
11	Bibliotheek van het Laboratorium voor Stralingsonderzoek
12 - 14	Auteurs
15	K.H. Chadwick
16	W. Slob
17	Hoofd afdeling Voorlichting & Public Relations
18	Bureau Rapportenregistratie
19	Bibliotheek RIVM
20	Depot van Nederlandse Publicaties en Nederlandse Bibliografie
21 - 22	Reserve exemplaren LSO
23 - 43	Bureau Rapportenbeheer

CONTENTS

SUMMARY

In this report a two stage carcinogenesis model is used to describe the development of lung tumours in rats exposed to radon and uranium ore dust. A tumour is taken to develop in two mutations steps, one from a normal cell to an intermediate cell, and one from an intermediate cell to a tumourous cell. The rates of these steps are influenced by radon and uranium dust exposure levels. Furthermore, the intermediate cell population experiences a net growth whose rate is also dependent on the radon and uranium concentrations. The relationship between mutation rate and the concentrations is investigated, and several biologically realistic possibilities are explored. The difficulties of incorporating the link between tumour incidence and death of the animal are considered. Tumours can be considered as incidental, fatal, or sometimes incidental and sometimes fatal. The effect of using these different interpretations is explored. Finally, a comparison between a much used approximation to this model, and the exact solution is made.

For the tumour development in the rats examined, it was found that the normal-intermediate mutation rate is dose rate dependent, and that the intermediate-malignant mutation rate is dose rate independent. Further, it was found that the intermediate cell growth rate increases with increasing radon concentrations. Finally, it was found that different sets of parameters values are able to describe the data equally well, but give different predictions of tumour incidence for radiation exposure patterns different to those found in the data.

SAMENVATTING

In dit rapport wordt een twee-staps carcinogenese model beschouwd om de ontwikkeling van longtumoren veroorzaakt door radon en uraniumertsstof te beschrijven. Een tumor wordt verondersteld via twee mutaties te ontstaan, één van een normale cel naar een intermediaire cel, en één van een intermediaire cel naar een kwaadaardige cel. De snelheid van de stappen is afhankelijk van de mate van blootstelling aan radon en uraniumstof. Ook wordt er verondersteld dat een intermediaire cel zichzelf vermenigvuldigt naar twee intermediaire cellen met een snelheid die afhankelijk is van de radon en uranium concentraties. Een aantal verschillende biologisch realistische mogelijkheden van deze afhankelijkheden wordt beschouwd. Ook wordt de link tussen de aanwezigheid van tumoren en het overlijden van een dier onderzocht. Tumoren kunnen als incidenteel, als fataal, of soms als incidenteel en soms als fataal worden behandeld. De invloed van deze mogelijke keuzes is hier onderzocht. Tenslotte wordt er een vergelijking gemaakt tussen een veelgebruikte benadering en de exacte oplossing van dit model.
Voor de tumorontwikkeling in de onderzochte ratten blijkt dat de snelheid van de intermediaire-malignant mutatie onafhankelijk van de radonconcentratie is. Ook blijkt dat de groei van de intermediaire cellen groter wordt bij grotere radonconcentraties. Tenslotte is gevonden dat verschillende waarden van de modelparameters de data even goed kunnen beschrijven, maar verschillende risicoschattingen geven in andere dan de onderzochte stralingsomstandigheden.

1 INTRODUCTION

The estimation of human cancer risks at low and chronic radiation exposures is, due to lack of better data, done on the basis of high exposure data. It is, a priori, unclear how one should include factors such as the exposure rate, age at exposure, and total exposure dosage in the risk analysis. Also, it is not clear how one should use the data to give risk estimates for low exposures. Sophisticated statistical analyses have been carried out in attempts to describe the epidemiological data, and to make risk estimates from the fits obtained. (See, for example Gilbert et al.[1] where such an analysis is done using the data we use in this paper.) In this paper we use a model based approach. The tumour incidence can be described in terms of parameters with natural biological interpretations. In such a scheme, the effects of different exposure patterns can be introduced in a natural way. The model describes the time and dose dependence of tumour incidence - something that otherwise must be found by statistical means. Finally, the model gives predictions which can be experimentally tested.

The model we use is a two-mutation carcinogenesis model which has previously been used to describe various tumour data, including the rat data used in this paper.[2,3]
The theoretical predictions of the model have been thoroughly investigated, as done, for example in Moolgavkar et al.[4] and in Moolgavkar and Lübeck[5]. The model describes qualitative aspects of tumour development well - for example, the age dependent radiation sensitivity, and the age dependence of tumour incidence. The actual model parameter values obtained depend on the method used to fit the data, and the interpretation of the data.

In this analysis, we use data obtained from an experiment on rats, conducted at the Pacific Northwest Laboratory (PNL). The experimental procedure is documented in Cross[6]. The study involved Wistar rats that were exposed to varying radon dosages (between 20 and 10000 WLM), administered at different dosage rates (from 0.7 to 80 WLM/d), starting at around 90 days of age. Rats exposed to radon were also exposed to uranium ore dust, of which the concentrations are also known. For each rat, we know the administered radon and ore dust dosage pattern, the age at death, if there was a tumour present at death, and whether this was judged as being the cause of death. Rats, contrary to humans, can live for an extended period with a lung tumour. This raises the question as to how to handle the different sorts (or possibly stages) of lung tumours (incidental and fatal).

The purpose of this report is four-fold.

1. Firstly, it gives a mathematical description of the model, and in doing so, raises several points which are important to understand on a theoretical level.
2. It clarifies mathematically the approximation of this model made in many papers, and discusses the validity of this approximation using parameter values found in this paper.
3. It investigates several different definitions of the "tumour incidence" and looks at the influence this has on the chosen optimal model and parameters.

4. It explores several different parameterisations of the model to get a feel for which aspects are important in risk determination.
This paper is meant to be a summarising link between the literature and the work carried out at RIVM.

The paper is arranged as follows. In Section 2 the Two-Mutation-Model is described mathematically, and expressions are obtained for the quantities of interest. All cumbersome calculations are relegated to the appendix. In Section 3 different dependencies of the mutation and growth rates on radon exposure levels are given, together with their biological interpretations. Section 4 is devoted to describing the ways that the model is fitted to the data. In doing this, it is discussed how one should handle information concerning the fatality of tumours found at the death of the animal. Once one has found the model parameters that optimally describe the data, one would like to have an absolute measure as to how good this optimal fit actually is. This is discussed in Section 5. Section 6, titled 'Results' is divided into several subsections. Subsections 6.1 to 6.4 are devoted to analysing the optimal fits obtained. Subsection 6.5 discusses the problem of the uniqueness of the fits, and the possible effects this could have on risk estimates made for low radon exposure levels. Subsection 6.6 compares tumour incidence predictions using an approximation made in many papers (mentioned above) and those using the exact solution. The final section of the paper is titled 'Conclusions' and summarises the results of the paper.

2 THE TWO MUTATION MODEL

We consider a model where a normal cell can become a tumourous cell via an intermediate stage. The transformation thus occurs in two stages, and we call the model a two-stage carcinogenesis model.

We consider a pool of normal susceptible cells, called *stem* cells. We assume that at time t there are $S(t)$ such cells, where S is a deterministic function of t. The stem cells generate *intermediate* cells according to a non-homogeneous Poisson process with rate $\mu_1(t)S(t)$. An intermediate cell may die (with rate $\nu(t)$), may divide and become two intermediate cells (with rate $\varepsilon(t)$), or may turn into a *malignant* cell and an intermediate cell (with rate $\mu_2(t)$). A malignant cell is not directly detectable - but must first multiply itself into a detectable tumour. We assume that a malignant cell becomes a detectable tumour after a certain time t_{lag}. In this paper we take t_{lag} to be a deterministic value. This means then that if we are interested in the presence of a tumour at time $t+t_{lag}$, we must look at the probability of a malignant cell at time t. This is represented pictorially in Figure 1.

The state of the system is then described by the random variables $(I(t),M(t))$: $I(t)$ is the number of intermediate cells at time t, and $M(t)$ is the number of malignant cells at time t. As we can determine the number of tumours at time $t+t_{lag}$ by determining the number of malignant cells at time t, we focus our attention on the distribution of the number of malignant cells $M(t)$.

We denote by $P_{i,m}(t)$ the probability that there are i intermediate and m malignant cells at time t. That is

$$P_{i,m}(t) = P_{i,m} := P(I(t) = i, M(t) = m).$$

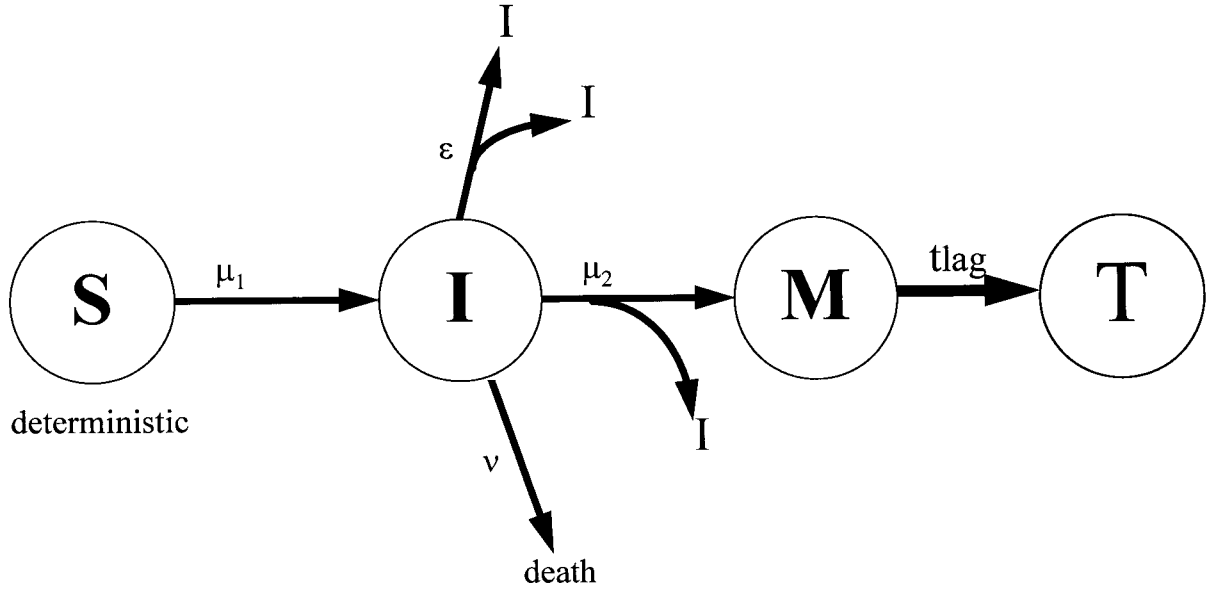

Figure 1 *Different stages in the development of a tumour from a stem cell.*

The time dependence of these probabilities is governed by the Kolmogorov forward equation:

$$\begin{aligned}\frac{\partial P_{i,m}}{\partial t} &= -i(\varepsilon(t)+\nu(t)+\mu_2(t))P_{i,m} \\ &\quad - S(t)\mu_1(t)P_{i,m} \\ &\quad + (i-1)\varepsilon P_{i-1,m} \\ &\quad + S(t)\mu_1(t)P_{i-1,m} \\ &\quad + (i+1)\nu P_{i+1,m} \\ &\quad + i\mu_2 P_{i,m-1}.\end{aligned} \tag{1}$$

This is a balance equation - the first two terms $-i(\varepsilon(t)+\nu(t)+\mu_2(t))P_{i,m} - S(t)\mu_1(t)P_{i,m}$ are the rates at which the state (i,m) is left, and the other terms are the rates at which the state (i,m) is entered. Henceforth we will sometimes neglect to explicitly write dependencies on t. The probability that there are i intermediate cells at time t, $P_i(t)$, is clearly $P_i(t) = P_i := P(I(t)=i) = \sum_m P_{i,m}(t)$ and the probability that there are m malignant cells at time t, $P^m(t)$, is

$P^m(t) = P^m = P(M(t)=m) = \sum_i P_{i,m}(t).$

Then, from (1) we see

$$\frac{\partial P_i}{\partial t} = -i(\varepsilon+\nu+\mu_2)P_i - S\mu_1 P_i + (i-1)\varepsilon P_{i-1} + S\mu_1 P_{i-1} + (i+1)\nu P_{i+1} + i\mu_2 P_i, \text{ and}$$

$$\begin{aligned}\frac{\partial P^m}{\partial t} &= \sum_i \left\{-i(\varepsilon+\nu+\mu_2)P_{i,m} - S\mu_1 P_{i,m} + (i-1)\varepsilon P_{i-1,m} + S\mu_1 P_{i-1,m} + (i+1)\nu P_{i+1,m} + i\mu_2 P_{i,m-1}\right\} \\ &= -\mu_2 E(I|\,M=m)P^m + \mu_2 E(I|\,M=m-1)P^{m-1}.\end{aligned} \tag{2}$$

We are interested in the fraction of animals that have (at least) one malignant cell at time t. We are thus interested in quantity $P(M(t)>0) = 1 - P^0(t)$. Equation (2) is not too complicated for the case where $m=0$, and we can easily write down the differential equation governing the development of $P^0(t)$.

$$\frac{\partial P^0}{\partial t} = -\mu_2 E(I|\,M(t)=0)P^0.$$

We call $\mu_2 E(I|\,M(t)=0)$ the *hazard, h(t),* at time t. We can solve this equation for P^0 to get

$$P^0(t) = \exp\left\{-\int_0^t \mu_2(s)E(I|\,M(s)=0)(s)ds\right\} = \exp\left\{-\int_0^t h(s)ds\right\}.$$

The hazard function has a simple intuitive meaning. Given that there are no malignant cells at time t, it is the rate at which the first one develops. This is an important quantity to be able to obtain from the model as it can be used to estimate the risk of tumour to the (up to that point) unaffected population. The function $1\text{-}P^0(t)$, called the *incidence,* is the fraction of the

population with (at least) one malignant cell at time *t*. If the tumours are fatal, this can be difficult to interpret. We will come back to this point later.

To actually calculate the model incidence or hazard, the differential equations in (1) must be solved. This is possible, but rather cumbersome. Often, a certain approximation is done (see, for example Venema and Leenhouts[7], or Leenhouts and Chadwick[8]) which simplifies the calculations greatly. It is, however, not always clear if this approximation is a good one[4].

The stochastic behaviour of the system can be removed by only calculating the expected number of intermediate and expected number of malignant cells as a function of time. It is easy to calculate *EI(t)* and *EM(t)* numerically from differential equations derived from (2).

$$\frac{dEI}{dt} = \sum_i i\frac{\partial P_i}{\partial t} = S(t)\mu_1 + (\varepsilon - \nu)EI,$$

$$\frac{dEM}{dt} = \sum_m m\frac{\partial P^m}{\partial t} = \mu_2 EI.$$

From the second differential equation we see that $EM(t) = \int_0^t \mu_2(s)EI(s)ds$. If we make the assumption that $EI(t) \approx E(I \mid M(t) = 0)$, we now see that we can approximate $P^0(t)$ by $P^0(t) \approx \exp(-EM(t))$. This approximation is valid in conditions when there are not many malignant cells. In these cases, the unconditional expected number of intermediate cells will not differ much from that given that there are no malignant cells present. Note that this approximation is equivalent to assuming that the number of malignant cells has a Poisson distribution.

In this paper we do not use this approximation, but, rather, use the exact solution obtained from much cumbersome calculation. We relegate the description of this calculation to the appendix as it is not needed in subsequent discussions in this paper.

3 MODEL PARAMETERISATION

We assume that the rates μ_1, μ_2, ε and ν are all functions of radon (and possibly ore dust) dosage, which, in turn, is a function of time. We choose the functions $\mu_1(d)$, $\mu_2(d)$, $\varepsilon(d)$ and $\nu(d)$ from biological considerations (d is the radon dose rate). We propose five models and use the data to investigate the validity of each model.

General Assumptions

There are several assumptions that we make for all the models considered. We assume that the number of stem cells, $S(t)$, is a deterministic function of time - in this paper we take $S(t)$ to be constant, equal to $5\text{x}10^5$ [3]. Leenhouts and Chadwick[8] and Venema and Leenhouts[7] take $S(t)$ to increase linearly from zero until a maximum value is reached (at adulthood). From this time on, they take $S(t)$ to be constant. Other time dependencies are also possible.

The mutations taking a stem cell to an intermediate cell, and taking an intermediate cell to a malignant cell are taken to be caused by DNA double strand breaks. The rate of a double strand break in a cell induced by radiation dose rate d is, in general $\propto \alpha.d + \beta.d^2$, where α is the rate that a single ionising particle produces a double strand break, and β is the rate per dose rate squared that two independent particles induce a double strand break. In our model, we take $\beta=0$. Radon produces α-radiation for which β is small. Also, taking β non-zero would lead to overparameterisation. We therefore consider models with $\mu_i(d) = q_i(1+\alpha_i d)$. When there is no radiation, μ_i takes on its 'background' value. We call q_i the spontaneous mutation rate.

Model 1

$$\mu_1 = q_1(1+\alpha_1.d)$$

$$\mu_2 = q_1$$

$$\delta = \varepsilon - \nu = q_3(1+\alpha_3.d)^{p_3}$$

$$\frac{\nu}{\varepsilon} = const =: \theta,$$

Following Lübeck et al.[3], in model 1, instead of fitting ε and ν, $\delta := \varepsilon - \nu$, and $\theta := \nu / \varepsilon$ are fit. Note that δ is assumed to increase with increasing dose, and that θ is dose independent. The estimated values of θ are near *1*, which gives the idea of homeostatic control in the intermediate cell population. The fact that δ increases with increasing dose can be explained by noting that an increase in radon dosage brings about an increase in intermediate cell killing (ν). To keep ν / ε constant, an increase in the intermediate cell multiplication must follow. Such a form for the δ dependency seems biologically unlikely, as one would expect radiation to cause a decrease in cell population, especially at high dose levels. In the optimal fits obtained, p_3 is small, and α_3 is large translating to a rapid increase in δ for low radon doses, and a flattening off for higher doses. Such a δ dependency can be plausibly explained in terms of the ore dust concentrations, which we discuss in section 6.1. Also, following Lübeck et al., μ_2 is taken to be dose independent. They report that this assumption does not degrade the quality of the fit. We will discuss this point further in section 6.

Model 2

$\mu_1 = q_1(1+\alpha_1.d)$

$\mu_2 = q_1(1+\alpha_2.d)$

$\delta := \varepsilon - \nu = q_3$

$\nu / \varepsilon = const =: \theta.$

In model 2 we assume that ε and ν are independent of radiation dose. There is no evident reason why radiation should promote cell division, thus the ε dose independence. For non-excessive radiation dosages, it is also believable that the intermediate cell death rate is reasonably independent of dose rate. We do not set α_2 to be zero in this fit.

Model 3

$\mu_1 = q_1(1+\alpha_1.d)$

$\mu_2 = q_1$

$\varepsilon = s_3(1+\gamma_3.d)e^{-r_3 d}$

$\nu = const.$

In model 3, we assume that ν is constant, and that ε increases with increasing radon dose to a maximum point, beyond which it then decreases to zero. At high radon dosages, it is likely that a cell is severely damaged, and that the cell multiplication is thus inhibited. As in model 1, we take $\alpha_2=0$.

Model 4a

$\mu_1 = q_1(1+\alpha_1.d)$

$\mu_2 = q_1$

$\varepsilon - \nu = q_3(1+\alpha_3.d)$

$\nu / \varepsilon = const = \theta.$

Model 4b

$\mu_1 = q_1(1+\alpha_1.d.u)$

$\mu_2 = q_1$

$\varepsilon - \nu = q_3(1+\alpha_3.d.u)$

$\nu / \varepsilon = const = \theta.$

In model 4, we try to introduce the ore dust concentration (u above) into the mutation rate formulae. Two models are considered that are the same except that the mutation rates in the first are only dependent on the radon dose rate, and those in the second are dependent on the product of the radon dose rate and the ore dust concentration. The mutation rates thus have a multiplicative dependence of the radon and ore dust concentration instead of a linear dependence as investigated by Lübeck et al..[3] The first model, is, in fact, the same as model 1, with $p_3=1$.

In comparing these similar models, it is easy to judge if introducing the ore dust concentration in this way improves the fit.
In all models t_{lag} is a parameter which can also be varied in an attempt to describe the data better.

4 MAXIMUM LIKELIHOOD METHOD OF DETERMINING PARAMETER VALUES

For each rat we know the exposure profile, time of death, whether or not a tumour was present at death, and if it was deemed fatal. We find the parameter values that optimally describe the data by calculating the *likelihood* of seeing what we see. Suppose rat i dies at time t_i with a tumour. We assign a likelihood that this occurs L_i, given the model. If there are n rats, we then say that the likelihood of seeing what we see is $L = \prod_{i=1}^{n} L_i$. Clearly, as we want to find the most likely parameter values given the data, we would like to make L as large as possible. That is, we want to find the parameters that maximise the likelihood. In practice, we consider the *loglikelihood =-log(L),* and thus find the parameters that minimise the loglikelihood.

The way we assign L_i depends on our interpretation of malignant cell incidence *P(t).* Recall that our model describes the state of a rat (the 'system') by the random variable vector *(I(t),M(t)),* - the number of intermediate cells, and the number of malignant cells at time *t.* The state of the system develops for all time, and the numbers of intermediate and malignant cells increase indefinitely. The incidence $P(t-t_{lag})$ is the probability that there is at least one tumour at time *t.* The hazard $h(t-t_{lag})$ is the instantaneous rate at which the first tumour develops, given that the animal has no tumours at time *t.* In reality, the rats die, possibly resulting from the presence of a tumour. We are interested in the time that a rat develops its first tumour. After a rat dies, its further tumour development is censored. If the rat dies without a tumour, then it cannot be said when the rat would have developed its first tumour. This makes the practical interpretation of *P(t)* difficult. If the times of death are independent of the presence or absence of a tumour, (the tumours are incidental), then $P(t-t_{lag})$ is the fraction of the remaining animals that have a tumour at time *t.* If the times of death are affected by the presence of a tumour, then the group of remaining animals at time *t* may have a different tumour distribution than would otherwise be the case. The other extreme of the above independence is to assume that all tumours are instantly fatal. Animals may, however, still die from other causes. In this case, the hazard, *h(t),* has a more intuitive meaning than *P(t).* If all the tumours are assumed to be instantly fatal, it gives the instantaneous rate at which a rat will die due to a tumour, given that it has survived up to now (translated in time by t_{lag}). In reality, some tumours may cause instant death, and others may remain harmless for some (possibly indefinite) time. This reality is difficult to model.

Likelihood 1

One possibility of assigning a likelihood to an animal death at time t_i is to assume that all tumours are *incidental*, and therefore that death of the animal is independent of the presence of tumours. Then, given an animal dies at time t_i with a tumour, we can only say that the tumour first materialised *before* time t_i-t_{lag}. This has probability (or likelihood) $P(t_i-t_{lag})$. The

probability that no tumour is present is then $1\text{-}P(t_i\text{-}t_{lag})$.

$$L_i(t_i) = \begin{cases} P(t_i - t_{lag}) & \text{if tumor present,} \\ 1 - P(t_i - t_{lag}) & \text{otherwise.} \end{cases}$$

Likelihood 2

Another possibility is only consider the tumours that were deemed fatal, and to assume that they all are *immediately* fatal. If an animal dies at time t_i with a tumour, then the first malignant cell must have first materialised *at* time $t_i\text{-}t_{lag}$. This has probability (likelihood) $P(t_i\text{-}t_{lag}+\Delta t)\text{-}P(t_i\text{-}t_{lag})$. ($\Delta t$ is the smallest time interval we consider - in our analysis it is one day.) The probability of not having a tumour at t_i is $1\text{-}P(t_i\text{-}t_{lag})$, and so this is the likelihood given to the event that the animal dies at time t_i without a tumour.

$$L_i(t_i) = \begin{cases} P(t_i - t_{lag} + \Delta t) - P(t_i - t_{lag}) & \text{if tumor present,} \\ 1 - P(t_i - t_{lag}) & \text{otherwise.} \end{cases}$$

Likelihood 3

In an attempt to use all the information available in the data, Lübeck et al.[3] constructed another loglikelihood.

$$L_i(t_i) = \begin{cases} P(t_i - t_{lag}) & \text{if incidental tumor present} \\ P(t_i - t_{lag} + \Delta t) - P(t_i - t_{lag}) & \text{if fatal tumor present} \\ 1 - P(t_i - t_{lag}) & \text{otherwise.} \end{cases}$$

Tumours may either be incidental or fatal. If an incidental tumour is present at death, the first malignant cell may have first appeared at any time before $t_i\text{-}t_{lag}$. A fatal tumour first appeared at the time of death. The probability of no tumour present at t_i is again $1\text{-}P(t_i\text{-}t_{lag})$. In this case the interpretation of $P(t)$ becomes very unclear. Fatal and non-fatal tumours are described together, without distinction, by the same $P(t)$. Knowing the value of $P(t)$ says nothing of the fraction of animals with fatal, or non-fatal tumours.

The parameter set that maximises the appropriate likelihood is found using a global fit - *Very Fast Simulated Reannealing* (VFSR)[10], followed by two local fits - *amoeba* and *powell*. These routines search within a given 'allowable' region of the parameter space for the parameter set that maximises the likelihood.

5 JUDGING THE GOODNESS OF FIT

Maximising the likelihood finds the parameter values that best describe the data. It gives no indication as to how good a fit this best fit is. To do this, we need to compare model predicted results with the experimental data. This can be done for the first two forms of the likelihood, but not for the third.

The first case is when all the tumours are taken to be incidental. For each dose group, the expected number of tumours and the observed number of tumours can be compared. Given the times of the death of the animals, t_1, t_2,...,t_k, one can calculate the model expected number of tumours in the sample. This is simply $\sum_{i=1}^{k} P(t_i - t_{lag})$. As there are many (37) dose groups, it is advantageous to express the agreement of these two sets of figures in one number. To do this we calculate the χ^2 value between the model and observed number of tumours in each dose group. This is calculated according to the following formula:

$\chi^2 = \sum_{i:groups} \frac{(obs_i - mod_i)^2}{\max(obs_i,1).\#groups}$. Here obs_i is the observed number of tumours in a dose group, and mod_i is the model predicted number of tumours for that dose group. A value of χ^2 of around *1* tells us that the model and observed values are in good agreement.

The second case is when only the fatal tumours are considered. In this case an empirical *P(t)* can be built up for each dose group from the empirical hazard. Recall that the hazard is the rate that the first tumour develops given that there are none up to that point. When an animal dies with a fatal tumour, we can say that the first malignant cell materialised t_{lag} days earlier. The times of death of the animals in the dose group are ordered $t_1 \le t_2 \le ... \le t_k$ and for each t_i the hazard $h(t_i\text{-}t_{lag})$ and the survival $(1\text{-}P(t_i\text{-}t_{lag}))$ can be calculated from what has already been calculated. If the *i*th animal dies with a tumour, then, the hazard at time $t_i\text{-}t_{lag}$ can be estimated by $h(t_i\text{-}t_{lag})=1/(k\text{-}i+1)$ where $k\text{-}i+1$ is the number of animals still alive at t_i. This is then the empirical probability that a tumour first develops at time $t_i\text{-}t_{lag}$ given that none has developed before this time. The survival function is then

$1 - P(t_i - t_{lag}) = (1 - h(t_i - t_{lag}))(1 - P(t_{i-1} - t_{lag}))$. In words this means that the probability that an animal has no malignant cells after just after time $t_i\text{-}t_{lag}$ is the probability that it had none just before this time, multiplied by the probability that it develops none at this time. If the *i*th animal dies without a tumour, then the survival function keeps the same value :

$1 - P(t_i - t_{lag}) = 1 - P(t_{i-1} - t_{lag})$. The empirical function *P(t)* can then be plotted against the model *P(t)* for the optimal parameter values found. The 'goodness of fit' is a good indication as to how well the model describes the data.

It is not possible to construct an empirical *P(t)* from the data in the case when all tumours are considered to be incidental. When an animal dies with a tumour, it is not possible to pinpoint precisely when the tumour first appeared. On the other hand, when all tumours are considered to be fatal, it is also not possible to construct a model expectation of the number of tumours given the times of the deaths. This is made impossible by the dependence of the times of deaths on the presence or absence of tumours. Nothing is known about the model distribution

of the times of death not due to tumours. It is not clear how to construct any way of checking the 'goodness of fit' for the optimal solution obtained by maximising the third likelihood.

6 RESULTS

6.1 Only Incidental Tumours

First we compare the optimal parameter values found by maximising likelihood 1. (This assumes that all tumours are incidental.) The dose rate dependence used in Lübeck's paper[3] (model 1) comes up with the lowest loglikelihood, and thus the highest likelihood. The different loglikelihoods are shown in Table 1.

Model	Loglikelihood
model 1	354.5
model 2	366.7
model 3	370
model 4a	371
model 4b	405.5

Table 1: The loglikelihoods for the different optimal model solutions, loglikelihood 1.

As mentioned earlier, all parameters result from first a global fit, and then a local fit being carried out. The model 3 fit is an exception to this. The global fit was unable to find a good solution, and so a local fit was carried out around a solution that was considered good. In Table 2 are the expected number of tumours occurring in each dose group according to models 1 and 2, and the actual numbers found. Comparing these two values gives a good indication of the goodness of the fit.

Total WLM	#days exp	# in group	observed #	Expected # Model 1 (mhfit136)	Expected # Model 2 (mhfit137)
0	0	32	0	0.2	0.2
0	0	32	0	0.16	0.18
0	0	32	1	0.24	0.23
0	0	32	0	0.26	0.25
0	0	63	0	0.4	0.4
0	0	95	2	0.79	0.74
0	0	140	1	1.31	1.2
0	0	128	0	0.95	0.92
20	2	246	2	3.4	3.54
40	5	445	9	8.15	8.82
80	10	382	15	9.2	11.28
80.1	114	383	8	11.52	10.91
160	20	191	6	7.91	9.64
315.5	11	31	1	1.32	2.1
319.8	19	30	2	1.83	2.8
320	43	95	2	6.39	7.91
320.3	43	127	8	10.66	11.83
321	458	128	19	20.15	8.54
321.6	4	151	12	7.93	12.25
325.5	13	32	5	1.9	3.09
638	87	96	6	14.18	14.85
638.9	8	88	5	7.15	11.9
641	86	64	18	10.97	10.26
642	58	32	7	3.6	4.54
642.3	35	95	13	9.2	13.31
644	112	32	5	2.73	3.2
1280	172	32	17	12.24	9.72
1280.1	17	56	11	9.41	12.83
2559	113	62	21	23.56	24.45
2560	355	32	22	20.33	11.74
2580	218	32	15	16.69	13.07
2601.3	35	56	16	17.11	23.56
5090	147	32	21	18.51	17.17
5114.9	71	56	23	28.21	33.18
5120	701	32	25	23.01	16.98
5130	147	96	47	45.28	41.63
10254.5	138	64	38	43.68	42.01

Table 2: Expected number of tumours in different dose groups as predicted by optimal model 1 and model 2 solutions (likelihood 1), and the observed number of tumours in each of these groups.

The χ^2 values for the two models above are *1.23* and *2.27* respectively. The equivalent values for the fits for models 3, 4a and 4b are *2.59, 2.26* and *3.68* respectively. From this, we conclude that the model 1 fit is significantly superior to the other fits.

The expected and actual values for all the different models are also given in Figures 2 to 5: the model predictions are plotted against the empirical values. On each graph, a different fit is plotted together with the model 1 fit. If the model were to agree perfectly with the data, such

a plot should give a straight line with slope 1

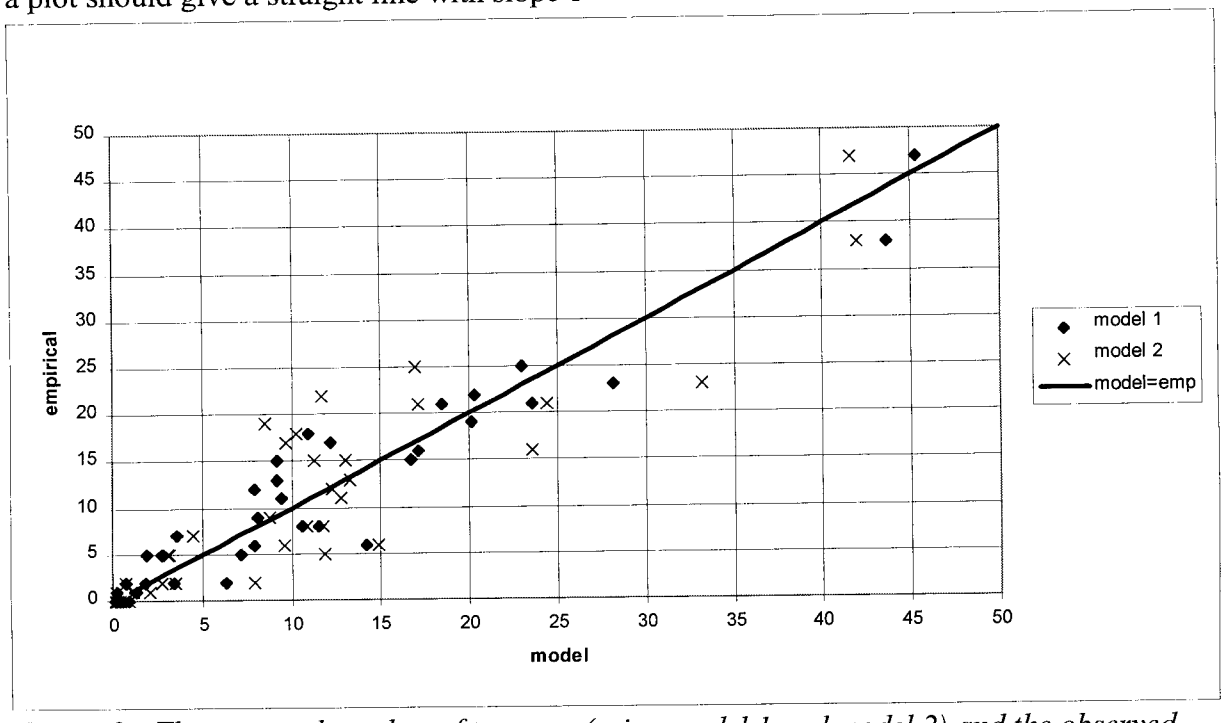

Figure 2: The expected number of tumours (using model 1 and model 2) and the observed number of tumours in each dose group. A perfect model would give points that lie on the straight line.

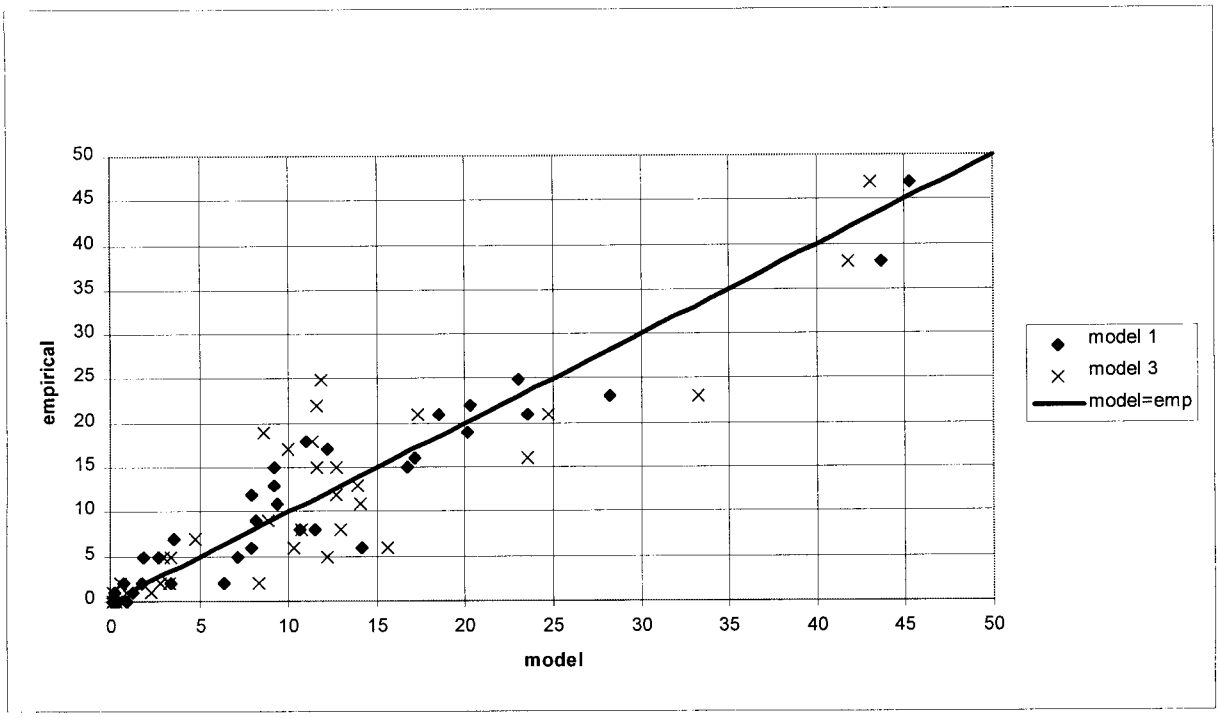

Figure 3: The expected number of tumours (using model 1 and model 3) and the observed number of tumours in each dose group. A perfect model would give points that lie on the straight line.

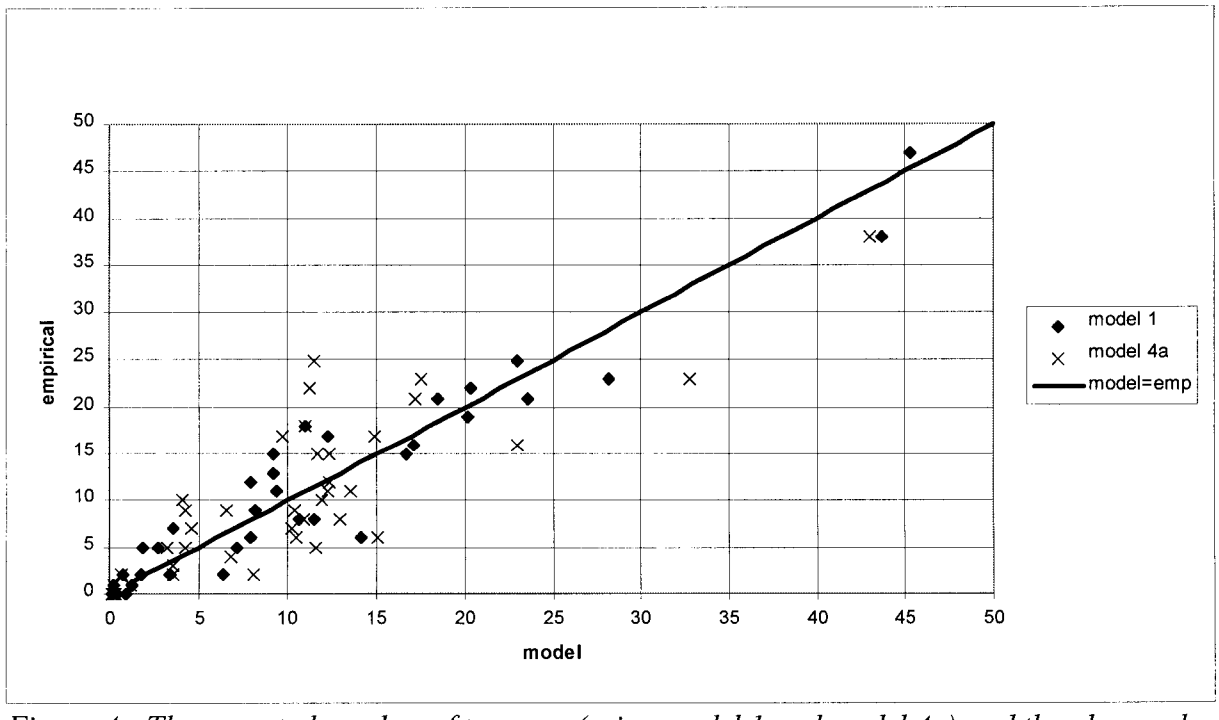

Figure 4: The expected number of tumours (using model 1 and model 4a) and the observed number of tumours in each dose group. A perfect model would give points that lie on the straight line.

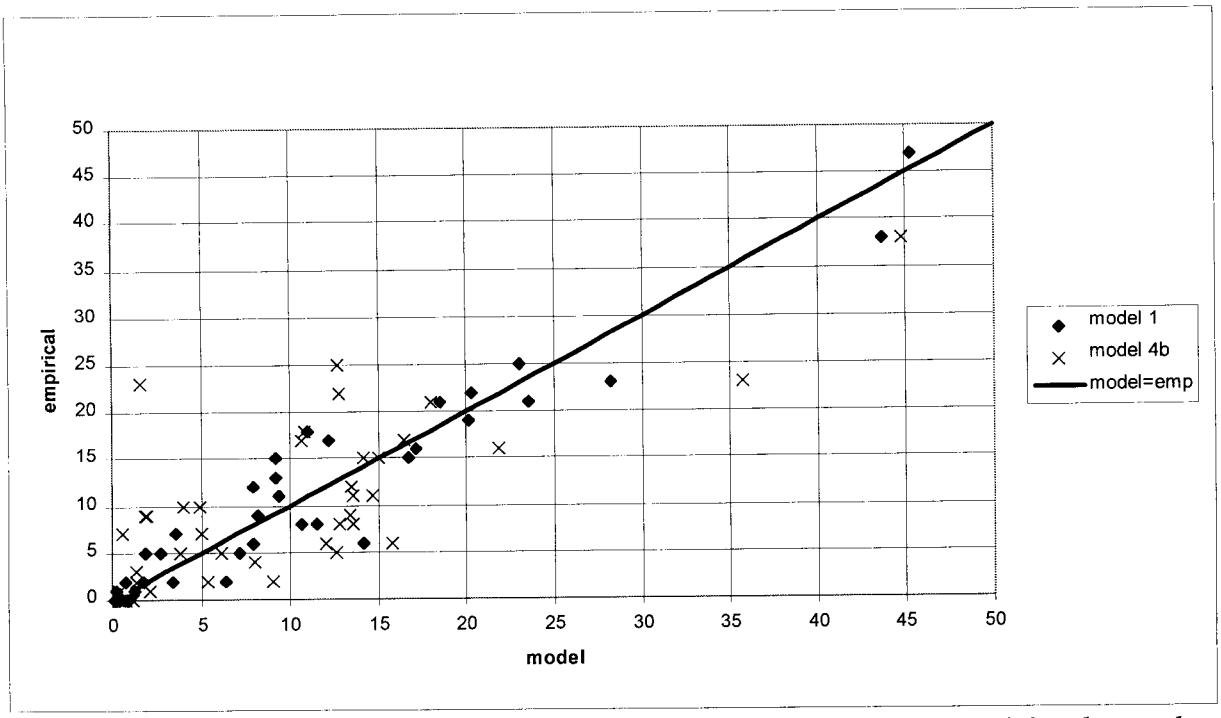

Figure 5: The expected number of tumours (using model 1 and model 4b) and the observed number of tumours in each dose group. A perfect model would give points that lie on the straight line.

Note that the data was divided into different dose groups for the model 4a and 4b calculations compared to the model 1-3 calculations. The data was divided into groups that had the same radon *and* ore dust dosage profile.

In Table 3 are the optimal parameters obtained from the different models.

	Model 1	Model 2	Model 3	Model 4a	Model 4b
Loglikelihood	354.5	366.7	370	371	405.5
q1 [1/d]	1.70E-07	5.93E-09	2.54E-07	2.70E-07	2.71E-07
q3 [1/d]	3.25E-03	2.44E-02		4.44E-06	6.75E-03
s3 [1/d]			7.62E-03		
p3 [ ]	9.15E-02				
r3 [ ]			4.24E-03		
α1 [d/WLM]	7.98E+00	1.37E+01	2.03E+01	1.56E+01	5.74E-01
α2 [d/WLM]		8.97E+00			
α3 [d/WLM]	9.44E+05			7.24E-03	4.13E-07
γ3 [d/WLM]			7.20E-04		
tlag [d]	1.14E+02	1.99E+00	1.61E+02	1.54E+02	1.44E+02
θ [ ]	9.97E-01	9.92E-01		9.72E-01	9.99E-01
ν [1/d]			7.86E-03		

Table 3: Optimal parameters found for the different models, using loglikelihood 1.

The Lübeck fit (model 1) has the lowest loglikelihood of all the fits, and thus can be considered to be the best. From the above graphs, and χ^2 values, we can conclude that the model 1 fit is indeed a good fit. Note that in this fit, p_3 is small and α_3 is large. This means that the intermediate cell overall growth rate δ is a particular value when there is no radon

exposure, and much higher when there is any radon exposure. After this initial jump, the overall growth rate is a slowly increasing function of the dose rate. Lübeck noted that rats were exposed to uranium dust only when they were also exposed to radon, and the uranium dust concentration was mostly fairly constant. He then argues that the initial sharp increase in this multiplication rate is due to the uranium dust, and then the further increase is due to the increases in radon concentration. Lübeck has also fitted a model where the ore dust presence gives a fixed increase in the self multiplication rate, and, on top of this, the multiplication rate increases linearly with radon dose. This model gave similar results to the model 1.

It is important to determine the rate determining steps in the process. A comparison of the relative sizes of α_1 and α_2 is important in determining the effect of chronic radon exposure on tumour incidence. During short exposures at the start of the animals life, due to the small number of intermediate cells present, a much increased second mutation rate (during radon exposure) does not have a large effect on the total tumour incidence. For radon exposures later in life, after the number of intermediate cells has had time to grow, α_2 plays a much more important rôle. As we wish to make predictions about chronic low radon exposures, it is important to compare the values of α_1 and α_2. Lübeck found that for his model, α_2 was much smaller than α_1, and could be set to zero without decreasing the goodness of fit. In model 1, α_2 is set to *0*, as in Lübeck's model. An additional run was carried out where α_2 was also allowed to vary. It was indeed found that α_2 was much smaller than α_1 (α_1=*1.18E+01*, α_2 =*4.68E-04*). The loglikelihood of this fit (*354.9*) was comparable to the run where α_2 was set to *0*. Another run was carried out with α_1 set to be equal to α_2. The resulting loglikelihood was *363.9*. This is substantially higher than that found in the model 1 optimal fit. From this we can conclude that, for model 1, the model with the best fit, α_2 is much smaller than α_1.

If the self multiplication of the intermediate cells is taken to be radiation independent (model 2), and α_2 is fitted, it is found that the optimal fit has comparable values of α_1 and α_2. A fit of similar quality was found, however, with α_1 much larger than α_2. (This is discussed further in section 6.5.) In this model, the background intermediate cell self multiplication rate is much higher than in the model 1 fit, but the background first and second mutation rates (q_1 and q_2) are much smaller. The fit here is of inferior quality to that from model 1, as seen in the above graphs, and from the higher loglikelihood. It appears from this that the intermediate cell kinetics is dose dependent.

In model 3 a different dose rate dependence is used for the intermediate cell kinetics. The intermediate cell death rate is assumed to be dose independent. The self multiplication rate is assumed to increase linearly with the radon dose, but, reaches a maximum, and decreases to zero for extremely high dose rates. This dose dependence gives an inferior fit to model 1, and a slightly inferior fit to model 2.

Models 4a and 4b are studied to consider a particular uranium ore dust concentration

dependence. This resulted in a far inferior fit, and so we do not discuss the optimal parameter values further. Models 4a and 4b were run in order to directly compare fits with and without the uranium dust incorporated. As seen from the resulting optimal loglikelihoods, a superior fit is obtained if the rates do not depend on the uranium dust concentration (at least not in this way).

6.2 Only Fatal Tumours

Next we compare the models where the optimal parameters are found by maximising likelihood 2. This is where we only consider the tumours that were deemed fatal by the pathologist. Incidental tumours are thus ignored. Although many animals died with a tumour, relatively few of these tumours were deemed to be the cause of death. Again, the dose rate dependence used in Lübeck's[3] paper (model 1) comes up with the lowest loglikelihood, and thus the highest likelihood.

model	loglikelihood
model 1	583.2
model 2	606.4
model 3	598.4

Table 4: The loglikelihoods for the different optimal model solutions, loglikelihood 2.

Recall that if only fatal tumours are considered, empirical incidence curves can be constructed for the different dose groups, and can be directly compared with the model incidence ($P(t)$) curves. To give an idea of the goodness of the different optimal fits, these curves were plotted against each other for several 'typical' dose groups. (See Figures 6 to 10.)

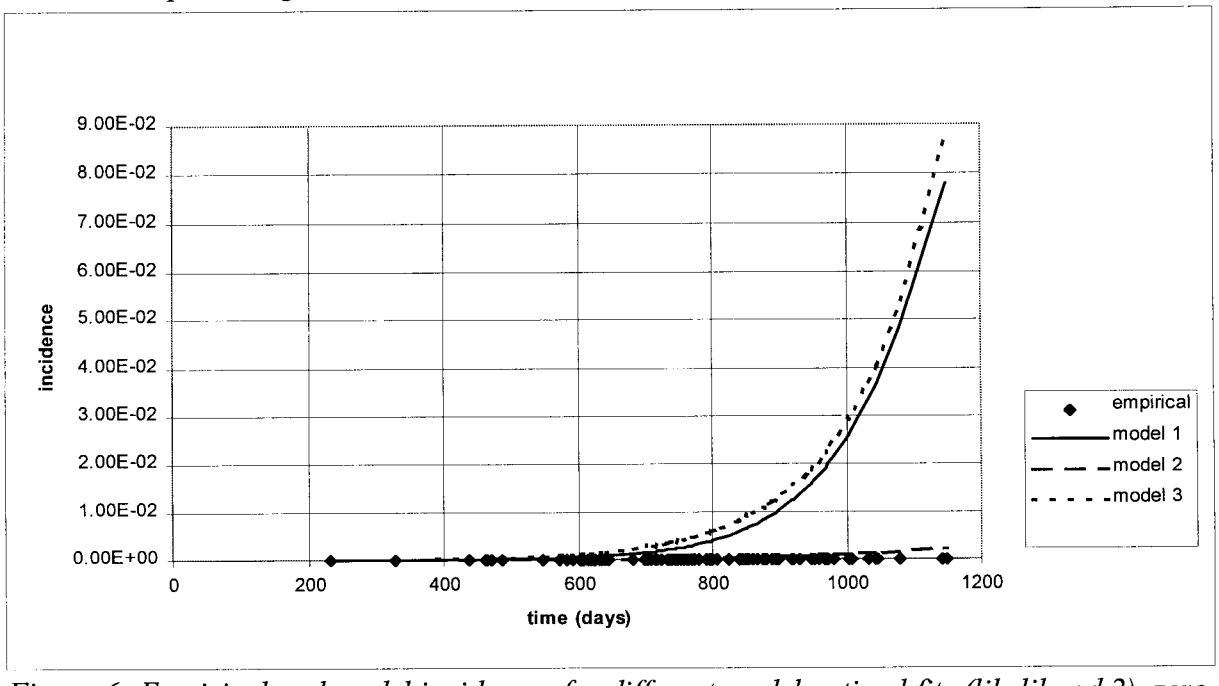

Figure 6: Empirical and model incidences for different model optimal fits (likelihood 2), zero dose.

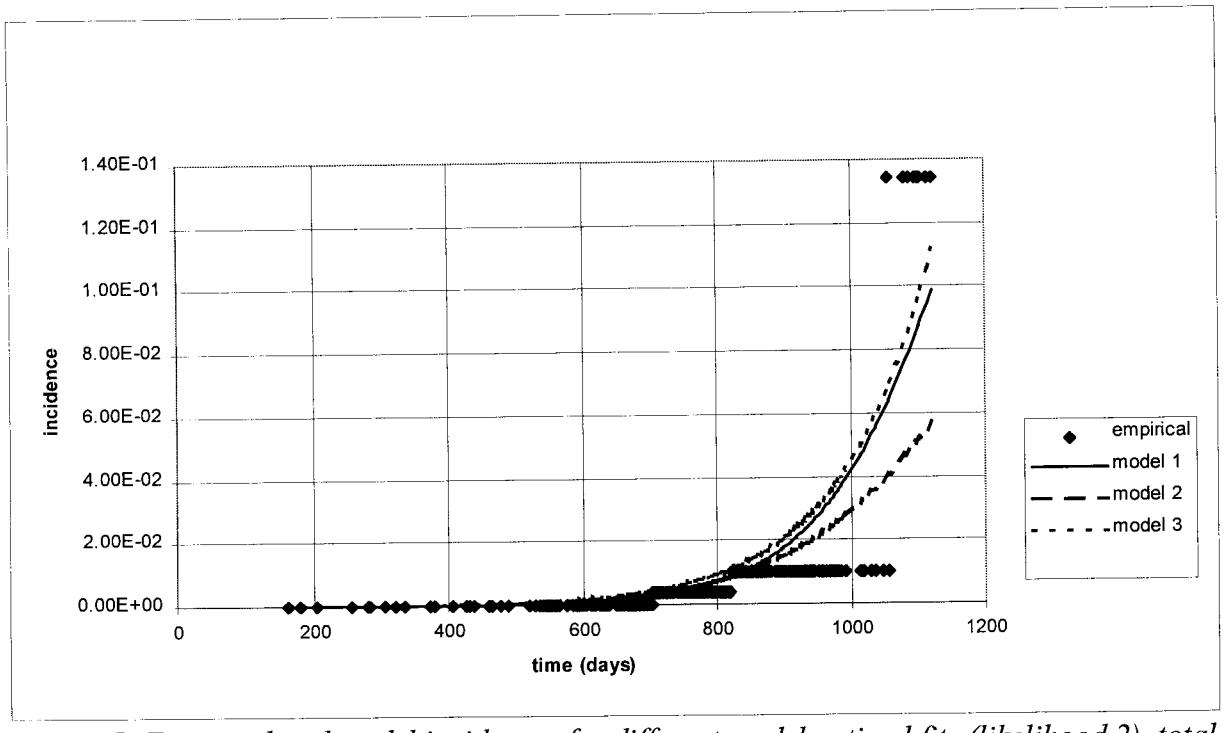

Figure 7: Empirical and model incidences for different model optimal fits (likelihood 2), total dose 80 WLM, rate 0.7 WLM/d, starting day 100.

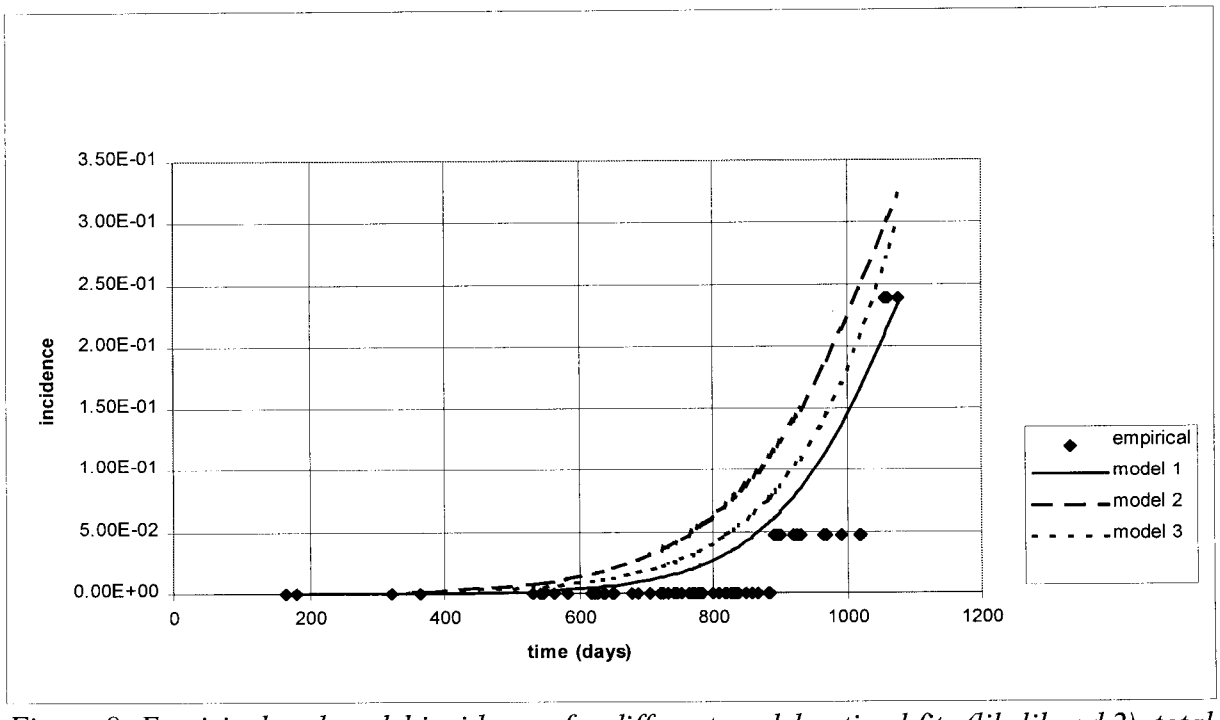

Figure 8: Empirical and model incidences for different model optimal fits (likelihood 2), total dose 638 WLM, rate 7.3 WLM/d, starting day 96.

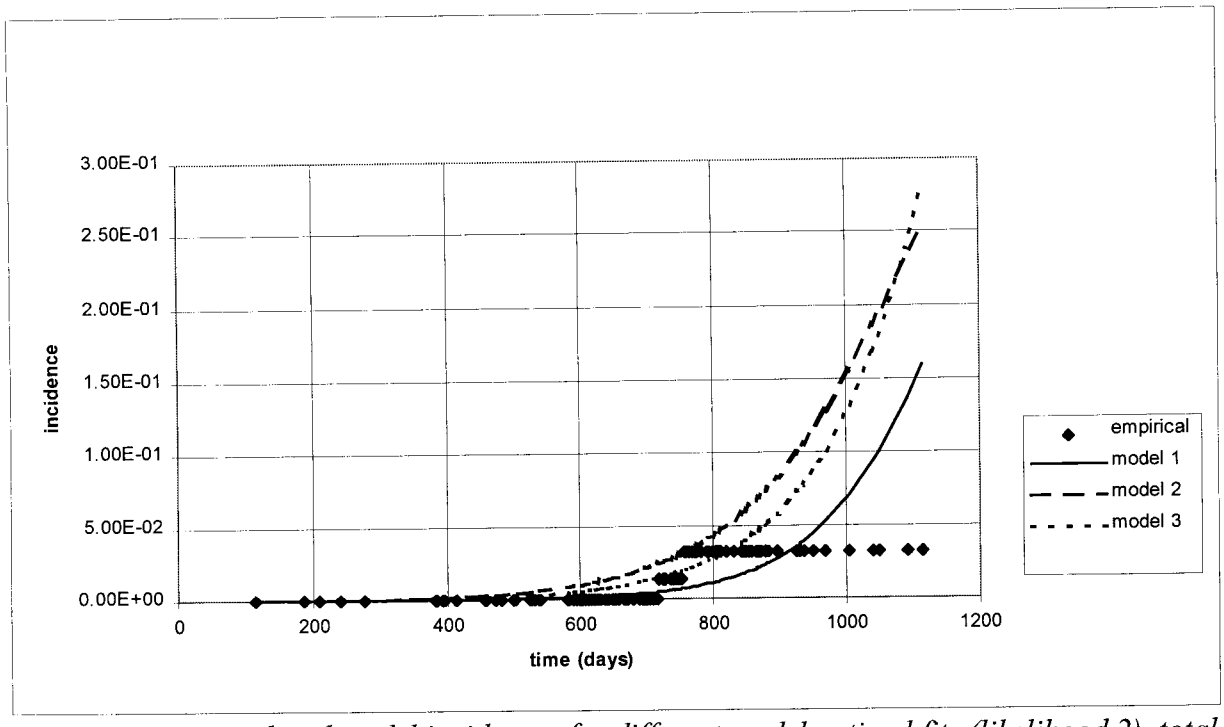

Figure 9: Empirical and model incidences for different model optimal fits (likelihood 2), total dose 321 WLM, rate 80.4 WLM/d, starting day 87.

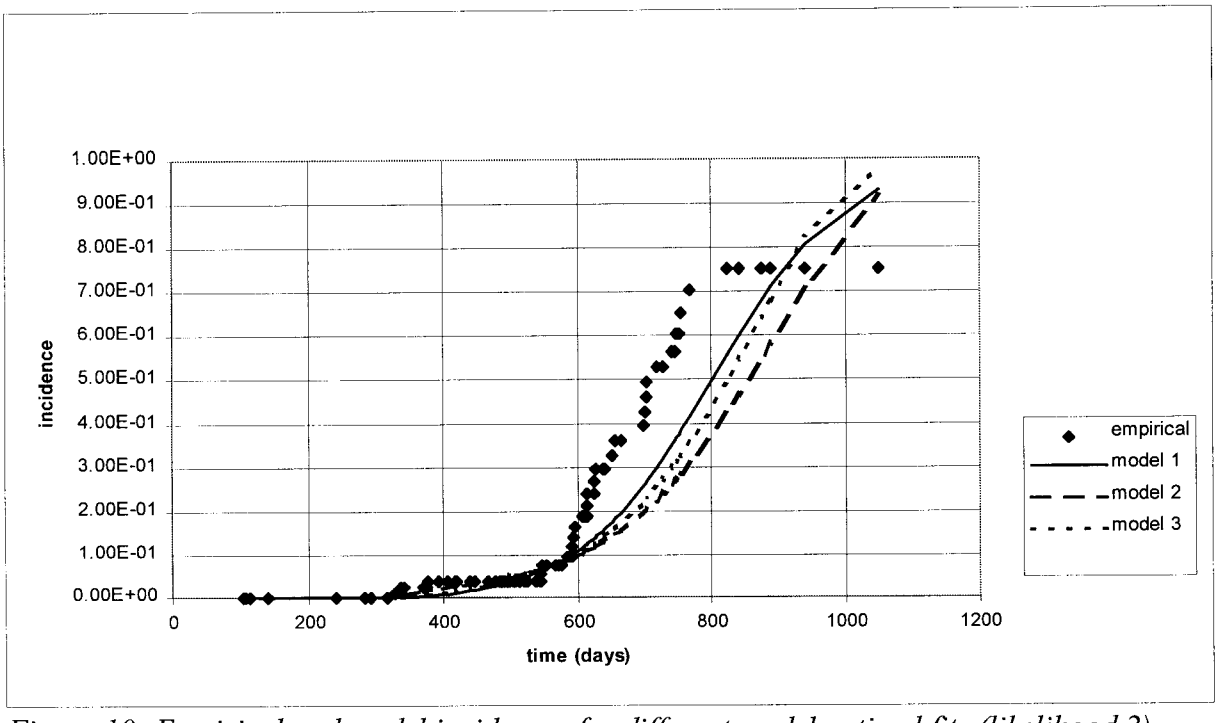

Figure 10: Empirical and model incidences for different model optimal fits (likelihood 2), total dose 5130 WLM, rate 34.8 WLM/d, starting day 87.

In Table 5 are the optimal parameters of the corresponding to the different models.

	Model 1	Model 2	Model 3
Loglikelihood	538.2	606.4	598.4
q1 [1/d]	3.66E-08	1.64E-08	6.98E-08
q3 [1/d]	9.87E-03	7.59E-03	
s3 [1/d]			7.81E-03
p3 [ ]	2.48E-01		
r3 [ ]			9.31E-03
α1 [d/WLM]	1.20E+00	1.31E+02	2.81E+00
α2 [d/WLM]		3.27E-02	
α3 [d/WLM]	4.65E+00		
γ3 [d/WLM]			6.22E-02
tlag [d]	1.51E+02	1.59E+02	1.54E+02
θ [ ]	9.49E-01	9.98E-01	
ν [1/d]			7.01E-06

Table 5: The optimal model 1, model 2 and model 3 parameters obtained using loglikelihood 2.

Here, again we find that the Lübeck model (model 1) gives the best loglikelihood. The Lübeck model has a higher value for ε-ν than for the other models, which is further enhanced for higher dose rates. It seems that the fit is better if ε increases with increasing dose rate, as in model 1. Also, it seems that α_2 can be taken to be much smaller than α_1, but, it is imaginable that there may be parameter sets with α_1 and α_2 of similar magnitude, which give similar fits. (See section 6.5.)
The optimal fit from model 3 is better than that in model 2, but still inferior to model 1 - Lübeck's model. This again tells us that it seems that, in order to describe the data well, we must assume that there is a dose dependence in ε. In model 3, ε reaches its maximum value at a dose rate of 91.33 WLM/d, (which is higher than most doses considered) and is linear for low doses. The 'step function' behaviour of ε in model 1 seems to describe the data better.
In general, the fits seem to be quite good. In the graphs with low dose rate profiles, it is important to note that the scale is quite small. In these graphs, large differences between the observed and model incidence are actually quite small. One should also notice that, as there are so few animals that die from a lung tumour, the data is rather difficult to fit. The empirical incidence curves are step functions. This is clearly a negative aspect of calculation with this likelihood.

6.3 Both Incidental and Fatal Tumours

Finally, we compare the models obtained by optimising the third form of the likelihood. This is the likelihood used in Lübeck's paper. As explained earlier, it is impossible to attach some sort of idea of the 'absolute' goodness of the fit. We can only compare the values of the loglikelihoods, and say which of the different optimal fits is the best. As for the two above forms of the likelihood, model 1 seems to give the best optimal fit. We also include the results obtained when Lübeck's optimal parameters were used. These are, of course, similar to the parameters found as the optimal parameters of model 1. A local fit around Lübeck's

optimal parameters returned the same Lübeck parameter values. The model 1 optimal parameters are slightly different.

model	loglikelihood
Lübeck	757.6
model 1	755.4
model 2	774.9
model 3	761

In Table 6 are the parameter values found from these optimal fits.

	Lübeck	Model 1	Model 2	Model 3
Loglikelihood	757.6	755.4	774.9	760
q1 [1/d]	1.24E-07	1.03E-07	1.21E-07	1.25E-07
q3 [1/d]	5.54E-03	5.33E-03	4.32E-03	
s3 [1/d]				8.41E-03
p3 [ ]	1.05E-01	7.38E-02		
r3 [ ]				2.52E-02
α1 [d/WLM]	1.09E+01	1.53E+01	2.56E+01	1.16E+01
α2 [d/WLM]			3.83E-09	
α3 [d/WLM]	5.87E+02	1.30E+04		
γ3 [d/WLM]				1.26E-01
tlag [d]	1.84E+02	1.82E+02	1.84E+02	1.79E+02
θ [ ]	9.93E-01	9.93E-01	5.80E-02	
ν [1/d]				3.40E-03

Table 6: The optimal model 1, model 2 and model 3 parameters obtained using loglikelihood 3.

As for the previous optimal fits, it appears that the best fits are obtained when the self multiplication rate ε increases with increasing radon dose rate, the best dependence being given in model 1. For this loglikelihood, it seems that α_2 is always smaller than α_1. To check this, an additional run was carried out where α_2 was also allowed to vary. It was indeed found that α_2 was much smaller than α_1 (α_1=*1.11E+01*, α_2=*1.58E-05*). The loglikelihood of this fit (*755.3*) was the same as in the run where α_2 was set to *0.*. Another run was carried out with α_1 set to be equal to α_2. The resulting loglikelihood was *800.0*. This is substantially higher than that found in the model 1 optimal fit. From this we can conclude that, for model 1, (the model with the best fit), α_2 is much smaller than α_1.

6.4 Comparison of These Different Approaches

As mentioned above, the interpretation of the incidence *P(t)* is different for the calculations with the different loglikelihoods. It is still informative to plot the incidences together on the same graph for different radon exposure profiles. We do this for model 1 fits. The fit with likelihood 2 only considers tumours that were considered fatal, where the fits with likelihood 1 and likelihood 3 use all the tumours discovered - incidental and fatal. Consequently, the

incidences found using likelihoods 1 and 3 are higher than those using likelihood 2.

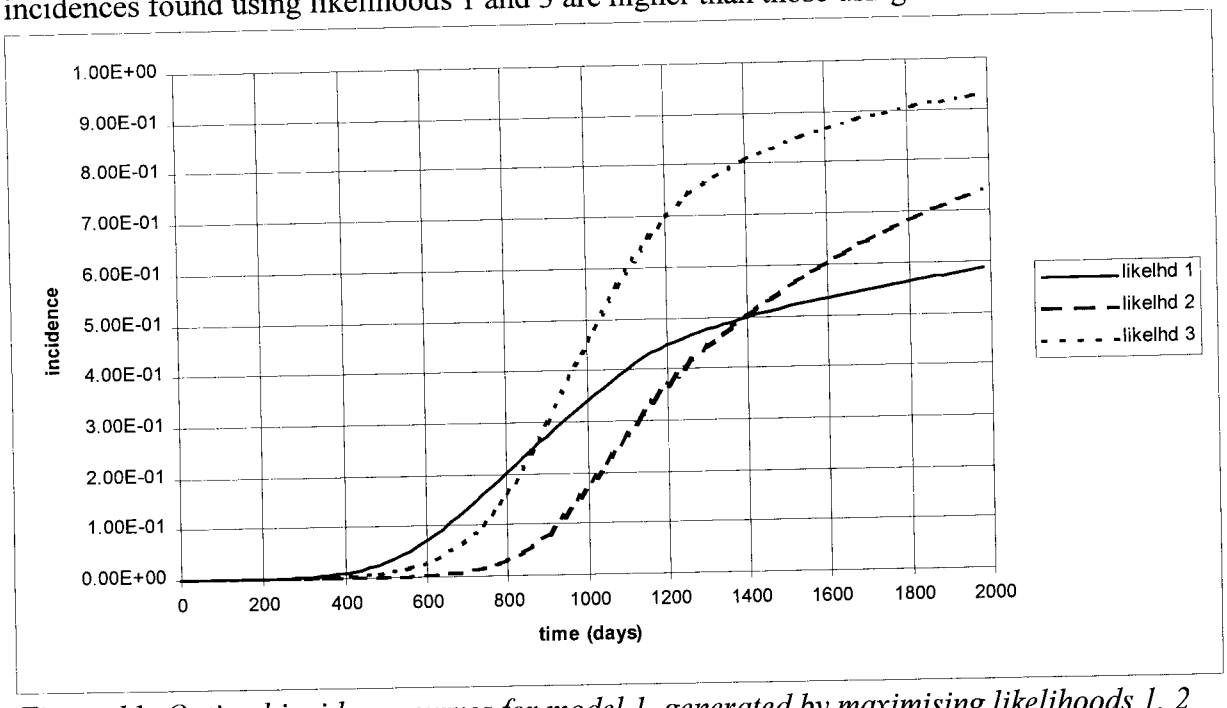

Figure 11: Optimal incidence curves for model 1, generated by maximising likelihoods 1, 2 and 3. Total dose 500 WLM, rate 0.5 WLM/d, starting day 1.

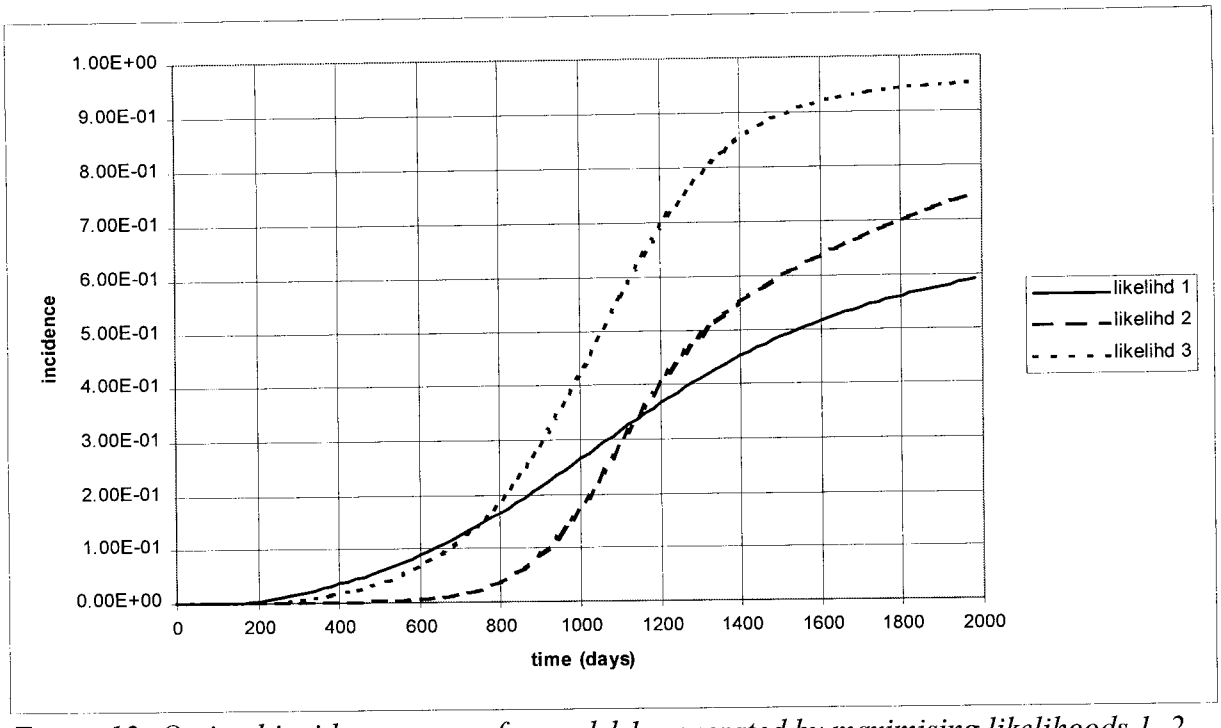

Figure 12: Optimal incidence curves for model 1, generated by maximising likelihoods 1, 2 and 3. Total dose 500 WLM, rate 5.0 WLM/d, starting day 1

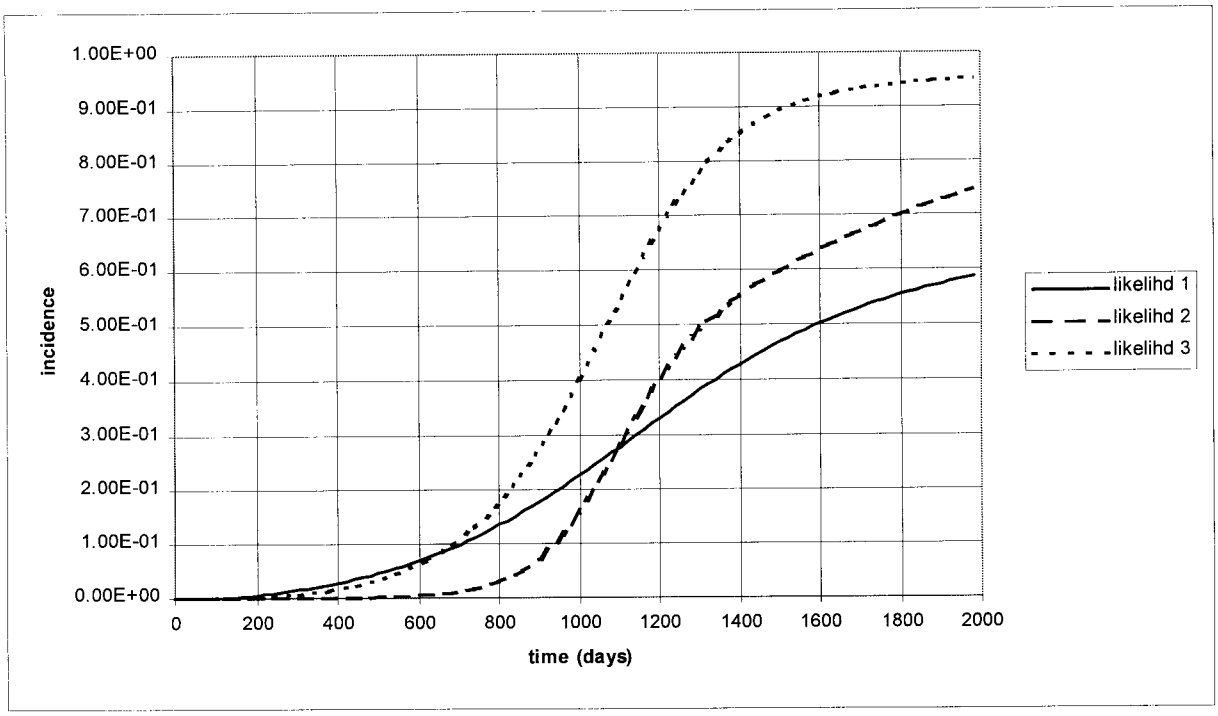

Figure 13: Optimal incidence curves for model 1, generated by maximising likelihoods 1, 2 and 3. Total dose 500 WLM, rate 50 WLM/d, starting day 1

6.5 Different Optimal Fits

As already mentioned, the parameter space is such that there are completely different sets of parameter values that give comparable fits. The following two parameter sets for model 2 give comparable values for likelihood 1.

	set 1	set 2
logliklihd	366.7	369.7
q1 [d-1]	5.92E-09	2.10E-07
α1 [d/WLM]	1.37E+01	1.67E+01
α2 [d/WLM]	8.97E+00	5.35E-06
ε [d-1]	3.24E+00	2.12E+01
ν [d-1]	3.21E+00	2.12E+01
tlag [d]	1.99E+00	1.45E+02

Table 7: Parameter values for two different, but comparable fits for likelihood 1, model 2.

We can examine the quality of fit for these two parameter sets by comparing the expected number of tumours with the observed in each dose group in the graph below. These fits only differ greatly in the dose group where the rats were irradiated with 7.3 WLM/d for 701 days. (The group with empirical value 25.)

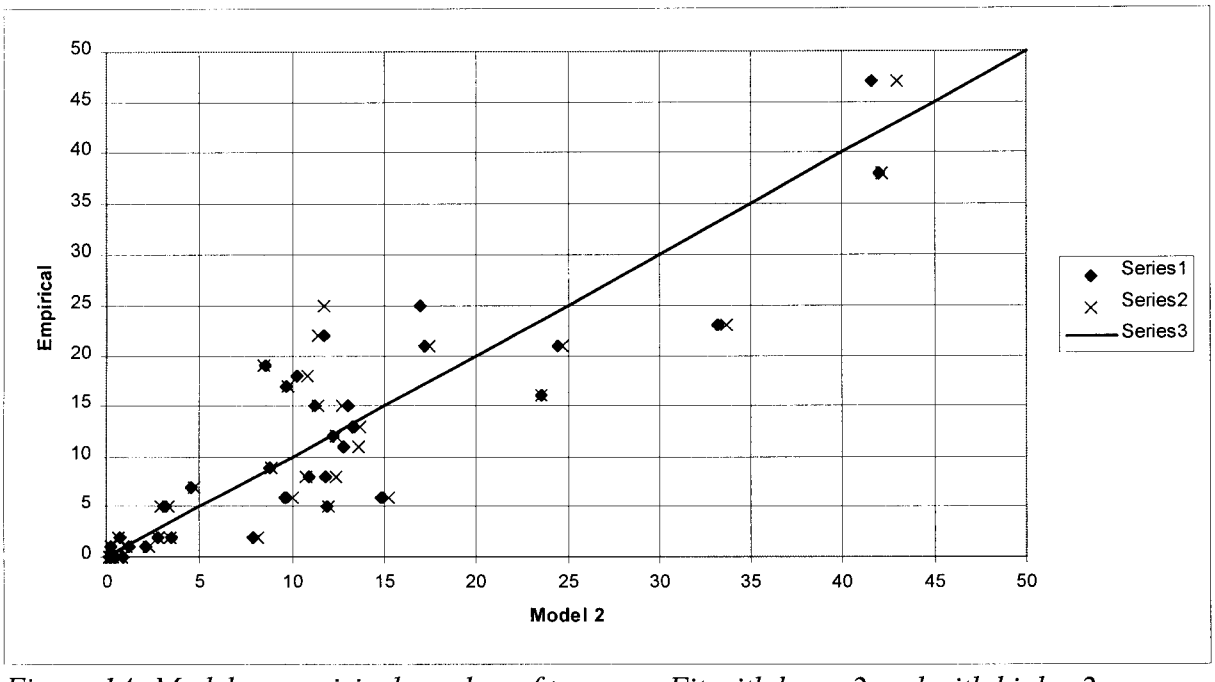

Figure 14: Model vs empirical number of tumours. Fit with low α2 and with high α2.

The incidence curves generated from the two different parameter sets agree for low level chronic radon dosages, but differ on moderate level chronic dosages and also for high level dosages. The model incidence for these two parameter sets are compared for three radon dose profiles in Figures 15, 16 and 17.

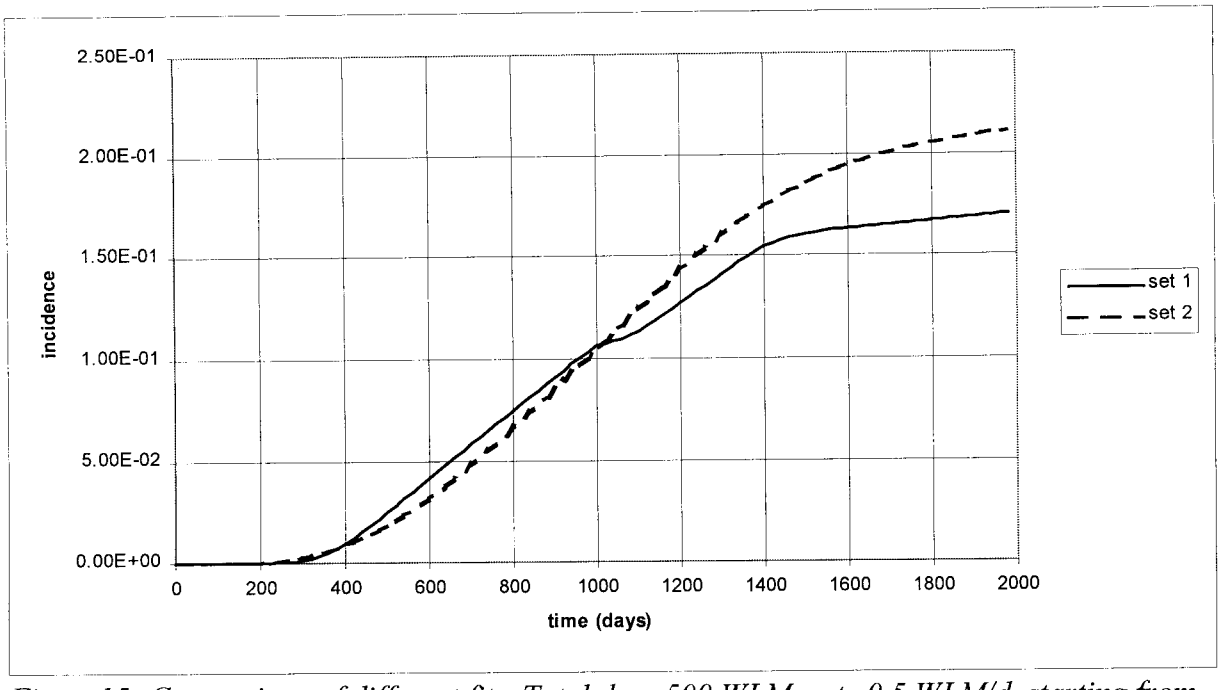

Figure15: Comparison of different fits. Total dose 500 WLM, rate 0.5 WLM/d, starting from day 1.

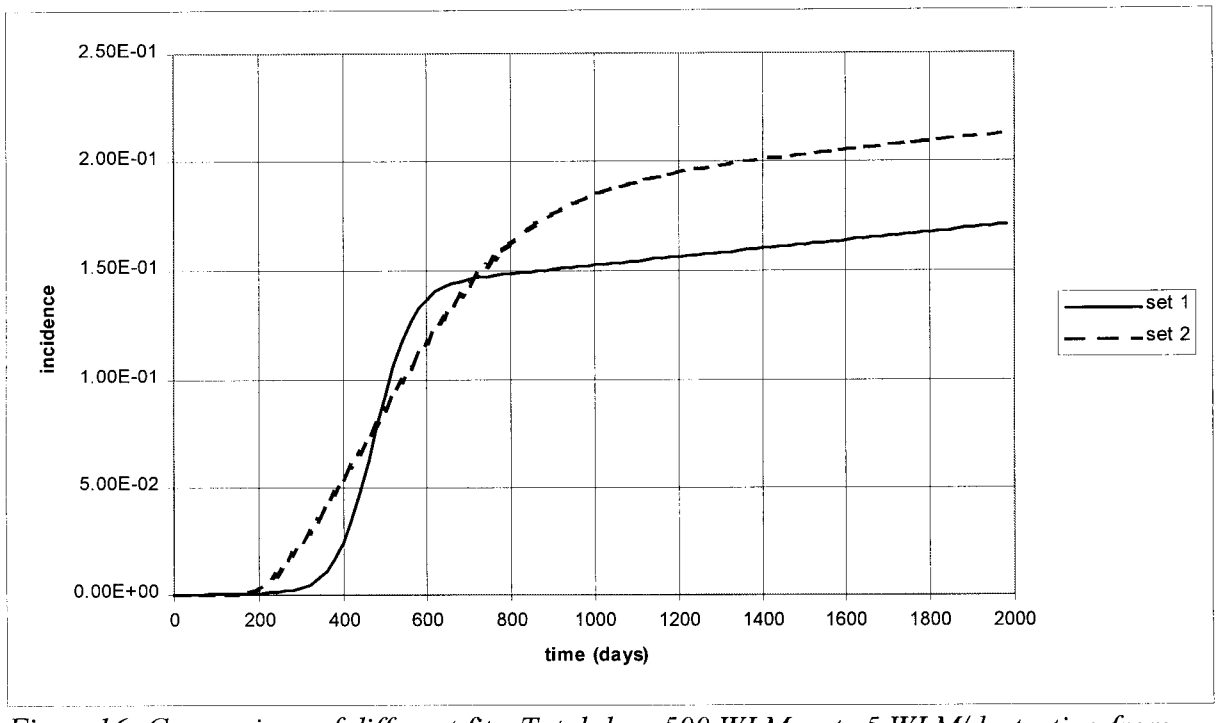

Figure16: Comparison of different fits. Total dose 500 WLM, rate 5 WLM/d, starting from day 1.

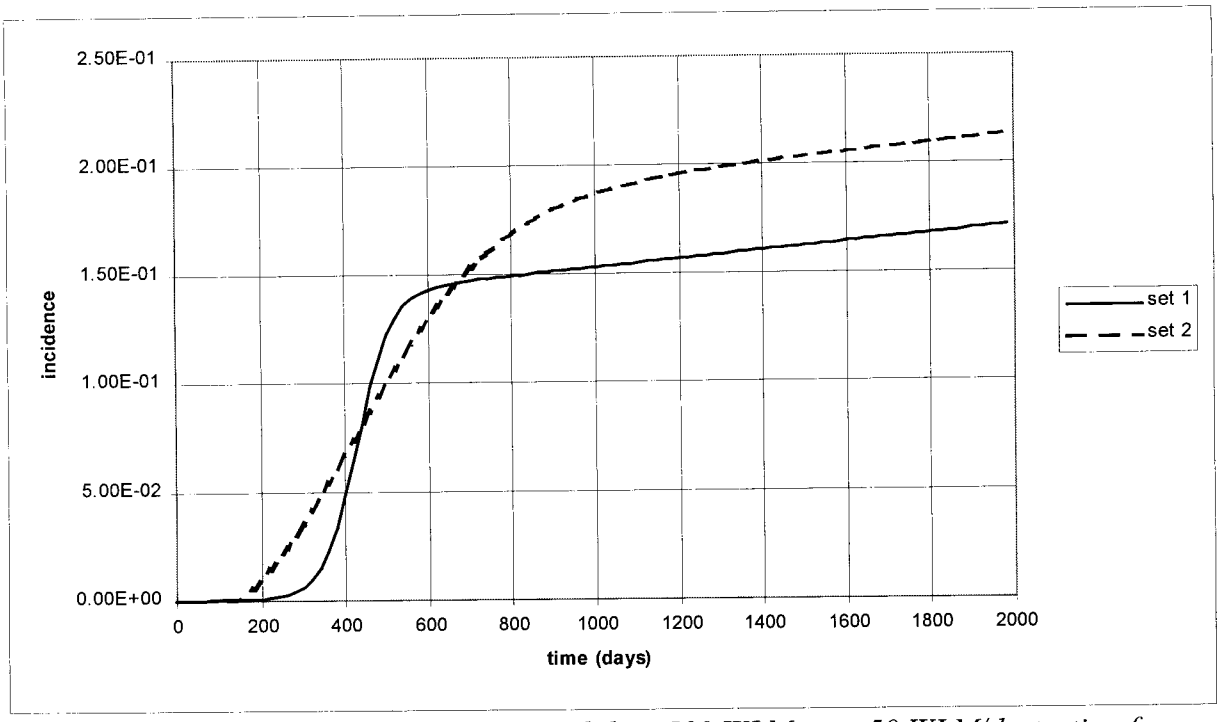

Figure17: Comparison of different fits. Total dose 500 WLM, rate 50 WLM/d, starting from day 1.

Due to the differences in the incidence curves generated from the two sets of parameters, we conclude that much care should be taken in using such models to estimate tumour risk. Two models that describe the data equally well may have different behaviours for the dosage profiles not covered by the data.

6.6 Exact vs Approximate Solution

As discussed in the introduction, in this paper we calculate *P(t)* according to the exact solution given in the appendix. This is different to the approximate calculation of *P(t)* used in Venema and Leenhouts[7], and Leenhouts and Chadwick[8]. In this section we will compare the exact and approximate solutions. Recall that the approximation uses $EI(t) \approx E(I \mid M(t) = 0)$ to estimate the hazard, and thus the incidence. In time, the probability of having no malignant cells decreases, and the approximation becomes less valid. For times with high malignant tumour incidence, the approximation gives an entirely inaccurate value for the hazard.

The two-stage carcinogenesis model is a stochastic model, where, at any given time, the number of intermediate and malignant cells in an animal is taken from a particular distribution. It is easiest to consider the distribution as being represented by a large group of animals with the same radon exposure profile. In calculating *EI(t),* the number of intermediate cells is averaged over *all* the animals. In calculating $E(I \mid M(t) = 0)$, the number of intermediate cells is only averaged over the animals with no malignant cells. When an animal with no malignant cells develops its first, it is removed from the population from which $E(I \mid M(t) = 0)$ is calculated. The rate of appearance of a malignant cell is highest amongst animals with many intermediate cells, and so, the animals removed from the population from which $E(I \mid M(t) = 0)$ is calculated are generally animals with many intermediate cells. In this way $EI(t) \geq E(I \mid M(t) = 0)$ always and the difference increases with time. As *EI(t)* increases for all time, the approximate hazard also increases (eventually, very sharply) for all time. The conditional expectation $E(I \mid M(t) = 0)$ and thus the exact hazard approaches an asymptotic value[9]. This can be understood intuitively as follows. There are two factors that affect $E(I \mid M(t) = 0)$. Animals with many intermediate cells that develop a tumour are removed from the population used in calculating $E(I \mid M(t) = 0)$, thus decreasing its value. The remaining animals develop more intermediate tumours from promotion of stem cells and self multiplication of intermediate cells. In time these two factors balance each other and the hazard remains constant.

The exact and approximate incidence functions are similar for early times, but often diverge for later times. Sometimes the disagreement only occurs after 1200 days, by which time all the rats are dead. The time and incidence value at which this occurs depends greatly on the parameter values and the radiation profile. Before plotting the curves, it is difficult to make a judgement on how long the approximate solution will remain valid. Moolgavkar et al.[4] find that the approximate and exact solutions agree for tumor incidences up to 60%. For our parameter values, we find that they disagree much earlier. On the whole, disagreement is strongest when the animal experiences chronic exposure. In Figures 18 to 25 are graphs comparing the exact and approximate curves for different radon exposure patterns - no exposure, chronic low exposure, medium exposure and high exposure. The parameter values used are from the optimal fit using likelihood 3 and model 1.

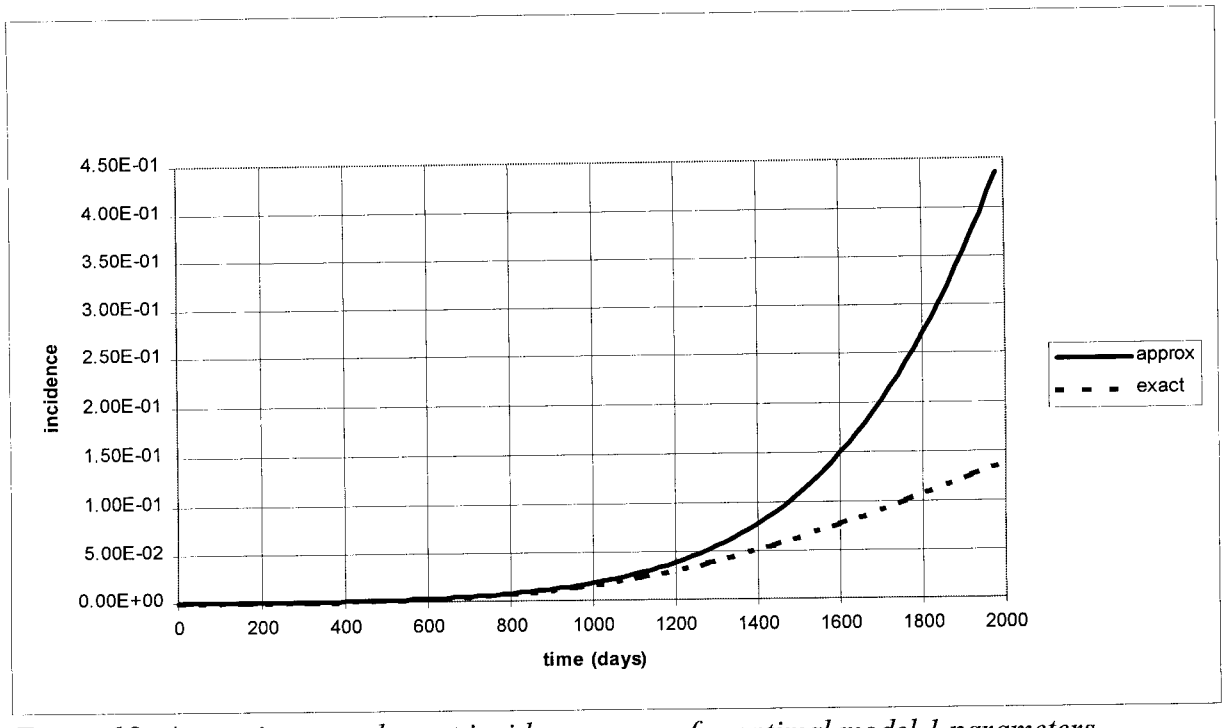

Figure 18: Approximate and exact incidence curves for optimal model 1 parameters, likelihood 3. No radon dose.

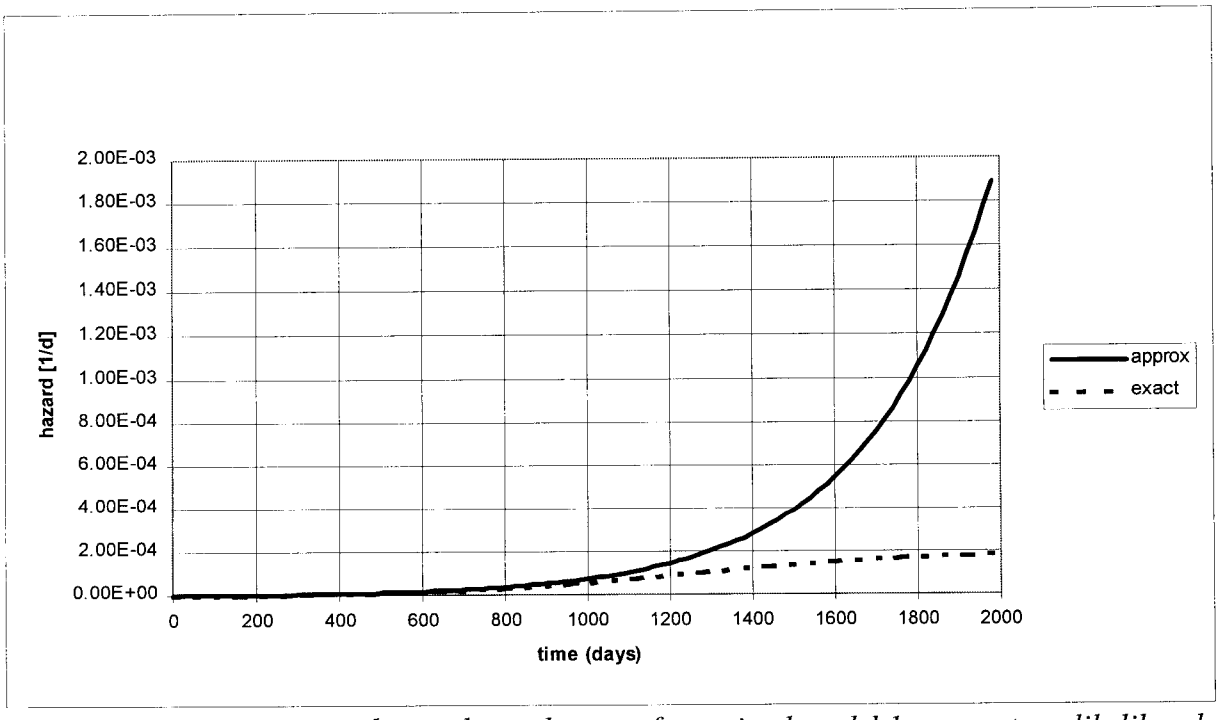

Figure 19: Approximate and exact hazard curves for optimal model 1 parameters, likelihood 3. No radon dose.

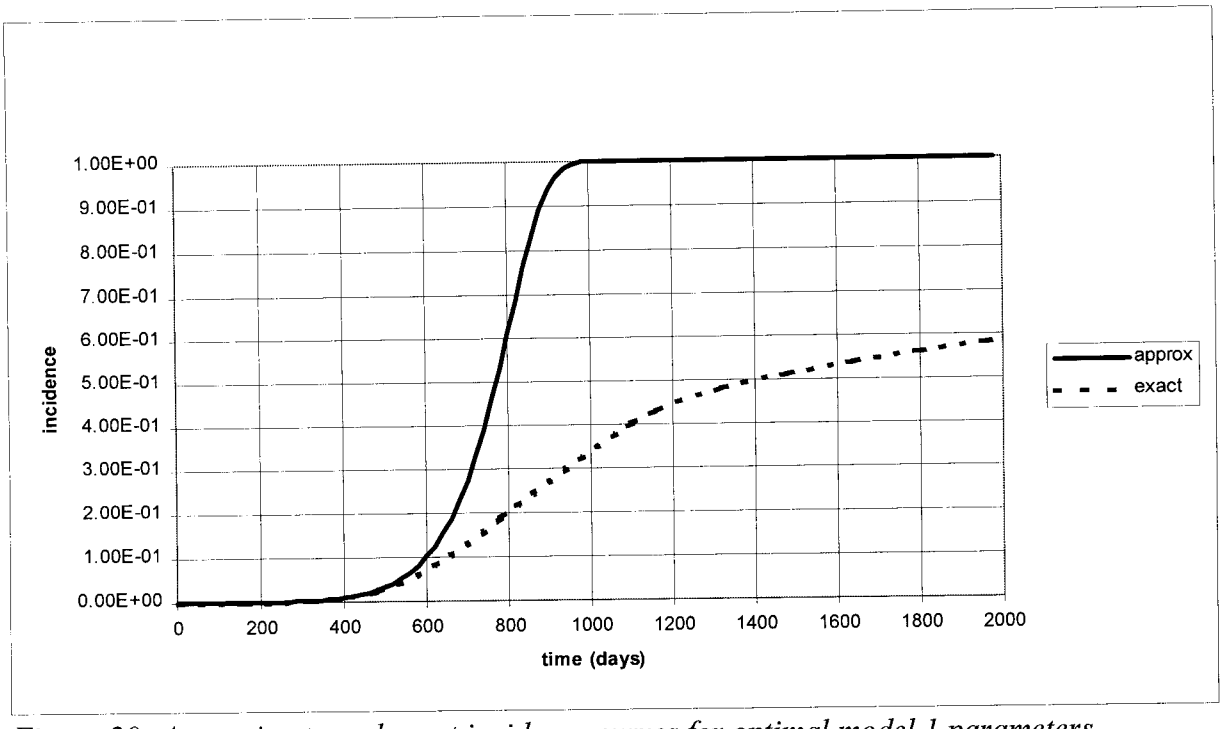

Figure 20: Approximate and exact incidence curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 0.5 WLM/d, starting day 1.

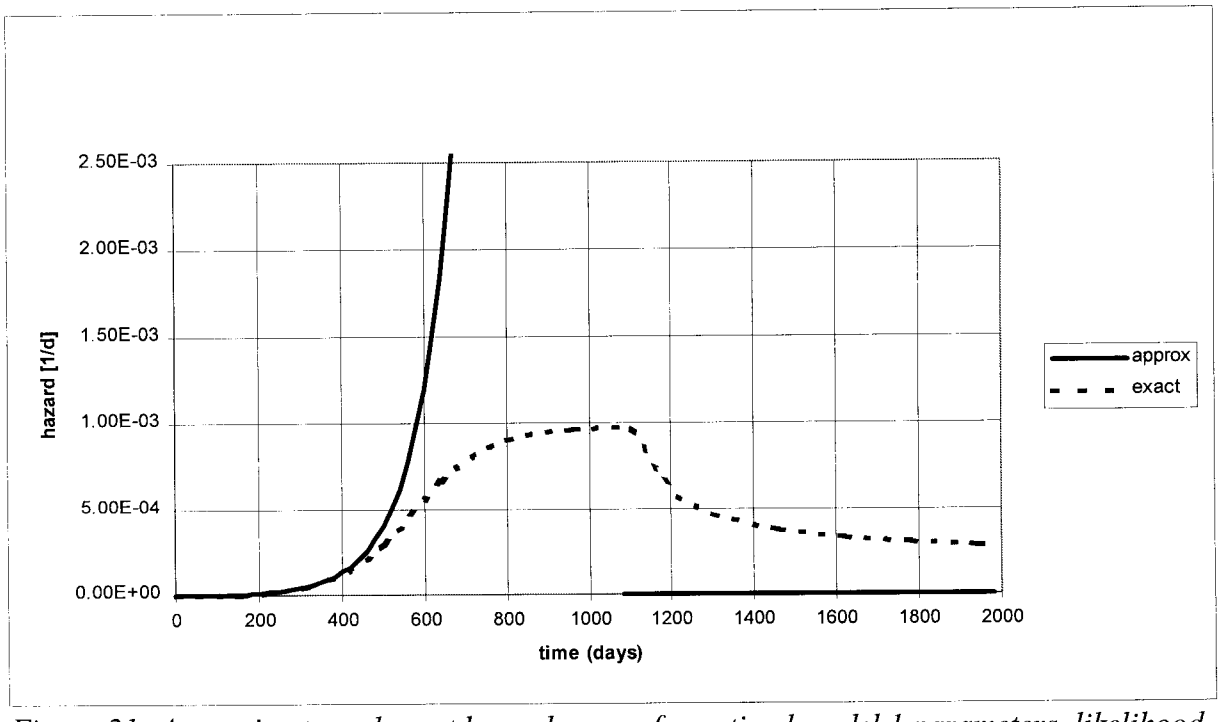

Figure 21: Approximate and exact hazard curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 0.5 WLM/d, starting day 1.

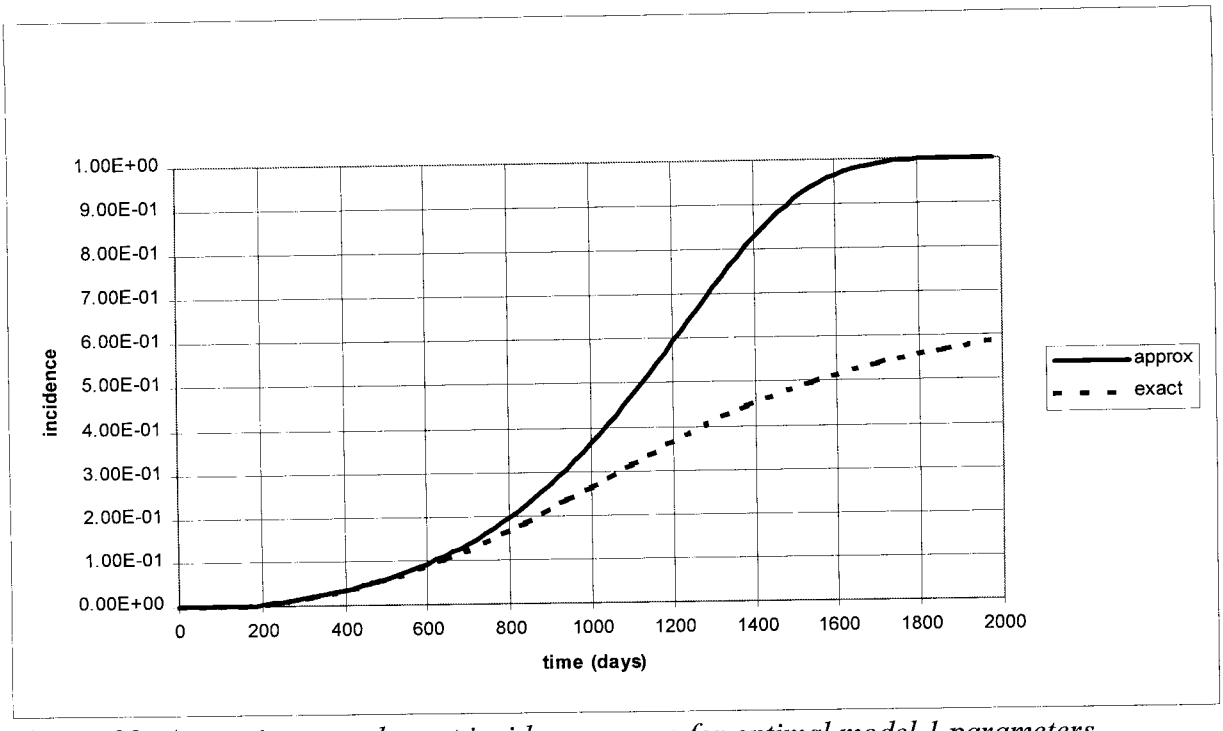

Figure 22: Approximate and exact incidence curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 5.0 WLM/d, starting day 1.

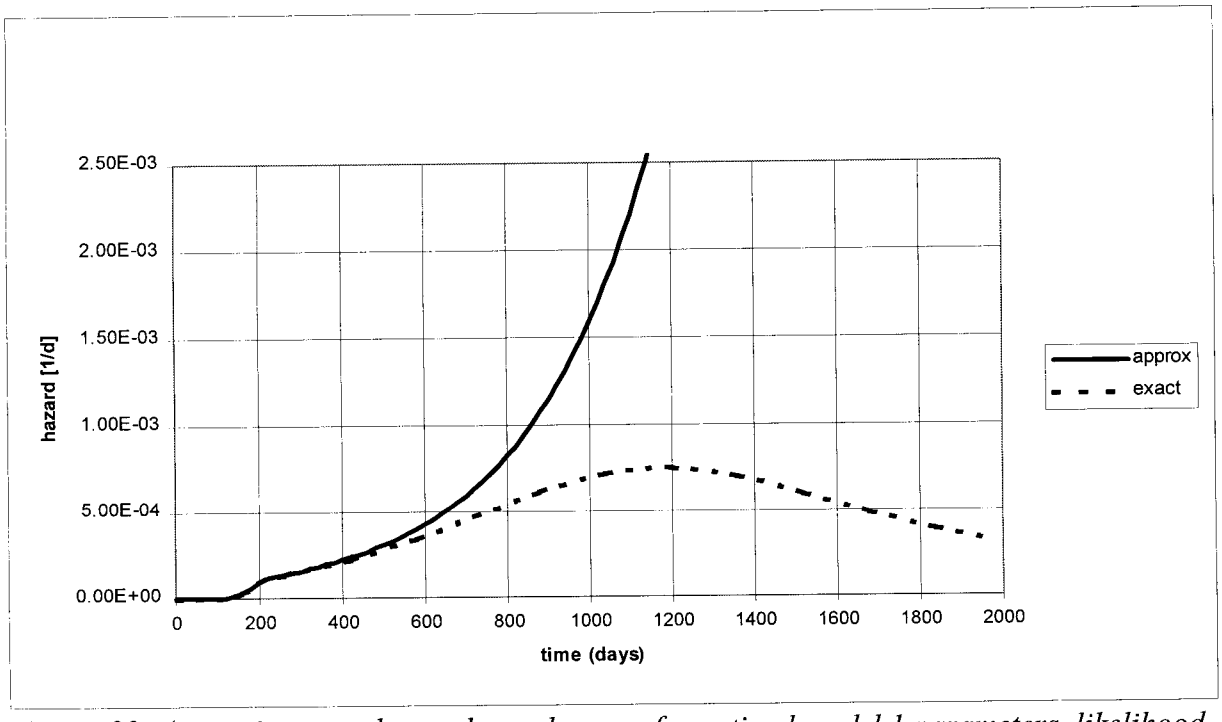

Figure 23: Approximate and exact hazard curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 5.0 WLM/d, starting day 1.

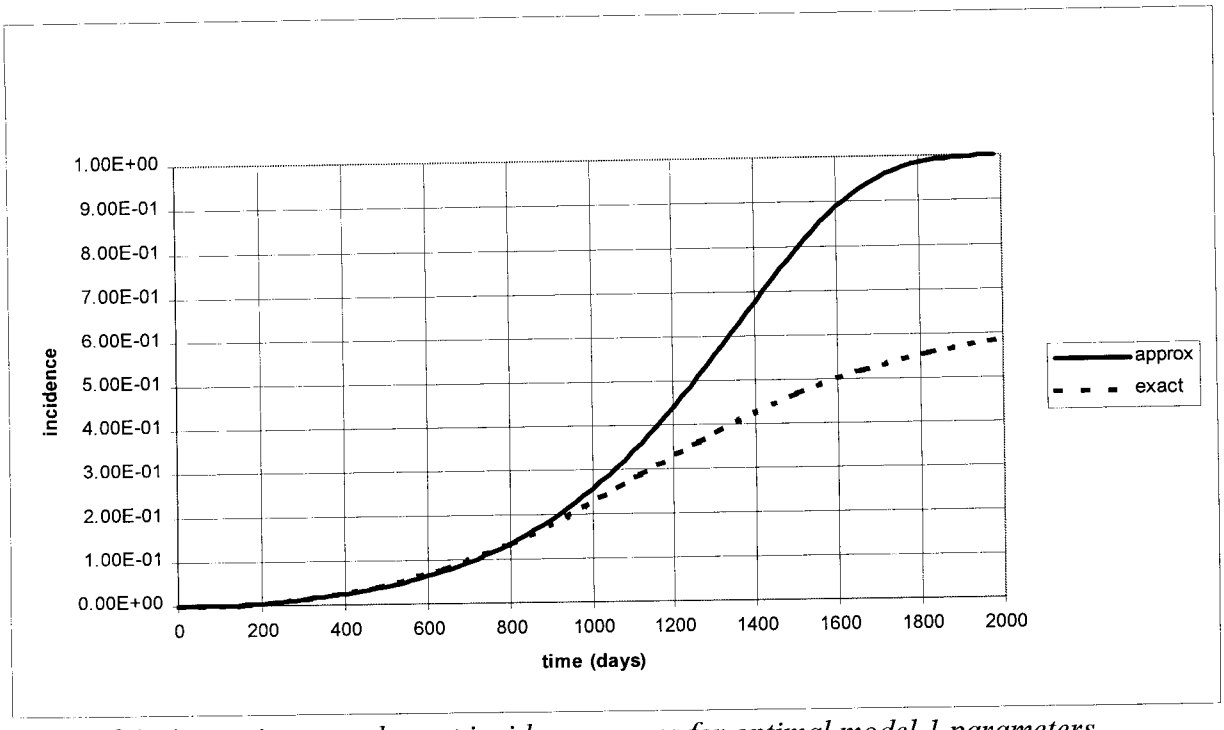

Figure 24: Approximate and exact incidence curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 50 WLM/d, starting day 1.

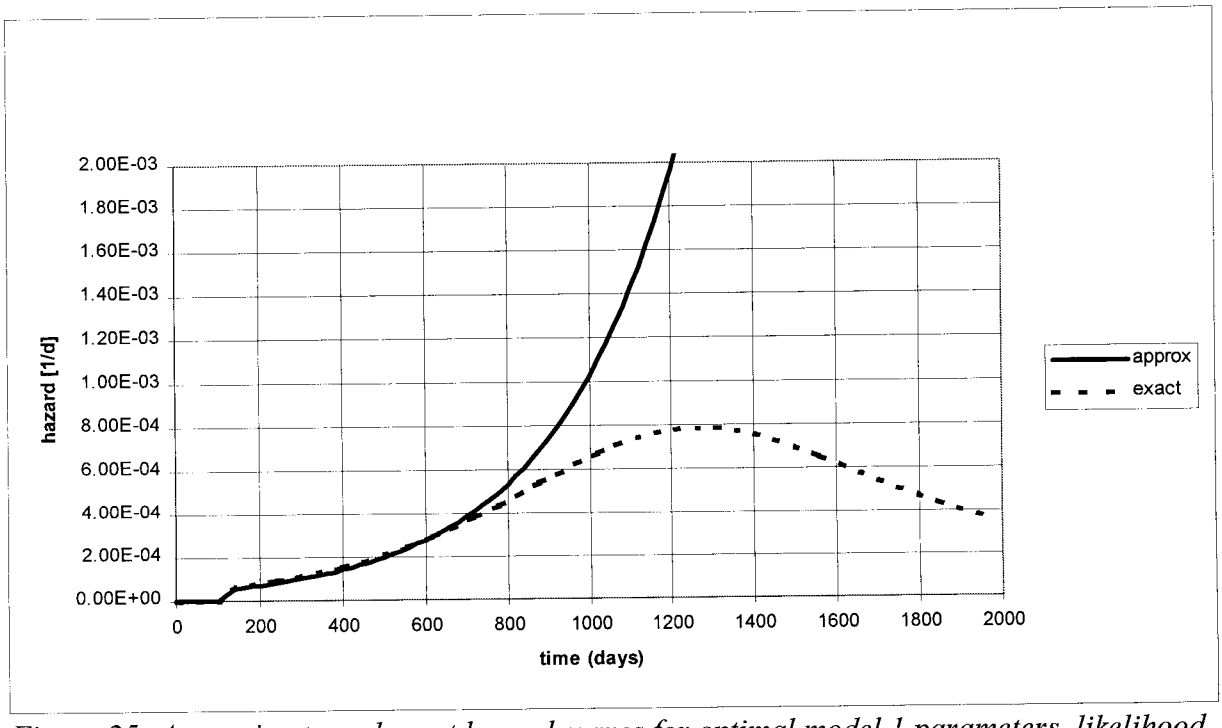

Figure 25: Approximate and exact hazard curves for optimal model 1 parameters, likelihood 3. Total dose 500 WLM, rate 50 WLM/d, starting day 1.

7 CONCLUSIONS

In this report, the two stage model for carcinogenesis is used in studying lung tumour development in rats due to radon and uranium ore dust exposure. Several aspects of this model are considered in this report.

1) Different biologically motivated dose-mutation and growth rate dependencies are examined, which describe the cellular effect of radon and uranium dust on the two stage carcinogenesis process. Fitting the model parameters to the data leads to a good description of the experimental results. A few general facts can be extracted from the results.
   - The second mutation rate seems to be independent of the exposure rate. The parameter α_2 is near zero, and is much smaller than α_1.
   - The intermediate cell self multiplication rate, and death rate seem to be dose dependent, and both increase with increasing dose.
   - The best intermediate cell kinetics seems to be described by model 1, where the self multiplication rate, ε, and the intermediate cell death rate, ν, are kept in the same ratio, and the overall intermediate cell growth rate $\delta=\varepsilon-\nu$ has a particular value for zero dosage rate, a higher value for non-zero small dosage rates, and increases slowly from this value for increasing dosage rates.
2) The tumour incidence has several different interpretations depending on how one incorporates the fatality of tumours into the model. These different interpretations result in different 'optimal' fits, and, thus, different tumour incidence predictions for other dose profiles not covered by the data. For all different interpretations, a particular model (model 1) was found to describe the data best.
3) The model has many parameters, and can exhibit many different behaviours for different parameter sets. The data can conceivably be equally well described by very different parameter sets, which give different incidence predictions for dosage profiles not covered by the data. We have found one example of when two different parameter sets give fits of similar quality as discussed in section 6.5. Thus, when making risk predictions for exposure profiles for which no data exists, one must take care that all possible 'good' models give similar predictions. In the example given in section 6.5, the two models give similar incidence curves for chronic low exposures.
4) The approximation to the model used in many papers[7,8] is examined for the models and parameter values from this report, and it was found the exact and approximate solutions can differ greatly in many realistic situations when the probability of tumour is no longer low.

Analysing the rat data using the two stage model gives insight to the working of the model, which can be used when working with human epidemiological data. We must recognise, however, that lung tumours in humans behave differently to those in rats. Rats, unlike humans, have a low spontaneous lung tumour incidence rate. Also, unlike in humans, lung

tumours may be present in rats without causing the death of the animal. As for in the rat analysis, care must be taken when using the model. Different parameter values or dose-rate models that describe the available data equally well may give different estimates for radiation induced tumour risks at low doses. The rat data was, however, described well by the model. The two stage carcinogenesis model is certainly worthy of receiving further attention.

REFERENCES

[1] E.S. Gilbert, F.T. Cross and G.E. Dagle,
Analysis of Lung Tumour Risks in Rats Exposed to Radon.
Radiat. Res. **145**, 350-360 (1996)

[2] S.H. Moolgavkar, S.H. Cross, G. Lübeck and G.E. Dagle,
A Two-Mutation Model for Radon-Induced Lung Tumours in Rats.
Radiat. Res. **121**, 28-37 (1990)

[3] E.G. Lübeck, S.B. Curtis, F.T. Cross and S.H. Moolgavkar,
Two-Stage Model of Radon-Induced Malignant Lung Tumours in Rats: Effects of Cell Killing.
Radiat. Res. **145**, 163-173 (1996)

[4] S.H. Moolgavkar, A. Dewanji and D.J. Venzon,
A Stochastic Two-Stage Model for Cancer Risk Assessment. I. The Hazard Function and the Probability of Tumour.
Risk Analysis **8**, 383-392 (1988)

[5] S.H. Moolgavkar and G. Lübeck,
Two-Event Model for Carcinogenesis: Biological, Mathematical and Statistical Considerations.
Risk Analysis **10**, 323-341 (1990)

[6] F.T. Cross,
Radiation Inhalation Studies in Animals,
Report DOE/FR-0396, U.S. Department of Energy,
Washington DC, (1988)

[7] L.B. Venema and H.P Leenhouts,
A Two-Mutation Model of Carcinogenesis: Application to ^{226}Ra Induced Osteosarcoma Prevalence in Beagles.
RIVM Report number 610065001 (1994)

[8] H.P. Leenhouts and K.H. Chadwick,
A Two-Mutation Model of Carcinogenesis: Analysis of Radiation Induced Lung Tumours in Animals and Implications for Risk Evaluation.
RIVM Report 749251001 (1993)

[9] S.H. Moolgavkar and D.J. Venzon,

Two-Event Models for Carcinogenesis: Incidence Curves for Childhood and Adult Tumours.
Mathematical Biosciences **47**, 55-77 (1979)

[10] I. Ingber,
Very Fast Simulated Reannealing.
Mathl. Comput. Modelling, **12** 967-973 (1989)

APPENDIX: EXACT INCIDENCE CALCULATIONS

Our aim is to obtain an expression for $P^0(t)$ as a function of the mutation, multiplication and death rates $\mu_1(t)$, $\mu_2(t)$, $\varepsilon(t)$ and $\nu(t)$. We do this using the *generating function* $\psi(y,z;t) := \sum_{i,m} P_{i,m}(t) y^i z^m$. By multiplying (1) by $y^i z^m$ and summing over i and m we obtain

$$\frac{\partial \psi}{\partial t} = (y-1)\mu_1 S\psi + (\mu_2 yz + \varepsilon y^2 + \nu - \varepsilon y - \nu y - \mu_2 y)\frac{\partial \psi}{\partial y}. \tag{3}$$

We are interested in $P(M(t)=0) = \psi(1,0;t)$. We can solve (3) to obtain this using the method of characteristics. The partial differential equation (3) can be transformed into an ordinary differential equation on a curve *(y(u),z(u),t(u))* (called a *characteristic*) parameterised by u. The differential equations describing the curve are

$$\frac{dy}{du} = -(\mu_2 yz + \varepsilon y^2 + \nu - \varepsilon y - \nu y - \mu_2 y), \frac{dz}{du} = 0, \frac{dt}{du} = 1. \tag{4}$$

(These are the coefficients of $\frac{\partial \psi}{\partial y}$, $\frac{\partial \psi}{\partial z}$ and $\frac{\partial \psi}{\partial t}$ respectively.) There is a whole family of such characteristics satisfying (4). A particular characteristic can be pinpointed by defining a point that lies on it. Now, ψ satisfies the following ordinary differential equation along each of the characteristics described above:

$$\frac{d\psi}{du} = (y-1)\mu_1 S\psi. \tag{5}$$

(The right hand side is just the coefficient of ψ in (3).) Note as $dt/du = 1$, we can take $t=u$. As $dz/du = 0$, z remains constant on these curves. We wish to find $\psi(1,0;t)$ and so we want to find a characteristic that passes through the point *(1,0;t)*. Thus, we require the characteristic *(y(u),z(u);t(u))* for which there exists a u_0 such that $(y(u_0),z(u_0);t(u_0))=(1,0;t)$. As $t(u)=u$, we therefore require the characteristic with the boundary conditions that $z(u)=0$ for all u, and $y(t)=1$. We write this characteristic $y(u,t)$. Like all characteristics, it is parameterised by u, and the t given in the argument determines the particular characteristic we are examining. Now, from (5):

$$P^0(t) = \psi(1,0;t) = \exp\int_0^t (y(u,t)-1)\mu_1(u)S(u)du.$$

This integration can be done numerically given the function $y(u,t)$. The function $y(u,t)$ can be evaluated exactly in the case where the rates μ_2, ε and ν are piecewise constant. (This is the case in the data we consider in this paper.) Let $0 = t_0 < t_1 < \ldots < t_k = t$, and suppose the parameters take on values μ_{2i}, ε_i and ν_i on the interval (t_{i-1},t_i). Suppose further that A_i and B_i are the two roots of the polynomial $\varepsilon_i x^2 - [\varepsilon_i + \nu_i + \mu_{2i}]x + \nu_i$. It is easy to show that $0 < A_i < 1 < B_i$. An expression for $y(u,t)$ can be found successively on each subinterval,

starting with the subinterval (t_{k-1},t_k), with initial condition $y(t_k,t_k)=1$. (This is the condition that $y(t)=1$.) The value $y(t_{k-1},t_k)$ found from the solution on this subinterval is used as a boundary condition for the subsequent subinterval - $y(t_{k-1},t_{k-1})=y(t_{k-1},t_k)$. The solution $y(u,t)$ is thus inductively built up. For $u \in (t_{i-1},t_i)$, we have

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

with $y(t_k,t)=1$.