Hiroshima J. Med. Sci.
Vol. 54, No. 1, 9~20, March, 2005
HIJM 54–2

Multi-target Models and their Application to Data Analysis of Cellular Mortality due to Radiation Exposure

I Made ARCANA[1] and Megu OHTAKI[2]

1) Graduate School of Biomedical Sciences, Hiroshima University, 1–2–3 Kasumi, Minami-ku, Hiroshima 734–8551, Japan

2) Department of Environmetrics and Biometrics, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1–2–3 Kasumi, Minami-ku, Hiroshima 734–8551

ABSTRACT

We consider multi-target models for use in analyzing data of the dose-response relationship. The target sizes we are concerned with here are both homogeneous, as assumed in the classical model, and heterogeneous, as simplified using geometric progression. We apply two models for establishing the multi-target models: a Poisson regression model constructed by assuming that the response variable Y follows Poisson distribution, and a gamma-frailty model as a Poisson mixture model derived by adding random common risks having a gamma distribution. Applying these models to experimental data relating the effects of miso fermentation-stages on the survival rate of cells of intestinal crypts of mice exposed to radiation yielded the result that there were substantial frailties associated with all miso fermentation-stages. Short-term and medium-term fermented miso provided similar effects, whereas long-term fermentation had the lowest relative risk value, indicating a significant protection of the crypts against exposure effects. A gamma-frailty model based on heterogeneous target size was more suitably applied when there were at least 3 dead stem cells having 10 target genes.

Key words: *Gamma-frailty model, Miso (fermented soy bean paste), Poisson regression model, Radio protective effects*

For more than fifty years, the multi-stage model proposed by Armitage and Doll[1] has continued to influence biomedical thinking on cellular changes, particularly in regard to the processes underlying carcinogenesis (e.g., see 2, 15, 17, 21, and 23). This model, however, is derived according to a probabilistic mechanism that is precisely described on the basis of reasonable assumptions[23], of which correspondence to actual events in biomedical fields has not been established. Therefore, Moolgavkar[15] noted that the multi-stage model needs to be embellished in various ways to accommodate our current thinking on carcinogenesis.

Building on the idea of removing stages from the multi-stage model of cell changes, the purpose of the present paper is to establish mathematical models for examining the relation of the dose response level of exposure to cellular changes on the pathway to cell death. These dose-response-based models, called multi-target models, consider a unit in exposed cells as a target.

Survival curves for most mammalian cells exposed to low-LET radiation, such as gamma and X-rays, show a shoulder-shaped curvature, in which there is less cell inactivation per unit dose at the initial low dose region and a tendency towards a constant slope at the higher dose. This constant final slope is caused by the effect of the repair of DNA single strand breaks during exposure. The shoulder region of the curve can be interpreted in two possible ways. Firstly, the given dose is considered as a total of dose fractions that are individually capable of repairing sub-lethal damage in between them[15], but become lethal damage when added together. In this situation, we assume that each dose fraction is given acutely and that the repair of the single strand breaks during the radiation can be ignored. Secondly, lesions are individually reparable, but when the efficiency of the enzymatic repair mechanisms diminishes due to the number of lesions, they are become irreparable and kill the cell. This means that it requires more than two targets getting exposure to radiation on the pathway to cell death. For these reasons, we applied multi-target models to cellular mortality due to radiation exposure.

Suppose that the probability of a target surviving after exposure at dose D is expressed by the survival function $S(D\,|\,\beta)=e^{-\beta D}$, where β is an unknown parameter describing the coefficient of exposure effects. Under such a condition, the probability of a target having vital damage is denoted by the failure distribution[12] $F(D\,|\,\beta)=1-e^{-\beta D}$. In the classical multi-target model, it is assumed that

lethal damage occurs after independent hits on a certain number of targets (say k), and that the target size is homogeneous. Here, the target size is related to its sensitivity level to radiation effects. The larger the target sizes the higher probability of it getting hit when exposed to radiation. The target survivor function is thus

$$S_k(D|\beta) = 1 - (1 - e^{-\beta D})^k. \quad (1)$$

However, the assumption of homogeneous target size is sometimes unrealistic. It is therefore appropriate to remove this assumption. A general extension of the target survivor function shown in equation (1) that allows each hit target to have a different exposure effect can be formulated by

$$S_k(D|\beta_1, \ldots, \beta_k) = 1 - \prod_{j=1}^{k}(1 - e^{-\beta_j D}), \quad (2)$$

where (unknown) parameter β_j represents the coefficient of sensitivity for the j-th target, j=1, 2, ... , k.

For establishing the multi-target models, we apply two models: a Poisson regression model constructed by assuming that the response variable Y follows Poisson distribution, and a gamma-frailty model as a Poisson mixture model derived by adding random common risks having a gamma distribution.

MULTI-TARGET MODELS

1. Geometric Structure on Heterogeneous Targets

By assuming that the parameter of the sensitivity coefficient for the j-th target (β_j) in classical theorem has regularity following geometrical progression, we constructed one of the simplest models for heterogeneity. That is, assume that $\beta_j=\beta\rho^{j-1}$, where the unknown parameter ρ describes the index of heterogeneity of the target size. Then, the survival function in equation (2) can be specified as

$$S_k(D|\beta, \rho) = 1 - \prod_{j=1}^{k}(1 - e^{-\beta\rho^{j-1}D}). \quad (3)$$

From a graphical point of view, in the case that the number of targets k=10 for a given exposure dose (D), the heterogeneity index value of ρ>1 provides significant decreases in the survival curve as compared with that of ρ=1, as shown in Fig. 1. Continued increases in the value of ρ may give accelerated increases in the hazard rate.

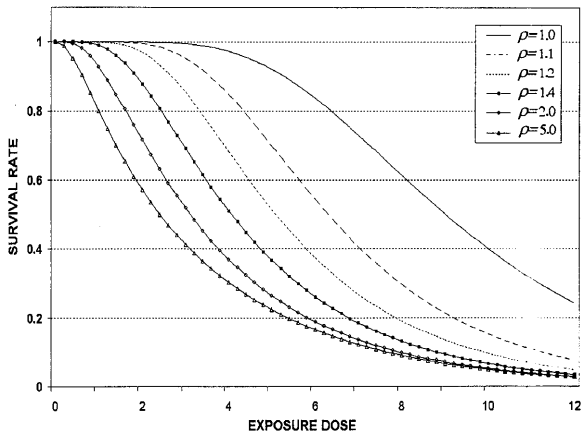

Fig. 1. Survival curves of the multi-target models at various values of ρ and given exposure dose (D), in a case where the number of targets k is equal to 10 and where the coefficient of exposure effects (β) is fixed. Significant decreases in the survival curve can be seen when the index value of ρ is greater than 1. A steady increase in the value of ρ results in an accelerated increase in the hazard rate.

2. Poisson Regression Model

Let Y be a response variable following Poisson distribution with mean $\mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)=\mu_0 S_k(De^{\gamma^T\boldsymbol{x}}|\beta, \rho)$, a function of exposure dose (D) for given covariates vector $\boldsymbol{x}$ and unknown parameters vector $\boldsymbol{\theta}^*=(k, \mu_0, \beta, \rho, \boldsymbol{\gamma}^T)^T$, where μ_0 denotes the baseline mean parameter and $S_k(De^{\gamma^T\boldsymbol{x}}|\beta, \rho)$ expresses the survival function, which is specified as

$$S_k(De^{\gamma^T\boldsymbol{x}}|\beta, \rho) = 1 - \prod_{j=1}^{k}(1 - e^{-\beta\rho^{j-1}De^{\gamma^T\boldsymbol{x}}}), \quad (4)$$

where $\gamma^T\boldsymbol{x} = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_p x_p$ is a linear combination of p covariates. Then,

$$P(y|D, \boldsymbol{x}, \boldsymbol{\theta}^*) = \frac{\{\mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)\}^y}{y!} e^{-\mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}. \quad (5)$$

Given a set of n independent samples $(y_i, D_i, \boldsymbol{x}_i), i$=1,2, ... , n, where y_i is the observed response, D_i is the exposure dose and $\boldsymbol{x}_i=(x_{i1}, x_{i2}, \ldots, x_{ip})^T$ is the covariates vector for the ith individual, we denote the actually observed data set by $\boldsymbol{d}_{(obs)}=(y, D, \boldsymbol{x}^T)^T$. Then, the likelihood function for estimating the unknown parameters based on the observed data set can be specified by

$$L(\boldsymbol{\theta}^*|\boldsymbol{d}_{(obs)}) = \prod_{i=1}^{n} P(y_i|D_i, \boldsymbol{x}_i, \boldsymbol{\theta}^*). \quad (6)$$

3. Gamma-frailty Model for Heterogeneous Background

It is important to take account of heterogeneity between individuals in population-based survival studies[9]. A systematic way of describing heterogeneity is by entering an unobserved quantity called frailty, here denoted by the letter Z. This quantity describes common risk factors, measurable or non-measurable, and is not included in the model[9,10,22].

Till now, most studies have used a frailty having a gamma distribution, which is mathematically convenient. Gamma distributions have been used for many years to generate a Poisson mixture model. From a computational point of view, they fit

very well with survival models, because it is easy to derive the formulas for any number of events[9]. For finite mean frailty distributions, it is required that the mean of the frailty be unity in order for the parameters of the model to be identifiable[10]. Furthermore, regarding the heterogeneous population, Hougaard[8] has examined the consequences of the difference between gamma distribution and inverse Gaussian distribution as the distribution of frailties, and remarked that the inverse Gaussian makes the population homogeneous with time, whereas for the gamma the relative heterogeneity is constant. For these reasons, we adopt the gamma distribution as the distribution of frailties.

Assume that the mean of the survival rate at given exposure dose D is

$$\mu_k(D, \boldsymbol{x}|Z, \boldsymbol{\theta}^*) = Z\mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*),$$

where Z denotes a random variable having gamma distribution with mean unity and variance σ (unknown). Then, the density function of Z can be described as

$$\varphi(z|\sigma) = \frac{\sigma^{-\sigma^{-1}}}{\Gamma(\sigma^{-1})} z^{\sigma^{-1}-1} e^{-\sigma^{-1}z}. \tag{7}$$

Let $\boldsymbol{d}=(y,D,\boldsymbol{x}^T,z)^T$ be the complete data set including the unobserved frailty term Z. The likelihood function can be formulated for a given complete data set $\boldsymbol{d}$ as

$$L(\boldsymbol{\theta}^*|\boldsymbol{d}) = \prod_{i=1}^{n} P(y_i|z_i, D_i, \boldsymbol{x}_i, \boldsymbol{\theta}^*), \tag{8}$$

where $P(y|z, D, \boldsymbol{x}, \boldsymbol{\theta}^*)$ denotes the probability density function of Poisson distribution with mean $\mu_k(D, \boldsymbol{x}|z, \boldsymbol{\theta}^*)$. Thus, the likelihood function based on the observed data set $\boldsymbol{d}_{(obs)}$ excluding the frailty term Z is obtained by integrating the likelihood function in equation (8) with respect to the density function of the frailty term of the ith individual, z_i. And, we have

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}) &= \int_0^\infty L(\boldsymbol{\theta}^*|\boldsymbol{d})\ \varphi(z_i|\sigma)\ dz_i \\ &= \prod_{i=1}^{n} \int_0^\infty P(y_i|z_i, D_i, \boldsymbol{x}_i, \boldsymbol{\theta}^*)\ \varphi(z_i|\sigma)\ dz_i \\ &= \prod_{i=1}^{n} f(y_i|D_i, \boldsymbol{x}_i, \boldsymbol{\theta}), \end{aligned} \tag{9}$$

where $f(y|D, \boldsymbol{x}, \boldsymbol{\theta})$ denotes the density function of negative binomial distribution with parameters vector $\boldsymbol{\theta} = (\boldsymbol{\theta}^{*T}, \sigma)^T$ expressed by

$$f(y|D, \boldsymbol{x}, \boldsymbol{\theta}) = \frac{\prod_{j=1}^{y}\{1+\sigma(j-1)\}}{y!} \left\{\frac{\mu(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{1+\sigma\mu(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}\right\}^y \{1+\sigma\mu(D, \boldsymbol{x}|\boldsymbol{\theta}^*)\}^{-\sigma^{-1}} \tag{10}$$

PARAMETER ESTIMATION

The maximum likelihood estimation method based on the log-likelihood function on the observed data set $\boldsymbol{d}_{(obs)}$ is applied for estimating unknown parameters $\boldsymbol{\theta}$. The function can be written as

$$\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}) = \log L(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}). \tag{11}$$

In many cases, an analytical method is not available for maximizing the function. Therefore, the maximization must be performed using a numerical method, often of an iterative character. The Newton-Raphson method, with its combination of simplicity and power, is the most widely used, although in general we know very little about its global convergence properties[11]. The method often becomes impractical in problems involving many parameters.

Ohtaki & Izumi[18], therefore, have proposed an algorithm called SPIDER for optimization without derivatives of the function. For the p-dimensional function, this alternative technique has iterative maximization procedures with cyclic fixing of groups of parameters, maximizing over the remaining parameters. (*The steps of the algorithm are presented in Appendix C*)

According to the general asymptotic theory, the maximum likelihood estimator has many useful properties, including consistency and sufficiency. The ability to achieve the Cramer-Rao minimum variance asymptotically is another remarkable property of the estimator. Under the regularity conditions, the vector of maximum likelihood estimators of $\boldsymbol{\theta}$ denoted by $\hat{\boldsymbol{\theta}}$ is best asymptotically normal (BAN) if $^\forall\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then $\hat{\boldsymbol{\theta}}$ is the approximation to the normal distribution with mean $\boldsymbol{\theta}$ and variance-covariance matrix $\frac{1}{n}I_1(\boldsymbol{\theta})^{-1}$ as n goes to infinity[14], or more explicitly we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, I_1(\boldsymbol{\theta})^{-1}),$$

where $I_1(\boldsymbol{\theta})$ is the Fisher information matrix of sample size 1. Furthermore, the Fisher information matrix of sample size n, that is $I_n(\boldsymbol{\theta})=nI_1(\boldsymbol{\theta})$, is given by the symmetric matrix expressed as a negative form of expectation of the Hessian matrix whose ij-th element is specified by

$$I_n(\boldsymbol{\theta})_{i,j} = -E\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\theta_i\, \partial\theta_j}\right].$$

Moreover, inverting the form of the information matrix yields a matrix containing the variances of the parameters on its diagonal and the asymptotic covariance in the off-diagonal positions. The Hessian matrix elements of the models are described in detail in Appendix B.

APPLICATION TO REAL DATA ANALYSIS

1. Data Set

As an example of an application of multi-target models in the biomedical field, we will attempt to analyze experimental data on the density of the small intestinal crypt of mice after exposure to gamma rays. The aim of the experiment conducted by Ohara et al[16] was to verify the effect of giving a diet supplemented with miso (Japanese fermented soy bean paste) at various fermentation stages on crypt survival.

For this experiment, the mice were fed a commercial diet MF alone or a diet supplemented with miso for one week before the exposure. The miso had been fermented for a short-term (immediate fermentation), medium-term (4 months) or long-term (6 months). Groups of mice (each 5 mice) were whole-body exposed to 7, 8, 10 or 12 Gy of X-rays without anaesthetization. The number of surviving crypts was counted in 10 gut cross sections in each mouse.

2. Model for Growth and Disappearance of Intestinal Crypt

Ohara et al[16] remarked that in the absence of surviving crypt stem cells, the crypts disappear. In both the large and small intestine, mutagen administration leads to the occurrence of isolated crypts that are completely populated by a mutated phenotype. Therefore, it has been proposed that crypts are maintained by a single stem cell.

On the other hand, the results of studies on the small intestine by Williams et al[24] lead them to question the previous assumption. They proposed an alternative hypothesis in regard to the number of stem cells required to maintain the crypts, and gave an explanation based on multiple crypt stem cells with random cell loss after stem cell division.

Consider that a crypt contains multiple stem cells, and let the (unknown) parameter be m. Suppose that all of the stem cells will disappear after k independent hits cause the crypt to cease growing. Then, for given exposure dose D and covariates vector $\boldsymbol{x}=(x_1,x_2,x_3)^T$, we can apply the survival function in equation (4) with a slight modification for the survivor crypt data:

$$S_{k,m}(De^{\gamma^T\boldsymbol{x}}|\beta,\rho)=1-\left\{\prod_{j=1}^{k}\left(1-e^{-\beta\rho^{j-1}De^{\gamma^T\boldsymbol{x}}}\right)\right\}^m, \quad (12)$$

where the covariates vector $\boldsymbol{x}$ is constructed by setting a dummy variable to account for the duration of fermentation:

$$x_1=\begin{cases}1, \text{ if "Early" (short-term fermentation),}\\ 0, \text{ otherwise}\end{cases}$$
$$x_2=\begin{cases}1, \text{ if "Medium" (medium-term fermentation),}\\ 0, \text{ otherwise}\end{cases}$$
$$x_3=\begin{cases}1, \text{ if "Long" (long-term fermentation),}\\ 0, \text{ otherwise}\end{cases}$$

3. Results

The results show that there are substantial frailties for all miso fermentation-stages. The Akaike Information Criterion (AIC) values as a fit-

Table 1. Estimated Parameter Values in the Non-Frailty Poisson Regression Model

A. Homogeneous multi-target model

Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
12	0.3163 (0.3092, 0.3233)	1.0	0.913	0.921	0.871	1253.12

B. Heterogeneous multi-target model with single stem cell assumption

Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
10	0.2609 (0.2602, 0.2617)	1.035	0.910	0.919	0.869	1261.14
20	0.2356 (0.2351, 0.2361)	1.069	0.912	0.921	0.871	1236.30
30	0.2357 (0.2354, 0.2361)	1.069	0.912	0.921	0.871	1235.28
40	0.2358 (0.2355, 0.2361)	1.069	0.912	0.921	0.871	1235.28

C. Heterogeneous multi-target model with multiple stem cell assumption

Number of stem cells	Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
2	10	0.2430 (0.2425, 0.2434)	1.144	0.912	0.921	0.871	1236.38
	15	0.2432 (0.2428, 0.2435)	1.143	0.912	0.921	0.871	1235.40
	20	0.2432 (0.2429, 0.2434)	1.144	0.912	0.921	0.871	1235.40
3	10	0.2505 (0.2501, 0.2508)	1.225	0.912	0.921	0.871	1235.61
	12	0.2504 (0.2501, 0.2507)	1.225	0.912	0.921	0.871	1235.60
4	8	0.2576 (0.2573, 0.2580)	1.314	0.912	0.921	0.871	1235.90
	10	0.2576 (0.2573, 0.2579)	1.314	0.912	0.921	0.871	1235.90

Note: Values in parentheses are the 95% confidence intervals

Table 2. Estimated Parameter Values in the Gamma-Frailty Model

A. Homogeneous multi-target model

Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
11	0.3088 (0.2937, 0.3240)	1.0	0.914	0.912	0.859	996.11

B. Heterogeneous multi-target model with single stem cell assumption

Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
10	0.2362 (0.2343, 0.2380)	1.067	0.910	0.912	0.856	996.29
20	0.2313 (0.2307, 0.2319)	1.075	0.912	0.912	0.859	993.06
30	0.2315 (0.2311, 0.2319)	1.075	0.913	0.912	0.859	992.98
40	0.215 (0.2312, 0.2318)	1.075	0.913	0.912	0.859	992.98

C. Heterogeneous multi-target model with multiple stem cell assumption

Number of stem cells	Number of targets	$\hat{\beta}$	$\hat{\rho}$	$\widehat{RR}_e$	$\widehat{RR}_m$	$\widehat{RR}_l$	*AIC*
2	10	0.2391 (0.2386, 0.2397)	1.157	0.912	0.912	0.859	993.08
	15	0.2393 (0.2389, 0.2397)	1.156	0.912	0.912	0.859	993.03
	20	0.2393 (0.2390, 0.2396)	1.156	0.912	0.912	0.859	993.03
3	8	0.2468 (0.2463, 0.2472)	1.248	0.912	0.912	0.859	993.04
	10	0.2468 (0.2465, 0.2471)	1.247	0.912	0.912	0.859	993.03
	12	0.2468 (0.2465, 0.2471)	1.247	0.912	0.912	0.859	993.03
4	10	0.2544 (0.2541, 0.2546)	1.349	0.912	0.912	0.859	993.08
	12	0.2543 (0.2540, 0.2546)	1.349	0.912	0.912	0.859	993.08

Note: $\hat{\sigma}^2$=0.006
Values in parentheses are the 95% confidence intervals

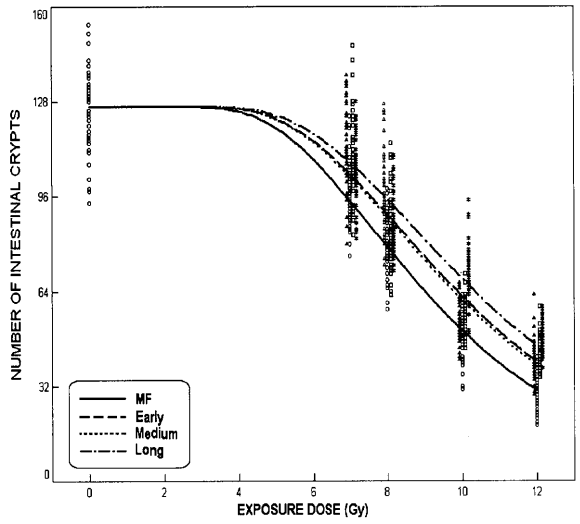

Fig. 2. Curve of the density of the intestinal crypt after an exposure event based on the Poisson regression model according to the fermented-stage of miso, with mice fed with a commercial diet of MF used as controls. The survival rate of crypts of mice fed long-term fermented miso has a higher rate indicated by the slope of the curve slightly decreasing as compared with the others. On the other hand, the short-term and medium-term fermentations confer almost the same level of protection on the crypts after exposure. In the scatter plot results for the mice exposed to 7, 8, 10 or 12 Gy of X-rays after being fed a commercial diet of MF marked by a circle or a diet supplemented with miso fermented for a short-, medium-, or long-term marked by a triangle, square, and asterisk respectively.

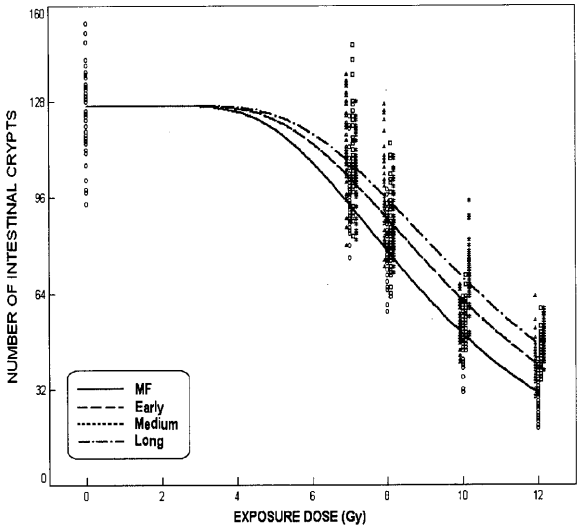

Fig. 3. Curve of the density of the intestinal crypt after an exposure event based on the gamma-frailty model according to the fermented-stage of miso, with mice fed a commercial MF diet used as controls. The survival curve of the crypts of mice fed long-term fermented miso has a slightly decreasing slope, indicating that the crypt-survival rate in this group was higher than in the other groups. On the other hand, the short-term and the medium-term fermentations confer exactly the same level of protection on the crypts after exposure. In the scatter plot results for the mice exposed to 7, 8, 10 or 12 Gy of X-rays after being fed a commercial diet of MF marked by a circle or a diet supplemented with miso fermented for a short-, medium-, or long-term marked by a triangle, square, and asterisk respectively.

ted model measurement were significantly lower when the gamma-frailty model was applied than when the Poisson regression model was used. For protecting the crypts after exposure, both the Poisson regression model and gamma-frailty model yielded similar results on short-term and medium-term fermented miso, as shown by the similar values of the relative risk corresponding to the fermentation terms $\widehat{RR}_e$ and $\widehat{RR}_m$, respectively. On the other hand, the relative risk values of the long-term group ($\widehat{RR}_l$) were a little lower than the others, indicating significant protection of the crypts against the exposure effects (see Table 1 and Table 2). Furthermore, from a graphical point of view, the survival curve of the long-term group has a slightly decreasing slope, which means that the rate of crypt survival of this group is higher than that of the other groups (see Fig. 2 and Fig. 3). Moreover, these results show that the gamma-frailty model based on assumed heterogeneity in the target size, as indicated by the values of the heterogeneity index, is more suitable for application to such empirical data, in which the number of targets was 30 genes and the AIC value was 992.98. Regarding the number of stem cells in the crypt, it was suggested that the fitted model could be obtained when m=3 and there were at least 10 genes, as indicated by the AIC value of 993.03.

DISCUSSION

Results of Data Analysis

As mentioned in the previous section, the results showed that the dose-incidence curves reached a plateau at about 3 dead cells per crypt section in the mouse small intestine. This result is close to the result reported by Hendry et al[6] of about 3 to 4 dead cells per crypt section. Furthermore, they pointed out that the production of apoptotic cells by low doses of gamma-rays was independent of the dose rate between 0.27 and 450 cGy per min. Moreover, Hendry and Potten[7] reported that the cells that die via apoptosis represent a very sensitive subpopulation of about 6 cells per crypt that may or may not be clonogenic. Fujikawa et al[5] similarly reported that 5.1 ± 0.3 somatic crossing-over mutations were induced by X-rays in *Drosophila melanogaster* and Takai et al[20] estimated that 4.3 ± 0.6 such mutations were induced by X-rays in medaka fish (*Oryzias latipes*).

Identifiability

Application of a gamma-frailty multi-target model to this experimental exposure data revealed that the survival curves flattened out when the number of targets was more than 10, as indicated by the relatively stable AIC values. This may indicate either that the model is less sensitive in identifying cell changes in more than 10 targets, or that the cell changes have no significant effect on the model. Furthermore, the index value ρ related to the survival rate indicated that as the exposure dose (D) approached infinity, the number of targets k does not affect the change of survival rate when the index value ρ is greater than 1. On the contrary, when ρ equals 1, the survival rate of k targets tends to be k times the survival rate of one target (*see Proposition 1 in Appendix A*).

Related Topics

There is a long history of attempts to establish a theoretical model of exposure-induced cell changes. The multi-stage model proposed by Armitage and Doll[1] based on the hypothesis of Fisher and Hollomon[4] has been used in biomedical fields for more than fifty years. This hypothesis assumed that carcinogenic transformation of cells in a tissue requires that independent changes occur in six or seven cells according to a specified form of relationship to age of the individual and for weighting concentration as a function of age in order to determine a hazard function. Thomas[21] remarked that the essence of this model is the peaked weighting function for exposure as a function of age, such that the later the sensitive stage of the model, the later the peak.

Currently, radiation exposures associated with human activity are expected to be low-dose, for example low dose-rate radiation from medical tests, waste cleanup and environmental isolation of materials associated with nuclear weapons and nuclear power production. An exposure-based event can cause a variety of damage scenarios: (1) the damage may be reparable if the damaged cells can repair themselves, and thus there will be no permanent damage; (2) millions of cells may die according to the natural processes of cell death; (3) mutations may occur if the damaged cells exhibit a change in their reproductive structure, resulting in potentially pre-cancerous cells. For such issues, in addition to the frailty model for heterogeneous background presented in this paper, we must consider a model of low-dose exposure based on risk factors describing heterogeneous sensitivity by assuming that each target before the exposure event contains random risk factors. We describe such a model in detail in Appendix B.3.

ACKNOWLEDGEMENTS

We thank Emeritus Professor Hiromitsu Watanabe for providing his unpublished data and Dr. Kenichi Satoh for his suggestions during the development of the models. We also thank the anonymous referees, whose comments were invaluable in the revision of this paper. This work was supported in part by a Grant in Aid (14380123) from the Japanese Ministry of Education, Science and Culture, and by a grant for a Research Program on Low-Dose Radiation Effects Based on Molecular Biology from the

Japan Atomic Energy Research Institute.

(Received November 5, 2004)
(Accepted January 31, 2005)

REFERENCES

1. **Armitage, P. and Doll, R.** 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br. J. Cancer **8:** 1–12.
2. **Dawson, S.V. and Alexeeff, G.V.** 2001. Multistage model estimates of lung cancer risk from exposure to diesel exhaust, based on a U.S. railroad worker cohort. Risk Analysis **21 (No. 1):** 1–18.
3. **Elkind, M.M. and Suton, H.** 1959. X-ray damage and recovery in mammalian cells in culture. Nature **184:** 1293–1295.
4. **Fisher, J.C. and Holloman, J.H.** 1953. A hypothesis for the origin of cancer foci. Cancer **7:** 916–918.
5. **Fujikawa, K., Hasegawa, Y., Matsuzawa, S., Fukunaga, A., Itoh, T. and Kondo, S.** 2000. Dose and dose-rate effects of X rays and fission neutrons on lymphocyte apoptosis in p53 (+/+) and p53 (–/–) mice. J. Radiat. Res. **41:** 113–127.
6. **Hendry, J.H., Potten, C.S., Cadwick, C. and Bianchi, M.** 1982. Cell death (apoptosis) in the mouse small intestine after low doses: effects of dose-rate, 14.7 MeV neutrons, and 600 MeV (maximum energy) neutrons. Int. J. Radiat. Biol. **42:** 611–620.
7. **Hendry, J.H. and Potten, C.S.** 1982. Intestinal cell radiosensitivity: a comparison for cell death assayed by apoptosis or by a loss of clonogenicity. Int. J. Radiat. Biol. **42:** 621–628.
8. **Hougaard, P.** 1984. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. Biometrika **71:** 75–83.
9. **Hougaard, P.** 2000. Analysis of multivariate survival data. Springer-Verlag, New York.
10. **Ibrahim, J., Chen, M.H. and Sinha, D.** 2001. Bayesian survival analysis. Springer-Verlag, New York.
11. **Jensen, S.T., Johansen, S. and Lauritzen, S.L.** 1991. Globally convergent algorithms for maximizing a likelihood function. Biometrika **78:** 867–877.
12. **Kleinbaum, D.G.** 1996. Survival analysis: a self-learning text. Springer-Verlag, New York.
13. **Lehmann, E.L.** 1983. Theory of point estimation. John Wiley & Sons. USA.
14. **Mood, A.M., Graybill, F.A. and Boes, D.C.** 1974. Introduction to the theory of statistics. McGraw-Hill. Singapore.
15. **Moolgavkar, S.H.** 2004. Commentary: Fifty years of the multistage model: remarks on a landmark paper. Int. J. Epidemiol. **33:** 7–8.
16. **Ohara, M., Lu, H., Shiraki, K., Ishimura, Y., Uesaka, T., Katoh, O. and Watanabe, H.** 2001. Radioprotective effects of miso (fermented soy bean paste) against radiation in B6C3F1 mice: increased small intestinal crypt survival, crypt lengths and prolongation of average time to death. Hiroshima J. Med. Sci. **50:** 83–86.
17. **Ohtaki, M., Fujita, S., Hayakawa, N., Kurihara, M. and Munaka, M.** 1985. The age distribution of human adult cancer and an initiation-manifestation model for carcinogenesis. Jpn. J. Clin. Oncol. **15 (Suppl. 1):** 325–343.
18. **Ohtaki, M. and Izumi, S.** 1999. Globally convergent algorithm without derivatives for maximizing a multivariate function. *In* Proceedings of Symposium on "Exploratory Methods and Analyses for Nonlinear Structures of Data with Random Variation" in Hiroshima.
19. **Sahu, S.K. and Dey, D.K.** 2000. A comparison of frailty and other models for bivariate survival data. Lifetime Data Anal. **6:** 207–228.
20. **Takai, A., Kagawa, N. and Fujikawa, K.** 2004. Dose- and time-dependent response for micronucleus induction by x-rays and fast neutrons in gill cells of medaka (*oryzias latipes*). Environ. Mol. Mutagen. **44:** 108–112.
21. **Thomas, D.C.** 1982. Temporal effects and interaction in cancer: Implications of carcinogenic models, p.107–121. *In* R. L. Prentice and A. S. Whittemore (eds.), Environmental epidemiology: Risk assessment, Philadelphia Society for Industrial and Applied Mathematics.
22. **Vaupel, J.W., Manton, K.G. and Stallard, E.** 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography **16:** 439–454.
23. **Whittemore, A.S.** 1977. The age distribution in human cancers for carcinogenic exposures of varying intensity. Am. J. Epidemiol. **106:** 418–432.
24. **Williams, E.D., Lowes, A.P., Williams, D. and Williams G.T.** 1992. A stem cell niche theory of intestinal crypt maintenance based on a study of somatic mutation in colonic mucosa. Am. J. Pathol. **141:** 773–776.

Appendix A

When the exposure dose (D) approaches infinity, the ratio of the survival rate of all targets to the survival rate of one target satisfies the following proposition.

Proposition 1. Let $\rho \geq 1$, and let

$$S_k(D|\beta,\rho) = 1 - \prod_{j=1}^{k}(1 - e^{-\beta\rho^{j-1}D})$$

for $k \geq 1$. Then, it holds that

$$\lim_{D\to+\infty} \frac{S_k(D|\beta,\rho)}{S_1(D|\beta,\rho)} = \begin{cases} k, & \text{if } \rho = 1, \\ 1, & \text{if } \rho > 1. \end{cases}$$

Proof:

If $\rho=1$, then

$$\begin{aligned}
\lim_{D\to+\infty} \frac{S_k(D|\beta,\rho)}{S_1(D|\beta,\rho)} &= \lim_{D\to+\infty} \frac{1-(1-e^{-\beta D})^k}{e^{-\beta D}} \\
&= \lim_{D\to+\infty} \frac{k(1-e^{-\beta D})^{k-1}(e^{-\beta D})(-\beta)}{(e^{-\beta D})(-\beta)} \\
&= \lim_{D\to+\infty} k(1-e^{-\beta D})^{k-1} \\
&= k.
\end{aligned}$$

If $\rho>1$, then

$$\begin{aligned}
\lim_{D\to+\infty} \frac{S_k(D|\beta,\rho)}{S_1(D|\beta,\rho)} &= \lim_{D\to+\infty} \frac{1-\prod_{j=1}^{k}(1-e^{-\beta\rho^{j-1}D})}{e^{-\beta D}} \\
&= \lim_{D\to+\infty} \frac{\sum_{j^*=1}^{k}\left[-\beta\rho^{j^*-1}e^{-\beta\rho^{j^*-1}D}\prod_{j\neq j^*}^{k}(1-e^{-\beta\rho^{j-1}D})\right]}{-\beta e^{-\beta D}} \\
&= \lim_{D\to+\infty} \frac{\sum_{j^*=1}^{k}\left[\rho^{j^*-1}e^{-\beta\rho^{j^*-1}D}\prod_{j\neq j^*}^{k}(1-e^{-\beta\rho^{j-1}D})\right]}{e^{-\beta D}} \\
&= \lim_{D\to+\infty} \left[\prod_{j=2}^{k}(1-e^{-\beta\rho^{j-1}D}) + \sum_{j^*=2}^{k}\left\{\rho^{j^*-1}e^{-\beta(\rho^{j^*-1}-1)D}\prod_{j\neq j^*}^{k}(1-e^{-\beta\rho^{j-1}D})\right\}\right] \\
&= 1.
\end{aligned}$$

Appendix B

B.1. Poisson Regression Model

The log-likelihood function of the model as shown in equation (6) is specified as

$$\ell(\boldsymbol{\theta}^*|\boldsymbol{d}_{(obs)}) = \sum_{i=1}^{n} \log P(y_i|D_i,\boldsymbol{x}_i,\boldsymbol{\theta}^*).$$

Then, elements of the Hessian matrix are given by

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta}^*|\boldsymbol{d}_{(obs)})}{\partial\theta_p^*} &= \sum_{i=1}^{n}\left\{\frac{y_i-\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}\right\}\left[\frac{\partial\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial\theta_p^*}\right], \\
\frac{\partial^2\ell(\boldsymbol{\theta}^*|\boldsymbol{d}_{(obs)})}{\partial\theta_p^*\,\theta_q^*} &= \sum_{i=1}^{n}\left\{-\frac{y_i\left[\frac{\partial\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial\theta_p^*}\right]\left[\frac{\partial\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial\theta_q^*}\right]}{[\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)]^2} + \frac{(y_i-\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*))\left[\frac{\partial^2\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial\theta_p^*\,\partial\theta_q^*}\right]}{\mu_k(D_i,\boldsymbol{x}_i|\boldsymbol{\theta}^*)}\right\}.
\end{aligned}$$

B.2. Gamma-frailty Model for Heterogeneous Background

The log-likelihood function of the model presented in equation (9) is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}) = \sum_{i=1}^{n} \log f(y_i|D_i, \boldsymbol{x}_i, \boldsymbol{\theta}).$$

Elements of the Hessian matrix are therefore specified as follows:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial \theta_p^*} &= \sum_{i=1}^{n} \frac{(y_i - \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_p^*}\right]}{\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial \theta_p^* \partial \theta_q^*} &= \sum_{i=1}^{n} \left\{ \frac{\sigma^2\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_p^*}\right]\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_q^*}\right]}{(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))^2} \right. \\
&\quad - \frac{(y_i(1+2\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)))\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_p^*}\right]\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_q^*}\right]}{\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)^2(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))^2} \\
&\quad \left. + \frac{(y_i - \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))\left[\frac{\partial^2 \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_p^* \partial \theta_q^*}\right]}{\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))} \right\}, \\
\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial \sigma} &= \sum_{i=1}^{n} \left\{ \frac{2\sigma y_i(1+\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))}{(-1+\sigma^2)(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))} + \frac{2}{\sigma^3}\log(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)) - \frac{2\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\sigma^3(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))} \right\}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial \sigma^2} &= \sum_{i=1}^{n} \left\{ \frac{2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)(3+5\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))}{(\sigma+\sigma^3\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))^2} - \frac{6\log(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))}{\sigma^4} \right. \\
&\quad \left. - \frac{2y_i(1+\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))(1+\sigma^2+\sigma^2(-1+3\sigma^2)\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))}{(-1+\sigma^2)^2(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))^2} \right\}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial \theta_p^* \partial \sigma} &= \sum_{i=1}^{n} \frac{2\sigma(-y_i + \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))\left[\frac{\partial \mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*)}{\partial \theta_p^*}\right]}{(1+\sigma^2\mu_k(D_i, \boldsymbol{x}_i|\boldsymbol{\theta}^*))^2}.
\end{aligned}$$

In the case of the model for a given exposure dose D, covariates vector $\boldsymbol{x}$=(x_1,x_2, ... ,x_p)^T, and unknown parameters $\boldsymbol{\theta}^* = (k, \mu_0, \beta, \rho, \boldsymbol{\gamma}^T)^T$ having a homogeneous target size expressed as

$$\mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*) = \mu_0 \left\{ 1 - \left(1 - e^{-\beta D e^{\boldsymbol{\gamma}^T \boldsymbol{x}}}\right)^k \right\},$$

we have

$$\begin{aligned}
\frac{\partial \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \mu_0} &= 1 - [F(D|\beta, \boldsymbol{\gamma})]^k, \\
\frac{\partial \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \beta} &= -\frac{\mu_0 D}{\beta} k f(D|\beta, \boldsymbol{\gamma})[F(D|\beta, \boldsymbol{\gamma})]^{k-1}, \\
\frac{\partial \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \gamma_p} &= -x_p \mu_0 D k f(D|\beta, \boldsymbol{\gamma})[F(D|\beta, \boldsymbol{\gamma})]^{k-1}, \\
\frac{\partial \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial k} &= -\mu_0 [F(D|\beta, \boldsymbol{\gamma})]^k \log[F(D|\beta, \boldsymbol{\gamma})], \\
\frac{\partial^2 \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \mu_0^2} &= 0, \\
\frac{\partial^2 \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \mu_0 \partial \beta} &= -\frac{k D}{\beta} f(D|\beta, \boldsymbol{\gamma})[F(D|\beta, \boldsymbol{\gamma})]^{k-1}, \\
\frac{\partial^2 \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \mu_0 \partial \gamma_p} &= -x_p\, k\, D\, f(D|\beta, \boldsymbol{\gamma})[F(D|\beta, \boldsymbol{\gamma})]^{k-1}, \\
\frac{\partial^2 \mu_k(D, \boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \mu_0 \partial k} &= -[F(D|\beta, \boldsymbol{\gamma})]^k \log[F(D|\beta, \boldsymbol{\gamma})],
\end{aligned}$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \beta^2} = -\frac{\mu_0 k D}{\beta^2} h(D|\beta,\gamma)\left\{\frac{k-1}{F(D|\beta,\gamma)} - k\right\} f(D|\beta,\gamma)[F(D|\beta,\gamma)]^{k-1},$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \beta \partial \gamma_p} = -\frac{x_p \mu_0 k D}{\beta}\left\{\frac{D(k f(D|\beta,\gamma) - h(D|\beta,\gamma))}{F(D|\beta,\gamma)} + 1\right\} f(D|\beta,\gamma)[F(D|\beta,\gamma)]^{k-1},$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \beta \partial k} = -\frac{\mu_0 D}{\beta}\{1 + \log[F(D|\beta,\gamma)]\} f(D|\beta,\gamma)[F(D|\beta,\gamma)]^{k-1},$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \gamma_p \partial \gamma_q} = -x_p x_q \mu_0 k\left\{\frac{D(k f(D|\beta,\gamma) - h(D|\beta,\gamma))}{F(D|\beta,\gamma)} + 1\right\} f(D|\beta,\gamma)[F(D|\beta,\gamma)]^{k-1},$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial \gamma_p \partial k} = -x_p \mu_0 D\{1 + k \log[F(D|\beta,\gamma)]\} f(D|\beta,\gamma)[F(D|\beta,\gamma)]^{k-1},$$

$$\frac{\partial^2 \mu_k(D,\boldsymbol{x}|\boldsymbol{\theta}^*)}{\partial k^2} = -\mu_0 [F(D|\beta,\gamma)]^k (\log[F(D|\beta,\gamma)])^2,$$

where $F(D|\beta,\gamma) = 1 - e^{-\beta D e^{\gamma^T \boldsymbol{x}}}$, $f(D|\beta,\gamma) = \beta e^{-\beta D e^{\gamma^T \boldsymbol{x}} + \gamma^T \boldsymbol{x}}$, and $h(D|\beta,\gamma) = \beta e^{\gamma^T \boldsymbol{x}}$.

B.3. Gamma-frailty Model for Heterogeneous Sensitivity (Low Dose)

Let us denote risk factors on the j-th target in each observed individual as z_j, j=1,2, ... , k. For a given exposure dose D, covariates vector $\boldsymbol{x}$=(x_1,x_2, ... ,x_p)^T, and risk factors Z, we construct a model having the form

$$\mu_k(D,\boldsymbol{x}|z,\boldsymbol{\theta}^*) = \mu_0\left\{1 - \prod_{j=1}^{k}(1 - e^{-\beta_j D e^{\gamma^T \boldsymbol{x}} z_j})\right\}. \tag{13}$$

If $\beta_j D$ comes close to zero for $^\forall j$=1,2, ... , k, then the model can be approximately expressed by

$$\mu_k(D,\boldsymbol{x}|z,\boldsymbol{\theta}^*) \simeq \mu_0\left\{1 - \prod_{j=1}^{k}(\beta_j D z_j e^{\gamma^T \boldsymbol{x}})\right\}. \tag{14}$$

By assuming that the sensitivity coefficient of the j-th target (β_j) has regularity following geometrical progression, that is, β_j=$\beta\rho^{j-1}$, the model will be specified by

$$\mu_k(D,\boldsymbol{x}|z,\boldsymbol{\theta}^*) \simeq \mu_0\left\{1 - (\beta\rho^* D \bar{z}^k e^{\gamma^T \boldsymbol{x}})\right\}, \tag{15}$$

where ρ^*=$\rho^{[k(k-1)/2]}$ and $\bar{z}^k$=$\prod_{j=1}^{k} z_j$. Thus, the likelihood function based on the complete data set $\boldsymbol{d}$ is

$$L(\boldsymbol{\theta}^*|\boldsymbol{d}) = \prod_{i=1}^{n} P(y_i|\bar{z}_i, D_i, \boldsymbol{x}_i, \boldsymbol{\theta}^*), \tag{16}$$

where $P(y|\bar{z}, D, \boldsymbol{x}, \boldsymbol{\theta}^*)$ expresses the probability density function of Poisson distribution with mean $\mu_k(D,\boldsymbol{x}|z,\boldsymbol{\theta}^*)$, as shown in equation (15). Integrating the form of the likelihood function in equation (16) with respect to the density function of Z in equation (7) provides the likelihood function based on observed data set $\boldsymbol{d}_{(obs)}$ given by

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}) &= \int_0^\infty L(\boldsymbol{\theta}^*|\boldsymbol{d})\ \varphi(z_i|\sigma)\ dz_i \\ &= \prod_{i=1}^{n} \int_0^\infty P(y_i|\bar{z}_i, D_i, \boldsymbol{x}_i, \boldsymbol{\theta}^*)\ \varphi(z_i|\sigma)\ dz_i \\ &= \prod_{i=1}^{n} g(y_i|D_i, \boldsymbol{x}_i, \boldsymbol{\theta}), \end{aligned} \tag{17}$$

where

$$g(y|D,\boldsymbol{x},\boldsymbol{\theta}) = \frac{\mu_0^y e^{-\mu_0}}{y!}\left\{\frac{1-(y+\mu_0\sigma)\beta D e^{\boldsymbol{\gamma}^T\boldsymbol{x}}\rho^{\frac{1}{2}k(k-1)}}{\{1-\beta D e^{\boldsymbol{\gamma}^T\boldsymbol{x}}\rho^{\frac{1}{2}k(k-1)}\}^{\frac{1}{\sigma}(1+\sigma)}}\right\}. \tag{18}$$

The log-likelihood function of the model as shown in equation (17) can be specified by

$$\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)}) = \sum_{i=1}^{n}\log g(y_i|D_i,\boldsymbol{x}_i,\boldsymbol{\theta}).$$

Then, elements of the Hessian matrix are

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0} &= \sum_{i=1}^{n}\left\{\frac{y_i}{\mu_0}-1+M(D_i|\boldsymbol{\theta}^*)\Big(\frac{Q(y_i|D_i,\boldsymbol{\theta})}{\mu_0}-\frac{\sigma^2R(y_i|D_i,\boldsymbol{\theta})}{y_i+\mu_0\sigma^2}\Big)\right\},\\
\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\beta} &= \sum_{i=1}^{n}\frac{1}{\beta}T(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\gamma_p} &= \sum_{i=1}^{n}x_pT(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\rho} &= \sum_{i=1}^{n}\frac{k(k-1)}{2\rho}T(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial k} &= \sum_{i=1}^{n}(k-\frac{1}{2})T(y_i|D_i,\boldsymbol{\theta})\log\rho,\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0^2} &= \sum_{i=1}^{n}\left\{\sigma^2M(D_i|\boldsymbol{\theta}^*)U(y_i|D_i,\boldsymbol{\theta})-\frac{y_i}{\mu_0^2}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0\partial\beta} &= \sum_{i=1}^{n}\frac{1}{\mu_0}U(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0\partial\gamma_p} &= \sum_{i=1}^{n}x_pU(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0\partial\rho} &= \sum_{i=1}^{n}\frac{k(k-1)}{2\rho}U(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\mu_0\partial k} &= \sum_{i=1}^{n}(k-\frac{1}{2})U(y_i|D_i,\boldsymbol{\theta})\log\rho,\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\beta^2} &= \sum_{i=1}^{n}\frac{(M(D_i|\boldsymbol{\theta}^*))^2}{\beta^2}\left\{\frac{\mu_0\sigma^2Q(y_i|D_i,\boldsymbol{\theta})}{1-\mu_0\sigma^2M(D_i|\boldsymbol{\theta}^*)}-\frac{(y_i+\mu_0\sigma^2)R(y_i|D_i,\boldsymbol{\theta})}{1-(y_i+\mu_0\sigma^2)M(D_i|\boldsymbol{\theta}^*)}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\beta\partial\gamma_p} &= \sum_{i=1}^{n}\frac{x_p}{\beta}V(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\beta\partial\rho} &= \sum_{i=1}^{n}\frac{k(k-1)}{2\beta\rho}V(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\beta\partial k} &= \sum_{i=1}^{n}\frac{(2k-1)\log\rho}{2\beta}V(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\gamma_p\partial\gamma_q} &= \sum_{i=1}^{n}x_px_qV(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\gamma_p\partial\rho} &= \sum_{i=1}^{n}\frac{k(k-1)x_p}{2\rho}V(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\gamma_p\partial k} &= \sum_{i=1}^{n}\frac{(2k-1)\log\rho}{2}V(y_i|D_i,\boldsymbol{\theta}),\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\rho^2} &= \sum_{i=1}^{n}\frac{k(k-1)}{4\rho^2}\left\{\Big((k-2)(k+1)+2\mu_0\sigma^2M(D_i|\boldsymbol{\theta}^*)\Big)V(y_i|D_i,\boldsymbol{\theta})-\frac{2y_i[M(D_i|\boldsymbol{\theta}^*)]^2}{1-(y_i+\mu_0\sigma^2)M(D_i|\boldsymbol{\theta}^*)}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial\rho\partial k} &= \sum_{i=1}^{n}\frac{k(k-1)(2k-1)}{4\rho^2}\left\{\Big(k(k-1)\log\rho+2(1-\mu_0\sigma^2M(D_i|\boldsymbol{\theta}^*))\Big)V(y_i|D_i,\boldsymbol{\theta})+\frac{2y_i[M(D_i|\boldsymbol{\theta}^*)]^2}{1-(y_i+\mu_0\sigma^2)M(D_i|\boldsymbol{\theta}^*)}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{d}_{(obs)})}{\partial k^2} &= \sum_{i=1}^{n}\frac{\log\rho}{4}\left\{\Big((2k-1)^2\log\rho+4(1-\mu_0\sigma^2M(D_i|\boldsymbol{\theta}^*))\Big)V(y_i|D_i,\boldsymbol{\theta})+\frac{4y_i[M(D_i|\boldsymbol{\theta}^*)]^2}{1-(y_i+\mu_0\sigma^2)M(D_i|\boldsymbol{\theta}^*)}\right\},
\end{aligned}$$

where:

$$
\begin{aligned}
M(D|\boldsymbol{\theta}^*) &= \beta D e^{\gamma^T \boldsymbol{x}} \rho^{\frac{1}{2}k(k-1)}, \\
Q(D|\boldsymbol{\theta}) &= \frac{(1+\sigma^2)\mu_0}{1-\mu_0\sigma^2 M(D|\boldsymbol{\theta}^*)}, \\
R(y|D,\boldsymbol{\theta}) &= \frac{y+\mu_0\sigma^2}{1-(y+\mu_0\sigma^2)M(D|\boldsymbol{\theta}^*)}, \\
T(y|D,\boldsymbol{\theta}) &= M(D|\boldsymbol{\theta}^*)\Big\{Q(D|\boldsymbol{\theta}) - R(y|D,\boldsymbol{\theta})\Big\}, \\
U(y|D,\boldsymbol{\theta}) &= M(D|\boldsymbol{\theta}^*)\left\{\frac{Q(D|\boldsymbol{\theta})}{\mu_0(1-\mu_0\sigma^2 M(D|\boldsymbol{\theta}^*))} - \frac{\sigma^2 R(y|D,\boldsymbol{\theta})}{(y+\mu_0\sigma^2)(1-(y+\mu_0\sigma^2)M(D|\boldsymbol{\theta}^*))}\right\}, \\
V(y|D,\boldsymbol{\theta}) &= M(D|\boldsymbol{\theta}^*)\left\{\frac{Q(D|\boldsymbol{\theta})}{1-\mu_0\sigma^2 M(D|\boldsymbol{\theta}^*)} - \frac{R(y|D,\boldsymbol{\theta})}{1-(y+\mu_0\sigma^2)M(D|\boldsymbol{\theta}^*)}\right\}.
\end{aligned}
$$

Appendix C

The algorithm of SPIDER proposed by Ohtaki & Izumi[18] are described as the following steps:

Step 1. Set initial values of the parameters for maximizing of p-dimensional function f, and let denote it as $\boldsymbol{\alpha}_0^{(0)}$.

Step 2. By starting with $\boldsymbol{\alpha}_0^{(s)}$, where s=0,1,2, ... , perform loop at the s stage. Define the function $f_\ell(t) = f(\boldsymbol{\alpha}_{\ell-1}^{(s)} + t\boldsymbol{\delta}_l)$for ℓ=1, ... , p, where $\boldsymbol{\delta}_\ell = (\delta_{\ell 1}, \delta_{\ell 2}, \cdots, \delta_{\ell p})^T$, a vector of Kronecker's delta. Optimize the function f_ℓ and set

$$
\begin{aligned}
t_\ell &= \arg \max_{t\in(-\infty,+\infty)} f_\ell(t) \\
\boldsymbol{\alpha}_\ell^{(s)} &= \boldsymbol{\alpha}_{\ell-1}^{(s)} + t_q \boldsymbol{\delta}_\ell
\end{aligned}
$$

Step 3. Calculate $\Delta_\ell = \| \boldsymbol{\alpha}_0^{(s)} - \boldsymbol{\alpha}_\ell^{(s)} \|$. If Δ_ℓ becomes small enough, then quit. Otherwise go back to Step 2 with $\boldsymbol{\alpha}_0^{(s+1)} = \boldsymbol{\alpha}_\ell^{(s)}$. Continue Step 2 and Step 3 until convergence.