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 $^\forall\boldsymbol{\theta} \in \Theta$, 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:

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

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

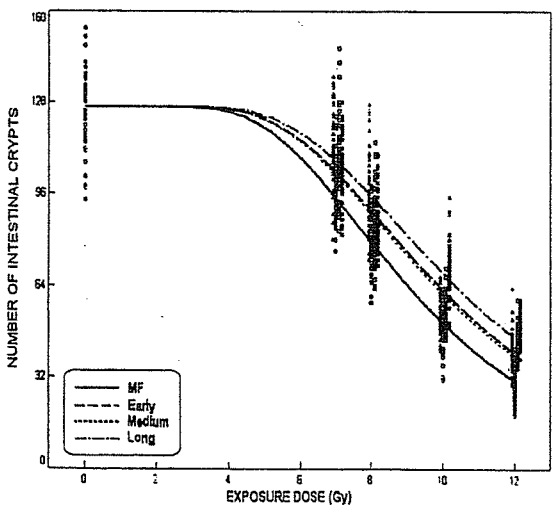

Fig. 2. Curve of the density of the intestinal crypt after an exposure event based on the Poisson regression model according to the fermented-stage of miso, with mice fed with a commercial diet of MF used as controls. The survival rate of crypts of mice fed long-term fermented miso has a higher rate 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.

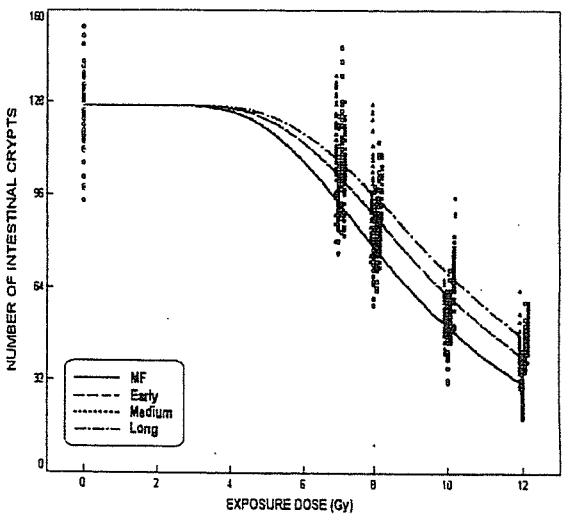

Fig. 3. Curve of the density of the intestinal crypt after an exposure event based on the gamma-frailty model according to the fermented-stage of miso, with mice fed a commercial MF diet used as controls. The survival curve of the crypts of mice fed long-term fermented miso has a slightly decreasing 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

$$\ell(\theta^*|d_{(obs)}) = \sum_{i=1}^{n} \log P(y_i|D_i, x_i, \theta^*).$$

Then, elements of the Hessian matrix are given by

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

B.2. Gamma-frailty Model for Heterogeneous Background

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

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

Elements of the Hessian matrix are therefore specified as follows:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}|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}|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}|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}|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}|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

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

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

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

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

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

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

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

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

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 $\boldsymbol{x}$=$(x_1,x_2, \ldots ,x_p)^T$, and risk factors Z, we construct a model having the form

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

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

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

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

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

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

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

where $P(y|\bar{z}, D, \boldsymbol{x}, \theta^*)$ expresses the probability density function of Poisson distribution with mean $\mu_k(D,\boldsymbol{x}|z,\theta^*)$, 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

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

where

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

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

$$\ell(\theta|d_{(obs)})=\sum_{i=1}^{n}\log g(y_i|D_i,x_i,\theta).$$

Then, elements of the Hessian matrix are

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

where:

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

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 $\alpha_0^{(0)}$.

Step 2. By starting with $\alpha_0^{(s)}$, where s=0,1,2, ... , perform loop at the s stage. Define the function $f_\ell(t) = f(\alpha_{\ell-1}^{(s)} + t\delta_l)$ for ℓ=1, ... , p, where $\delta_\ell = (\delta_{\ell 1}, \delta_{\ell 2}, \cdots, \delta_{\ell p})^T$, a vector of Kronecker's delta. Optimize the function f_ℓ and set

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

Step 3. Calculate $\Delta_\ell = \| \alpha_0^{(s)} - \alpha_\ell^{(s)} \|$. If Δ_ℓ becomes small enough, then quit. Otherwise go back to Step 2 with $\alpha_0^{(s+1)} = \alpha_\ell^{(s)}$. Continue Step 2 and Step 3 until convergence.

Japanese Journal of Biometrics Vol. 26, No. 1, 31–48 (2005)

Preliminary Report

Model-based analysis of microarray data: Exploration of differentially expressed genes between two cell types based on a two-dimensional mixed normal model

Megu Ohtaki[*1,†], Keiko Otani[*2], Kenichi Satoh[*1], Toshihiko Kawamura[*3], Keiko Hiyama[*4] and Masahiko Nishiyama[*4]

[*1]Department of Environmetrics and Biometrics,
Research Institute for Radiation Biology and Medicine,
Hiroshima University, Hiroshima, Japan
[*2]Japan Biological Informatics Consortium
[*3]Hiroshima Cancer Therapy Development organization
[*4]Department of Translational Cancer Research,
Research Institute for Radiation Biology and Medicine,
Hiroshima University, Hiroshima, Japan
[†]e-mail: ohtaki@hiroshima-u.ac.jp

Inference on gene expression change between two different samples is considered. We develop a mathematical model assuming that there exist two different functional states of a gene: "ON" and "OFF" . Each measured sample-specific gene expression intensity is described by an additive model, which accounts for fluctuations in absolute gene expression intensity and measurement error, to which a two-dimensional mixed normal model with four components considering the joint distribution of the sample "sum" and "difference" is approximated. We can successfully identify genes that are differentially expressed between two samples using posterior probabilities, while avoiding declaring false differences. The proposed methods are applicable to cDNA microarray data with two fluorescent dyes and to oligonucleotide data.

Key words: cDNA microarray; Empirical Bayes; Gene Expression; Mixed Normal Distribution; Normalization; Oligonucleotide microarray

1. Introduction

DNA microarray technology is presently the most effective high-throughput tool for identifying specific genes among tens of thousands of background genes (Gerhold et al., 2002; Schena et al., 1995). In cDNA microarray experiments, genes differentially expressed between query and reference samples (e.g. cancer and noncancerous cells) are identified through many processes using two different fluorescent dyes. The microarray experimental procedure proceeds as follows. **RNA isolation:** Two mRNA samples to be compared are isolated from the query and reference samples and reverse transcribed into cDNA.

Received December 2004. Revised May 2005. Accepted June 2005.

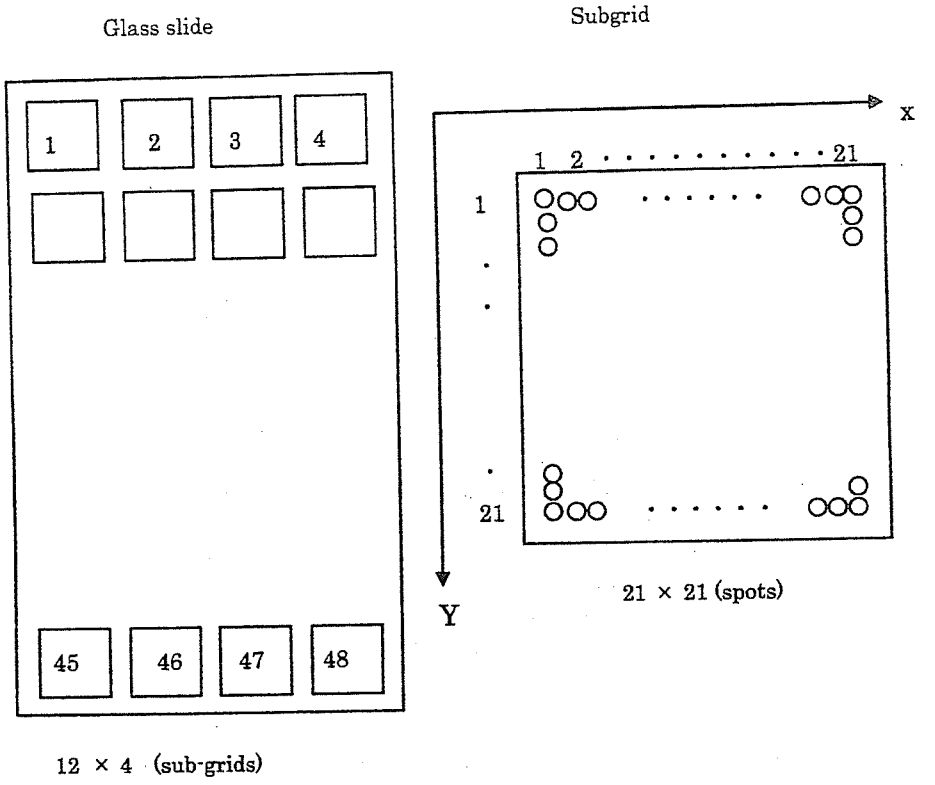

Fig. 1. A diagram of the microarray glass slide is shown schematically. It contains 48 subgrids, each containing 441 spots. In total, 21168 genes are spotted in a microarray.

Sample labeling: The cDNA from each sample is labeled using either a red fluorescent dye (cy5) or a green fluorescent dye (cy3).

Hybridization: Equal quantities of the two differentially labeled samples are mixed and hybridized to a microarray containing 21168 cDNA probes.

Array preparation: Array preparation depends on the experimental design. There are many variations in the geometrical setup of the microarray glass slide and gridding head used for spotting. We used a glass slide, Riken human 21K array, with a subgrid of 48 blocks. Each block contained 441 spots (21 × 21). Each spot on the slide is numbered from 1 to 21168 (48 × 441), which we call gene IDs. An (i, j) coordinate indicates the location of a spot at the i-th row and j-th column in a subgrid. The midpoint of a subgrid is denoted by (m_r, m_c). Figure 1 shows a schematic diagram of the microarray slide.

Data collection: The slides are imaged using a scanner and fluorescence measurements are made separately at each spot on the array by channels 1 and 2 for the two dyes. The data obtained from each channel consists of foreground and background intensities.

The purpose of our study is to successfully identify differentially expressed genes between the two samples, while avoiding false positives (i.e. declaring unchanged genes as differentially expressed). We propose a new mathematical model for microarray data based on a hypothesis

for functional status of the genes, and a method for estimating the probability of a gene being expressed differentially in the two samples being compared.

In Section 2, we introduce the mathematical model for microarray data. So far several model-based approaches have been proposed for identification of differentially expressed genes. They are divided roughly into the following two methods: ANOVA based on the fixed effects linear model (Churchill et al., 2002; Kerr et al., 2000; Lee et al., 2000) and the empirical Bayes method (Baldi et al., 2001; Kendziorski et al., 2003; Long et al., 2001; Newton et al., 2001). An empirical Bayes approach that treats the gene expression intensities as arising from some population was originally proposed by Newton et al. (2001). We also employ the two-group empirical Bayes method to infer differentially expressed genes between two samples. The point in which our approach differs from theirs is that the functional status of a gene is introduced into the mathematical model using a binary variable. We assume there exist two different functional states of a gene: "ON" and "OFF". Biologically, "ON" means the gene produces mRNA and "OFF" means the gene does not produce it. In cases where a gene is "ON", the gene expression intensity is regarded as the "sum" of a random variable that obeys the log normal distribution and a measurement error. In cases where a gene is "OFF", the gene expression intensity is regarded as measurement error only.

We use the "sum" and "difference" simultaneously, for which the variable transformation named S-D transform is introduced in Section 2.1. The acronym "S-D" is used for an abbreviation of "sum" and "difference" of gene expression levels in the query and reference samples. Here the scatter plot using these variables is called an S-D plot. In Section 2.2, a two-dimensional mixed normal density function having at most four components is introduced as the joint distribution of the S-D transformed variables. Section 2.3 describes the exploration of differentially expressed genes. The probability of gene i expressing differently between query and reference samples is obtained as a posterior probability. In Section 3, we explain the implementation of our method to detect differentially expressed genes using real cDNA microarray data. Because a massive amount of data is generated by cDNA microarray experiments, there could be large experimental variations that affect the resultant estimated gene expression levels. Many researchers stress the importance of normalization before carrying out the statistical analysis (Dudoi et al., 2002; Fan et al., 2004; Saviozzi et al., 2003; Schudhhardt et al., 2002; Yang et al., 2002; Wu et al., 2001). We describe a procedure of normalization that is based directly on the mathematical model in Section 3.1. Section 3.2 describes the parameter-estimation procedure.

Though we focus on cDNA microarrays in the present development, the proposed model is also applicable for analyzing a pair of expression data from oligonucleitide arrays (Irizarry et al., 2003: Li et al., 2001).

2. Mathematical model for microarray data

Suitably transformed (the logarithmic transformation in this study) and normalized gene expression measurements are expressed with a simple additive mathematical model that accounts for measurement error and fluctuations in absolute gene expression levels for each channel. We consider two types of measurement error in microarray data. One is common to channels 1 and 2, the other is detected independently between the two channels. Let $Y_i^{(1)}$ and $Y_i^{(2)}$ be the expression intensities of gene g_i in the query and reference samples, respectively, which are suitably transformed and normalized. The mathematical model for microarray data can be described as follows:

$$\begin{cases} Y_i^{(1)} = \tau_i^{(1)}\alpha_i\rho_i + \beta_i + \varepsilon_i^{(1)}, \\ Y_i^{(2)} = \tau_i^{(2)}\alpha_i\rho_i + \beta_i + \varepsilon_i^{(2)}. \end{cases} \tag{1}$$

The symbols $\tau_i^{(1)}$ and $\tau_i^{(2)}$ represent the expression status of gene g_i in the query and reference samples, respectively, which are defined by

$$\tau_i = \begin{cases} 1 & \text{if } g_i \text{ is "ON"}, \\ 0 & \text{if } g_i \text{ is "OFF"}. \end{cases}$$

The symbol α_i represents the expression level of gene g_i when it is "ON". The symbol ρ_i represents the volume of cDNA probe, which includes the fluctuation in probe volume. Since α_i and ρ_i are not identifiable unless repeated measurement are available, we replace $\alpha_i\rho_i$ by α_i for simplicity. We regard it as a positive random variable having log-normal distribution with mean $\log\mu - \frac{\lambda^2}{2}$ and variance λ^2 (*i.e.* $\log\alpha_i \sim N(\log\mu - \frac{\lambda^2}{2}, \lambda^2)$). Thus, $E(\alpha_i) = \mu$ and $Var(\alpha_i) = \mu^2(e^{\lambda^2} - 1)$. The symbols β_i denote random errors common to channels 1 and 2, which obey a normal probability density function with mean 0 and variance σ_β^2 (i.e. $\beta_i \sim i.i.d.\ N(0, \sigma_\beta^2)$). With oligonucleotide maicroarrays, each microarray measures a single sample and provides an absolute measurement level for each RNA molecule (Butte, 2002). Therefore the term β_i should be negligible in the case of oligonucleotide microarray. The symbols $\varepsilon_i^{(1)}$ and $\varepsilon_i^{(2)}$ indicate random errors, which are mutually independent and have normal probability density functions with mean zero and variance σ_ε^2 (i.e. $\varepsilon_i^{(1)}, \varepsilon_i^{(2)} \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$).

2.1 S-D transformation

The S-D transformation of paired expression intensities ($Y_i^{(1)}$, $Y_i^{(2)}$) into the "sum" and "difference". (U_i, V_i) can be described mathematically as follows:

$$\begin{cases} U_i = Y_i^{(1)} + Y_i^{(2)} = (\tau_i^{(1)} + \tau_i^{(2)})\alpha_i + 2\beta_i + \varepsilon_i^{(1)} + \varepsilon_i^{(2)}, \\ V_i = Y_i^{(1)} - Y_i^{(2)} = (\tau_i^{(1)} - \tau_i^{(2)})\alpha_i + \varepsilon_i^{(1)} - \varepsilon_i^{(2)}. \end{cases} \tag{2}$$

Then the conditional mean and variance of U and V are as follows:

$$E[U|(\tau^{(1)}, \tau^{(2)})] = \begin{cases} 0, & (\tau^{(1)}, \tau^{(2)}) = (0, 0), \\ 2\mu, & (\tau^{(1)}, \tau^{(2)}) = (1, 1), \\ \mu, & (\tau^{(1)}, \tau^{(2)}) = (1, 0), \\ \mu, & (\tau^{(1)}, \tau^{(2)}) = (0, 1), \end{cases} \tag{3}$$

$$Var[U|(\tau^{(1)}, \tau^{(2)})] = \begin{cases} 4\sigma_\beta^2 + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (0, 0), \\ 4\mu^2(e^{\lambda^2} - 1) + 4\sigma_\beta^2 + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (1, 1), \\ \mu^2(e^{\lambda^2} - 1) + 4\sigma_\beta^2 + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (1, 0), \\ \mu^2(e^{\lambda^2} - 1) + 4\sigma_\beta^2 + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (0, 1), \end{cases} \tag{4}$$

$$E[V|(\tau^{(1)}, \tau^{(2)})] = \begin{cases} 0, & (\tau^{(1)}, \tau^{(2)}) = (0, 0), \\ 0, & (\tau^{(1)}, \tau^{(2)}) = (1, 1), \\ \mu, & (\tau^{(1)}, \tau^{(2)}) = (1, 0), \\ -\mu, & (\tau^{(1)}, \tau^{(2)}) = (0, 1), \end{cases} \tag{5}$$

$$Var[V|(\tau^{(1)}, \tau^{(2)})] = \begin{cases} 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (0, 0), \\ 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (1, 1), \\ \mu^2(e^{\lambda^2} - 1) + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (1, 0), \\ \mu^2(e^{\lambda^2} - 1) + 2\sigma_\varepsilon^2, & (\tau^{(1)}, \tau^{(2)}) = (0, 1). \end{cases} \tag{6}$$

The concept of the S-D transformation for the four possible states of $(\tau_i^{(1)}, \tau_i^{(2)})$ is visually illustrated in Figure 2. A gene inside region A is "OFF" in both samples, a gene inside region B is "ON" in both samples, and a gene in region C or D is "ON" in one sample and "OFF" in the other.

Under the condition that $\tau_i^{(1)} = \tau_i^{(2)}$ (i.e. gene g_i is "ON" or "OFF" in both samples), the variable V follows the normal distribution $N(0, 2\sigma_\varepsilon^2)$, in which the value V_i shows only measurement error. Then the equations

$$E(V|U = u) = 0, \tag{7}$$

$$Var(V|U = u) = 2\sigma_\varepsilon^2, \tag{8}$$

hold regardless of the value U.

Using the inverse S-D transformation, we obtain the equations

$$\begin{cases} Y_i^{(1)} = \frac{1}{2}(U_i + V_i), \\ Y_i^{(2)} = \frac{1}{2}(U_i - V_i). \end{cases} \tag{9}$$

2.2 Two-dimensional mixed normal model

We introduce a two-dimensional normal mixture model having at most four components as the joint distribution of (U, V), whose density function is given by

$$f(u, v|\boldsymbol{p}, \boldsymbol{\theta}) = \sum_{s,t \in \{0,1\}} p_{st} f_{st}(u, v|\boldsymbol{\theta}), \tag{10}$$

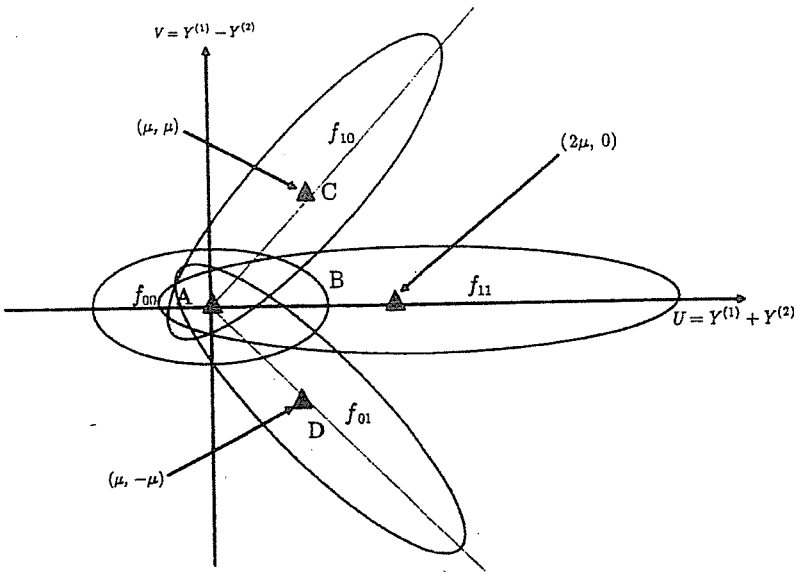

Fig. 2. The two-dimensional mixed normal model is shown schematically. The X-axis indicates the sum of the expression intensities in the query and reference samples; the Y-axis indicates the difference. The states of genes inside the regions A, B, C and D are (OFF, OFF), (ON, ON), (ON, OFF) and (OFF, ON).

where f_{st} denotes the density function under $\tau_1 = s$ and $\tau_2 = t$, p_{st} denotes its mixture rate under $\tau_1 = s$ and $\tau_2 = t$, $\theta = (\mu, \lambda^2, \sigma_\beta^2, \sigma_\epsilon^2)$ and p=$(p_{00}, p_{01}, p_{10}, p_{11})$ with $p_{00} + p_{01} + p_{10} + p_{11} = 1$. In this study, each function $f_{st}(u, v)$ is approximated by the two-dimensional normal density function with the moments specified by (4), (5), (6) and (7), respectively. Thus, they are described as

$$f_{00}(u,v|\,\theta) = \phi(u|\,0,\, 2\sigma_\epsilon^2 + 4\sigma_\beta^2)\phi(v|\,0,\, 2\sigma_\epsilon^2),$$
$$f_{11}(u,v|\,\theta) = \phi(u|\,2\mu,\, 4\mu^2(e^{\lambda^2} - 1) + 2\sigma_\epsilon^2)\phi(v|\,0,\, 2\sigma^2),$$

where ϕ denotes a one-dimensional normal density function,

$$f_{10}(u,v|\,\theta) = \phi_2((u,v)|\,\mu(1,1)',\, \Sigma_{10}),$$
$$f_{01}(u,v|\,\theta) = \phi_2((u,v)|\,\mu(1,-1)',\, \Sigma_{01}),$$

where ϕ_2 denotes a two-dimensional normal density function,

$$\Sigma_{10} = \begin{pmatrix} \mu^2(e^{\lambda^2} - 1) + 4\sigma_\beta^2 + 2\sigma_\epsilon^2 & , & \mu^2(e^{\lambda^2} - 1) \\ \mu^2(e^{\lambda^2} - 1) & , & \mu^2(e^{\lambda^2} - 1) + 2\sigma_\epsilon^2 \end{pmatrix}$$

and $\Sigma_{01} = \begin{pmatrix} \mu^2(e^{\lambda^2} - 1) + 4\sigma_\beta^2 + 2\sigma_\epsilon^2 & , & -\mu^2(e^{\lambda^2} - 1) \\ -\mu^2(e^{\lambda^2} - 1) & , & \mu^2(e^{\lambda^2} - 1) + 2\sigma_\epsilon^2 \end{pmatrix}$.

The parameters θ and p are estimated by maximum likelihood using the Newton-Raphson method and the Spider algorithm (Arcana and Ohtaki, 2005). The flow of estimation is described in Section 3.3.

2.3 Posterior probabilities

Given estimates $\hat{\theta}$ and $\hat{p}$, and normalized data (u_i, v_i), the posterior probabilities with respect to the status of gene expression can be expressed as

$$Pr(\tau^{(1)} = 1 | (u,v), \hat{\theta}) = \frac{\hat{p}_{10} f_{10}(u,v|\hat{\theta}) + \hat{p}_{11} f_{11}(u,v|\hat{\theta})}{f(u,v|\hat{\theta})},$$

$$Pr(\tau^{(2)} = 1 | (u,v), \hat{\theta}) = \frac{\hat{p}_{01} f_{01}(u,v|\hat{\theta}) + \hat{p}_{11} f_{11}(u,v|\hat{\theta})}{f(u,v|\hat{\theta})},$$

$$Pr(\tau^{(1)} \neq \tau^{(2)} | (u,v), \hat{\theta}) = \frac{\hat{p}_{01} f_{01}(u,v|\hat{\theta}) + \hat{p}_{10} f_{10}(u,v|\hat{\theta})}{f(u,v|\hat{\theta})}. \quad (11)$$

If a gene has a relatively large value of $Pr(\tau^{(1)} \neq \tau^{(2)})$, then it is assumed to be expressed differentially between the two samples.

3. Implementation of data analysis

In this section we illustrate how to implement the exploration of differentially expressed genes between two cell types based on the proposed model, using a set of real microarray data.

3.1 Background correction

Let $(y_i^{(\ell)}, b_i^{(\ell)})$, $(\ell = 1,2)$ be a pair of foreground and background intensities measured by both channels and let $y_i^{(\ell)*}$ be a background corrected value for the i-th gene. So far the background correction has commonly been done by $y_i^{(\ell)*} = y_i^{(\ell)} - b_i^{(\ell)}$ (Eisen et al., 2002) and its logarithmic transformed value is taken as a transformed background corrected intensity. When $y_i^{(\ell)} - b_i^{(\ell)}$ is negative, it is truncated and replaced by an appropriate small positive value, which yields frequently that an unreasonably large dispersion appears at low expression region in the logarithmic transformed background corrected intensities. In this study, we alternatively adopt the equation that is expressed by

$$y_i^{(\ell)*} = y_i^{(\ell)} / b_i^{(\ell)} \quad (12)$$

to correct background effects.

3.2 Normalization

In the first row of Figure 3, (a1) and (a2), whole 21168 gene expression levels of the transformed sample data are shown in order of gene ID. The left side is channel 1(the query sample) and the right side is channel 2 (the reference sample). (a2) shows that the intensity level varies as a wave depending on gene ID, suggesting that some spatial dependent difference may exist between the subgrids. Step N1 aims to remove it (see 'Flow of normalization'). The magnifications of the first four subgrids are shown in the second row of Figure 3, (b1) and (b2). A similar periodic pattern appears in every subgrid, implying that the intensity level of gene expression varies depending on location of the spot in the subgrid. If the spots are printed on the array randomly, the profiles of plots should not have such a systematic tendency. The purpose of Step

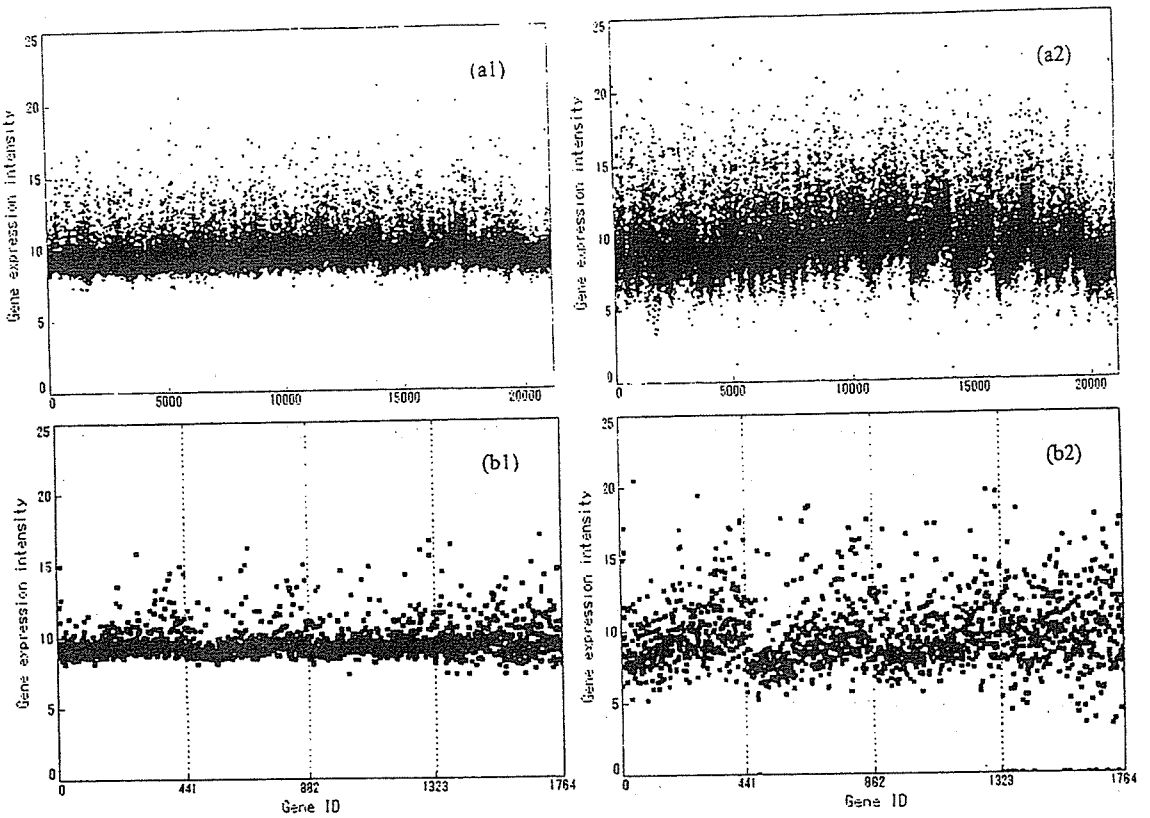

Fig. 3. (a1) and (a2) show the log-transformed gene expression intensities of channels 1 and 2, respectively. (b1) and (b2) are the magnifications of the first four subgrids of both channels.

N2 is to minimize biases associated with location of spots in a subgrid (see 'Flow of normalization'). The S-D plot of the sample data is shown in the left of Figure 5. If the efficiency of the two dyes were the same, the points should distribute almost symmetrically around the X-axis. This dependent dye bias can be removed by Step N3 (see 'Flow of normalization').

Such aberrant trends in the original observations should be removed by applying a combination of global normalization using all of the data and a local one using physical subsets of the data (Quackenbush, 2002). We propose the normalization procedure which is applied to the S-D transformed variables U and V rather than to the original variables $Y^{(1)}$ and $Y^{(2)}$. The procedure is based on the following two assumptions derived from the proposed model. Firstly, most of the u_i are presumed to be generated from the normal mixture density function with two components. Secondly, most of the v_i show only measurement error. Details are described in the Discussion. We can describe the flow of this normalization briefly as follows.

[Flow of normalization]

Step N1. [Adjustment among the subgrids]

Let the values $u_k(i, j)$ and $v_k(i, j)$ be those of the gene located at coordinates (i,j) in the k-th subgrid ($i = 1, \cdots, 21$, $j = 1, \cdots, 21$, $k = 1, \cdots, 48$, $\ell = 1, 2$). Then the data are updated

as follows:

$$u_k(i,j) := u_k(i,j) - Q_k^{(u)}(35) + Q_*^{(u)}(35),$$
$$v_k(i,j) := v_k(i,j) - Q_k^{(r)}(50),$$

where $Q_k^{(u)}(35)$ and $Q_k^{(r)}(50)$ indicate the 35% point of u and 50% point of v in the k-th subgrid, respectively, and $Q_*^{(u)}(35)$ indicates the 35% point of u using all spots on the slide.

Step N2. [Adjustment among spots in a subgrid]

Let $u^*(i,j)$ and $v^*(i,j)$ be the leveled $u(i,j)$ and $v(i,j)$ using all subgrids, which are defined by

$$\begin{cases} u^*(i,j) = \frac{1}{48}\sum_{k=1}^{48} u^{(k)}(i,j), \\ v^*(i,j) = \frac{1}{48}\sum_{k=1}^{48} v^{(k)}(i,j), \end{cases}$$

respectively. Assume that

$$\begin{cases} u^*(i,j) - \bar{u} = a_u(i) + b_u(j) + c_u((i-m_r)(j-m_c)) + \varepsilon_u(s,t), \\ v^*(i,j) - \bar{v} = a_r(i) + b_r(j) + c_r((i-m_r)(j-m_c)) + \varepsilon_r(s,t), \end{cases}$$

where

$$\bar{u} = \frac{1}{21\times 21}\sum_i\sum_j u^*(i,j), \quad \bar{v} = \frac{1}{21\times 21}\sum_i\sum_j v^*(i,j),$$

and $a_u(i)$, $b_u(j)$, $c_u((i-m_r)(j-m_c))$, $a_r(i)$, $b_r(j)$ and $c_r((i-m_r)(j-m_c))$ are fixed parameters satisfying

$$\sum_i a_u(i) = 0, \ \sum_j b_u(j) = 0, \quad \sum_i\sum_j c_u(i-m_r)(j-m_c)) = 0.$$

Then estimate the functions a_u, b_u, c_u, a_r, b_r and c_r nonparametrically through the ACE algorithm (Breiman and Friedman, 1985). Given the estimates $\hat{a_u}$, $\hat{b_u}$, $\hat{c_u}$, $\hat{a_r}$, $\hat{b_r}$ and $\hat{c_r}$, update $u(i,j)$ and $v(i,j)$ using the following formula:

$$\begin{cases} u(i,j) := u(i,j) - \hat{a}_u(i) + \hat{b}_u(j) + \hat{c}_u((i-m_r)(j-m_c)), \\ v(i,j) := v(i,j) - \hat{a}_r(i) + \hat{b}_r(j) + \hat{c}_r((i-m_r)(j-m_c)). \end{cases}$$

Step N3. [Removal of dye dependent bias]

Apply the non-parametric regression model,

$$v_i = \phi(u_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0,$$

to the S-D plots to obtain the trend of v_i depending on u_i, and estimate the trend function ϕ by using a moving average method. Then update v_i by $v_i - \hat{\phi}(u_i)$. The normalized channel specific gene expression intensity is given by

$$\begin{cases} \hat{y}_i^{(1)} = \frac{1}{2}\{u_i + v_i - \hat{\phi}(u_i)\}. \\ \hat{y}_i^{(2)} = \frac{1}{2}\{u_i - v_i + \hat{\phi}(u_i)\}. \end{cases}$$

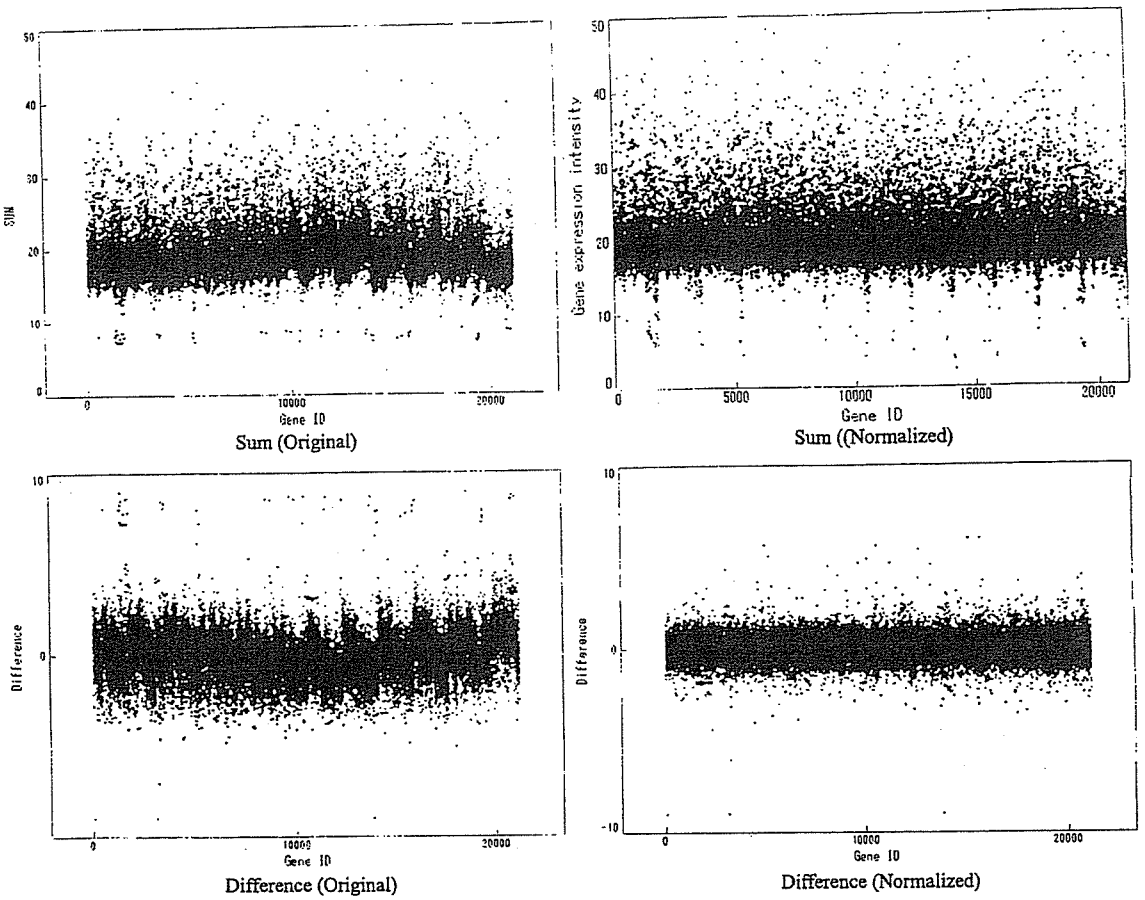

Fig. 4. The scatter plot of the sum of the two channels is shown above and that of the difference is shown below. The scatter plots of original data are shown on the left, and those of normalized data on the right.

Repeat the normalization procedures from Step N1 to Step N3 until the systematic error is removed. Figures 4 and 5 show the efficiency of normalization graphically by comparing the original data (left side) with the normalized data (right side). The Y-axis shows the "sum" of the two channels or the "difference" of the two cannels. Figure 5 shows the S-D plot.

3.3 Parameter estimation

In the ordinary microarray examination, a very low frequency, such as less than three percent, is expected for the heterogeneous components. Therefore they can be negligible at the estimation of the initial value of μ. The model with four components is used in the step E3. The flow of parameter estimation can be described by the following steps.

Step E1. Fit the following mixed normal model with two components to data along the U-axis,

$$(1-\xi)\phi(u-\mu_0|\sigma_0^2)+\xi\phi(u-\mu_1|\sigma_1^2),$$

where μ_0 and μ_1 denote means, σ_0^2 and σ_1^2 denote variances, ϕ is the normal density function given by $\phi(t|\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{t^2}{2\sigma^2}}$, and ξ denotes mixing proportion. We assume here that