

Statistical Analysis of Bivariate Current Status Data with Informative Censoring Using Frailty Effects

Yang-Jin Kim¹

¹Department of Statistics, Sookmyung Women's University

(Received November 2011; Revised January 2012; Accepted January 2012)

Abstract

In animal tumorigenicity data, tumor onsets occur at several sites and onset times cannot be exactly observed. Instead, the existence of tumors is examined only at death time or sacrifice time of the animal. Such an incomplete data structure makes it difficult to investigate the effect of treatment on tumor onset times; in addition, such dependence should be considered when censoring due to death is related with tumor onset. A bivariate frailty effect is incorporated to model bivariate tumor onsets and to connect death with tumor. For the inference of parameters, EM algorithm is applied and a real NTP(National Toxicology Program) dataset is analyzed as an illustrative example.

Keywords: Bivariate current status data, bivariate frailty effect, informative censoring, tumorigenicity.

1. Introduction

Current status data occur when subject is not under continuous examinations, instead, the occurrence of event is checked at only one examination time. This kind of data is also known as case I interval-censored data and was reviewed in detail by Sun (2006). In this paper, our interest is to consider bivariate current status data with informative censoring. For example, in a carcinogenicity study, tumors can occur at two different sites. However, instead of exact tumor onset times, the available data for the i^{th} subject are composed of $(C_i, D_i, \Delta_{i1}, \Delta_{i2})$, where C_i denotes a censoring time at the death or sacrifice time and D_i has a value one for a natural death and zero for sacrifice. Δ_{i1} and Δ_{i2} are indicator variables showing that tumor 1 and tumor 2 are found at C_i , respectively. Most studies assume that the censoring time is independent of the event occurrence time. For example, when the censoring time is predetermined, this assumption is valid. In the

This research was supported by Sookmyung Women's University Research Grants 2011.

¹Assistant Professor, Department of Statistics, Sookmyung Women's University, Seoul 122-807, Korea.

E-mail: yjin@sookmyung.ac.kr

tumorigenicity study, however, the censoring can happen at death and this requires investigation of the relation between death occurrence and tumor onset. As related studies, a three-state model with no tumor, tumor and death state has been applied to univariate current status data (Lindsey and Ryan, 1994; French and Ibrahim, 2002; Kim *et al.*, 2011). They constructed a likelihood for four possible outcomes depending on tumor onset and death. Dunson and Dinse (2002) considered informative censoring for multivariate current status data and developed a MCMC method under conditional independence using frailty effect. Chen *et al.* (2009) and Hens *et al.* (2009) applied a univariate frailty effect for multivariate current status data without death. In this paper, we extend the three state model to analyze bivariate current status data with informative censoring and assume the occurrence of two events are conditionally independent given a bivariate frailty.

We introduce notations and related likelihood in Section 2 and an estimation procedure is developed in detail in Section 3. The suggested method is applied to a real dataset in Section 4 and related comments are remarked in Section 5.

2. Data and Likelihood

We extend a three state model into a multi-state model in order to model bivariate current status data with informative censoring. For the $i (= 1, \dots, n)^{th}$ subject, let X_{ij} denote the $j (= 1, 2)^{th}$ unobservable tumor occurrence time and $\Delta_{ij} = I(X_{ij} \leq C_i)$ indicates whether the j^{th} tumor occurs before censoring time. Hereafter, lowercase letters are applied for observed data, $(c_i, d_i, \delta_{i1}, \delta_{i2}, z_i)$. There are two features to consider for modeling in our study. The first one is the association between tumor onsets. Even though their exact onset times are unobservable, their occurrences are considered in the context of multivariate survival times. Therefore, a bivariate frailty effect is adopted to model such association. That is, given a bivariate frailty $u_i = (u_{i1}, u_{i2})$, the conditional independence between two tumor occurrences is assumed. With covariate vector $z_i (r \times 1)$, the hazard functions of tumor 1 and tumor 2 are defined, respectively,

$$\alpha_{i1}(s|u_{i1}) = \alpha_0(s)\exp(z_i'\beta_1 + u_{i1}), \quad \alpha_{i2}(s|u_{i2}) = \alpha_0(s)\exp(z_i'\beta_2 + u_{i2}),$$

where $u_i = (u_{i1}, u_{i2})$ is assumed to follow a bivariate normal distribution with zero means and the following variance-covariance,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

The second feature of this data is the relation between tumor onset and death occurrence. In the presence of tumors, we assume that tumor 1 and tumor 2 have different lethality and the hazards of death is denoted by

$$\lambda_{i1}(t|u_{i1}) = \lambda_0(t)\exp(z_i'\gamma_1 + u_{i1}), \quad \lambda_{i2}(t|u_{i2}) = \lambda_0(t)\exp(z_i'\gamma_2 + u_{i2}).$$

Also, it is still possible for death to occur without tumors and corresponding hazards are given as

$$\lambda_{i3}(t) = \lambda_0(t)\exp(z_i'\gamma_3).$$

Table 2.1. Likelihood for (δ_1, δ_2, d)

(δ_1, δ_2, d)	Likelihood
$(0, 0, 0) + (0, 0, 1)$	$[e^{-\Lambda_3(0,c) - A_1(0,c) - A_2(0,c)}]^{(1-\delta_1)(1-\delta_2)} \lambda_3(c)^{(1-\delta_1)(1-\delta_2)d}$
$(1, 0, 0) + (1, 0, 1)$	$[e^{-\Lambda_3(0,c) - A_1(0,x_1) - A_2(0,c)} \alpha_1(x_1) e^{-\Lambda_1(x_1,c)}]^{(1-\delta_1)(1-\delta_2)} \lambda_1(c)^{\delta_1(1-\delta_2)d}$
$(0, 1, 0) + (0, 1, 1)$	$[e^{-\Lambda_3(0,c) - A_1(0,c) - A_2(0,x_2)} \alpha_2(x_2) e^{-\Lambda_2(x_2,c)}]^{(1-\delta_1)\delta_2} \lambda_2(c)^{(1-\delta_1)\delta_2 d}$
$(1, 1, 0) + (1, 1, 1)$	$[e^{-\Lambda_3(0,c) - A_1(0,x_1) - A_2(0,x_2)} \alpha_1(x_1) \alpha_2(x_2) e^{-\Lambda_1(x_1,c) - \Lambda_2(x_2,c)}]^{(1-\delta_1)\delta_2} (\lambda_1(c) \lambda_2(c))^{\delta_1 \delta_2 d}$

For the convenience, the subject script is dropped hereafter. A likelihood is composed of four parts depending on tumor existence and death occurrence shown in Table 2.1. $A_1, A_2, \Lambda_1, \Lambda_2$ and Λ_3 are the cumulative hazard functions of $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ and λ_3 , respectively. Here, the cumulative hazard functions are defined as

$$A_k(s, t) = \int_s^t \alpha_k(v|u_k)dv, \quad \Lambda_k(s, t) = \int_s^t \lambda_k(w|u_k)dw, \quad k = 1, 2, \quad \text{and} \quad \Lambda_3(s, t) = \int_s^t \lambda_3(w)dw.$$

Then a likelihood is written as,

$$L_c = \prod_{i=1}^n e^{-\Lambda_3(0,c_i)} \lambda_3(c_i)^{(1-\delta_{i1})(1-\delta_{i2})d_i} e^{-A_1(0,c_i)(1-\delta_{i1})} e^{-A_2(0,c_i)(1-\delta_{i2})} \\ \times \left\{ e^{-A_1(0,x_{i1}) - \Lambda_1(x_{i1},c_i)} \alpha_1(x_{i1}) \right\}^{\delta_{i1}} \left\{ e^{-A_2(0,x_{i2}) - \Lambda_2(x_{i2},c_i)} \alpha_2(x_{i2}) \right\}^{\delta_{i2}} \lambda_1(c_i)^{\delta_{i1} d_i} \lambda_2(c_i)^{\delta_{i2} d_i}.$$

For easy and robust estimation of baseline functions, we adopt a piecewise constant hazard function. Let $0 = s_0 < s_1 < \dots < s_J$ denote pre-specified breakpoints resulting in the following intervals, $I_j = (s_{j-1}, s_j]$, $j = 1, \dots, J$ which have constant intensities. For a baseline tumor onset hazard $\alpha_0(t)$, define,

$$\alpha_0(t) = \tilde{\alpha}_l, \quad t \in (s_{l-1}, s_l].$$

Similar definitions are assigned for baseline death hazard $\tilde{\lambda} = \{\tilde{\lambda}_1, \dots, \tilde{\lambda}_J\}$ corresponding $\lambda_0(t)$. Thus, $(\tilde{\alpha}, \tilde{\lambda}) = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_J, \tilde{\lambda}_1, \dots, \tilde{\lambda}_J)$ are unknown baseline intensity parameters. Let $\theta = (\tilde{\lambda}, \tilde{\alpha}, \beta, \gamma, \Sigma)$ be the parameters of interest. Then a complete data likelihood is redefined as

$$L(\theta) = \prod_{j=1}^J \tilde{\lambda}_j^{NB_j + a_j + b_j} \tilde{\alpha}_j^{B_{1j} + B_{2j}} \\ \times \exp \left(-\tilde{\lambda}_j \sum_{i=1}^n \left(T_{ij}^{NB T} e^{z'_i \gamma_3} + T_{ij}^{B1} e^{z'_i \gamma_1 + u_{i1}} + T_{ij}^{B2} e^{z'_i \gamma_2 + u_{i2}} \right) \right. \\ \left. - \tilde{\alpha}_j \sum_{i=1}^n \left(T_{ij}^{NB1} e^{z'_i \beta_1 + u_{i1}} + T_{ij}^{NB2} e^{z'_i \beta_2 + u_{i2}} \right) \right) \\ \times \exp \left\{ \sum_{i=1}^n z'_i [(1 - \delta_{1i})(1 - \delta_{2i})d_i \gamma_3 + \delta_{1i} \beta_1 + \delta_{2i} \beta_2 + \gamma_1 \delta_{1i} d_i + \gamma_2 \delta_{2i} d_i] \right. \\ \left. + \sum_{i=1}^n [\delta_{1i}(1 + d_i)u_{i1} + \delta_{2i}(1 + d_i)u_{i2}] \right\},$$

where $(NB_j, a_j, b_j, B_{1j}, B_{2j}, T_{ij}^{NB T}, T_{ij}^{B1}, T_{ij}^{B2}, T_{ij}^{NB1}, T_{ij}^{NB2})$ are defined as,

NB_j : the number of subjects dying without both tumors at I_j ,

a_j, b_j : the number of subjects dying with the tumor 1 and tumor 2 at I_j ,

$B1_j, B2_j$: the numbers of tumor 1 and tumor 2 onsets at I_j ,

$T_{ij}^{NB1}, T_{ij}^{NB2}$: the times the i^{th} subject spends without tumor 1 and tumor 2 at I_j , respectively,

T_{ij}^{B1}, T_{ij}^{B2} : the times the i^{th} subject spends with tumor 1 and tumor 2 at I_j , respectively,

$T_{ij}^{NB1}, T_{ij}^{NB2}$: the times the i^{th} subject spends without tumor 1 and tumor 2 at I_j , respectively,

where NB_j, a_j, b_j can be calculated while $B1_j, B2_j, T_{ij}^{NB1}, T_{ij}^{B1}, T_{ij}^{NB2}$ and T_{ij}^{B2} are uncalculable owing to unobservable tumor onset times.

3. EM Algorithm

We apply the EM algorithm (Dempster *et al.*, 1977) to estimate unknown parameters. There are two types of missing information caused by both frailty effects and current status data. Therefore, a two-stage procedure is applied in order to recover unknown quantities. In the first stage, the frailty effects are assumed to be known and then unknown quantities related with unobservable tumor onset times are estimated using the following conditional expectations with $O_i = (\delta_{i1}, \delta_{i2}, c_i, z_i)$ (Lindsey and Ryan, 1994),

$$\begin{aligned}
 E(B1_j | O_i, \hat{\theta}) &= \sum_{i=1}^n \delta_{1i} p_{1j}(c_i), \\
 E(B2_j | O_i, \hat{\theta}) &= \sum_{i=1}^n \delta_{2i} p_{2j}(c_i), \\
 E(T_{ij}^{NB1} | O_i, \hat{\theta}) &= \begin{cases} I(c_i \in I_j)(c_i - s_{j-1}) + \sum_{I(c_i \in I_k, k > j)} (s_j - s_{j-1}), & \delta_{1i} \neq 1, \\ I(c_i \in I_j)(E_j(x_1 | c_i) - s_{j-1})p_{1j}(c_i) + \sum_{I(c_i \in I_k, k > j)} (s_j - s_{j-1})p_{1j}(c_i), & \delta_{1i} = 1, \end{cases} \\
 E(T_{ij}^{NB2} | O_i, \hat{\theta}) &= \begin{cases} I(c_i \in I_j)(c_i - s_{j-1}) + \sum_{I(c_i \in I_k, k > j)} (s_j - s_{j-1}), & \delta_{2i} \neq 1, \\ I(c_i \in I_j)(E_j(x_2 | c_i) - s_{j-1})p_{2j}(c_i) + \sum_{I(c_i \in I_k, k > j)} (s_j - s_{j-1})p_{2j}(c_i), & \delta_{2i} = 1, \end{cases}
 \end{aligned}$$

where $p_{1j}(c_i)$ is the conditional probability of the i^{th} subject acquired tumor 1 in I_j , given it died or sacrificed at c_i .

$$p_{1j}(c_i) = \begin{cases} \frac{\int_{s_{j-1}}^{c_i} q(x, c_i) dx}{\int_0^{c_i} q(x, c_i) dx}, & c_i \in I_j, \\ \frac{\int_{s_{j-1}}^{s_j} q(x, c_i) dx}{\int_0^{c_i} q(x, c_i) dx}, & s_j < c_i, \end{cases} \quad E_j(x_1 | c_i) = \begin{cases} \frac{\int_{s_{j-1}}^{c_i} xq(x, c_i) dx}{\int_{s_{j-1}}^{s_j} q(x, c_i) dx}, & c_i \in I_j, \\ \frac{\int_{s_{j-1}}^{s_j} xq(x, c_i) dx}{\int_{s_{j-1}}^{c_i} q(x, c_i) dx}, & s_j < c_i, \end{cases}$$

where $q(x_1, c_i) = \alpha_1(x_1) \exp[-\int_0^{x_1} (\alpha_1(s) + \lambda_3(s)) ds] \lambda_1(c_i) \exp[-\int_{x_1}^{c_i} \lambda_1(s) ds]$. For tumor 2, p_{2j} and $E_j(x_2 | c_i)$ are defined similarly. Therefore, $T_{ij}^{B1} = T_{ij}^{tot} - T_{ij}^{NB1}$, $T_{ij}^{B2} = T_{ij}^{tot} - T_{ij}^{NB2}$ and

$T_{ij}^{NBT} = T_{ij}^{tot} - \max(T_{ij}^{B1}, T_{ij}^{B2})$. Once the missing values are replaced by the expected values, the next stage is to calculate the functions of the frailties based on the following conditional probabilities,

$$E(h(u_i)|O_i, \theta) = \frac{\int h(u_i)L_i(\theta)f(u_i|\Sigma)du_i}{\int L_i(\theta)f(u_i|\Sigma)du_i},$$

where

$$\begin{aligned} L_i(\theta) = & \prod_{j=1}^J \tilde{\lambda}_j^{d_i(1-\delta_{1i})(1-\delta_{2i})I(c_i \in S_j) + d_i\delta_{1i}I(c_i \in S_j) + d_i\delta_{2i}I(c_i \in S_j)} \times \tilde{\alpha}_j^{\delta_{1i}p_{1j}(c_i) + \delta_{2i}p_{2j}(c_i)} \\ & \times \exp\left[-\tilde{\lambda}_j\left(T_{ij}^{NBT}e^{z_i\gamma_3} + T_{ij}^{B1}e^{z_i\gamma_1+u_{i1}} + T_{ij}^{B2}e^{z_i\gamma_2+u_{i2}}\right) - \tilde{\alpha}_j\left(T_{ij}^{NB1}e^{z_i\beta_1+u_{i1}} + T_{ij}^{NB2}e^{z_i\beta_2+u_{i2}}\right)\right] \\ & \times \exp\left\{z_i[(1-\delta_{1i})(1-\delta_{2i})d_i\gamma_3 + \delta_{1i}\beta_1 + \delta_{2i}\beta_2 + \delta_{1i}d_i\gamma_1 + \delta_{2i}d_i\gamma_2] + \delta_{1i}(1+d_i)u_{i1} + \delta_{2i}(1+d_i)u_{i2}\right\}. \end{aligned}$$

The calculation of integration can be done by Gaussian quadrature techniques (Abramowitz and Stegun, 1972). In M-step to obtain updated estimates, we maximize the following log likelihood replaced with the expectations of quantities calculated in E-step,

$$\begin{aligned} l(\theta) = & \sum_{j=1}^J \left[(NB_j + a_j + b_j)\log\tilde{\lambda}_j + (B1_j + B2_j)\log\tilde{\alpha}_j \right. \\ & \left. - \tilde{\lambda}_j \sum_{i=1}^n T_{ij}^{NBT} e^{z_i\gamma_3} + T_{ij}^{B1} e^{z_i\gamma_1+u_{i1}} + T_{ij}^{B2} e^{z_i\gamma_2+u_{i2}} - \tilde{\alpha}_j \sum_{i=1}^n \left(T_{ij}^{NB1} e^{z_i\beta_1+u_{i1}} + T_{ij}^{NB2} e^{z_i\beta_2+u_{i2}} \right) \right] \\ & + \sum_{i=1}^n \left[(1-\delta_{1i})(1-\delta_{2i})d_i\gamma_3 + \delta_{1i}\beta_1 + \delta_{2i}\beta_2 + \delta_{1i}d_i\gamma_1 + \delta_{2i}d_i\gamma_2 + \delta_{1i}(1+d_i)u_{i1} + \delta_{2i}(1+d_i)u_{i2} \right]. \end{aligned}$$

We use a Newton-Raphson method to estimate parameters θ . The first and second derivatives are defined as follows, $j = 1, \dots, J$,

$$\begin{aligned} \frac{\partial l}{\partial \tilde{\alpha}_j} &= \frac{B1_j + B2_j}{\tilde{\alpha}_j} - \sum_{i=1}^n \left(T_{ij}^{NB1} e^{z_i\beta_1+u_{i1}} + T_{ij}^{NB2} e^{z_i\beta_2+u_{i2}} \right), \\ \frac{\partial l}{\partial \tilde{\lambda}_j} &= \frac{NB_j + a_j + b_j}{\tilde{\lambda}_j} - \sum_{i=1}^n \left(T_{ij}^{NBT} e^{z_i\gamma_3} + T_{ij}^{B1} e^{z_i\gamma_1+u_{i1}} + T_{ij}^{B2} e^{z_i\gamma_2+u_{i2}} \right), \\ \frac{\partial l}{\partial \beta_1} &= \sum_{i=1}^n z_i \left[- \sum_{j=1}^J \tilde{\alpha}_j T_{ij}^{NB1} e^{z_i\beta_1+u_{i1}} + \delta_{1i} \right], \\ \frac{\partial l}{\partial \beta_2} &= \sum_{i=1}^n z_i \left[- \sum_{j=1}^J \tilde{\alpha}_j T_{ij}^{NB2} e^{z_i\beta_2+u_{i2}} + \delta_{2i} \right], \\ \frac{\partial l}{\partial \gamma_1} &= \sum_{i=1}^n z_i \left[- \sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{B1} e^{z_i\gamma_1+u_{i1}} + \delta_{1i}d_i \right], \\ \frac{\partial l}{\partial \gamma_2} &= \sum_{i=1}^n z_i \left[- \sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{B2} e^{z_i\gamma_2+u_{i2}} + \delta_{2i}d_i \right], \\ \frac{\partial l}{\partial \gamma_3} &= \sum_{i=1}^n z_i \left[- \sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{NBT} e^{z_i\gamma_3} + (1-\delta_{1i})(1-\delta_{2i})d_i \right], \end{aligned}$$

Table 4.1. Summary of tumor occurrence of 100 male rats in the NTP data

Treat	No Tumor	Adrenal Tumor	Lung Tumor	Both Tumor
Control	29	19	2	0
High dose	31	13	1	3

and the second derivatives are

$$\begin{aligned}
\frac{\partial^2 l}{\partial \tilde{\alpha}_j^2} &= -\frac{B1_j + B2_j}{\tilde{\alpha}_j^2}, & \frac{\partial^2 l}{\partial \tilde{\lambda}_j^2} &= -\frac{NB_j + a_j + b_j}{\tilde{\lambda}_j^2}, \\
\frac{\partial^2 l}{\partial \tilde{\alpha}_j \partial \beta_1} &= -\sum_{i=1}^n z_i \left(T_{ij}^{NB1} e^{z'_i \beta_1 + u_{i1}} \right), & \frac{\partial^2 l}{\partial \tilde{\alpha}_j \partial \beta_2} &= -\sum_{i=1}^n z_i \left(T_{ij}^{NB2} e^{z'_i \beta_2 + u_{i2}} \right), \\
\frac{\partial^2 l}{\partial \tilde{\lambda}_j \partial \gamma_1} &= -\sum_{i=1}^n z_i \left(T_{ij}^{B1} e^{z'_i \gamma_1 + u_{i1}} \right), & \frac{\partial^2 l}{\partial \tilde{\lambda}_j \partial \gamma_2} &= -\sum_{i=1}^n z_i \left(T_{ij}^{B2} e^{z'_i \gamma_2 + u_{i2}} \right), \\
\frac{\partial^2 l}{\partial \tilde{\lambda}_j \partial \gamma_3} &= -\sum_{i=1}^n z_i \left(T_{ij}^{NBT} e^{z'_i \gamma_3} \right), \\
\frac{\partial^2 l}{\partial \beta_1^2} &= -\sum_{i=1}^n \left[\sum_{j=1}^J \tilde{\alpha}_j T_{ij}^{NB1} e^{z'_i \beta_1 + u_{i1}} \right] z_i z'_i, & \frac{\partial^2 l}{\partial \beta_2^2} &= -\sum_{i=1}^n \left[\sum_{j=1}^J \tilde{\alpha}_j T_{ij}^{NB2} e^{z'_i \beta_2 + u_{i2}} \right] z_i z'_i, \\
\frac{\partial^2 l}{\partial \gamma_1^2} &= -\sum_{i=1}^n \left[\sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{B1} e^{z'_i \gamma_1 + u_{i1}} \right] z_i z'_i, & \frac{\partial^2 l}{\partial \gamma_2^2} &= -\sum_{i=1}^n \left[\sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{B2} e^{z'_i \gamma_2 + u_{i2}} \right] z_i z'_i, \\
\frac{\partial^2 l}{\partial \gamma_3^2} &= -\sum_{i=1}^n \left[\sum_{j=1}^J \tilde{\lambda}_j T_{ij}^{NBT} e^{z'_i \gamma_3} \right] z_i z'_i.
\end{aligned}$$

For the variance-covariance estimation of the estimator $\hat{\theta}$, the inverse of the observed information matrix, $I(\hat{\theta})$, is used. The estimation includes the calculation of piece-wise baseline parameters whose length can be increased with sample size. To find appropriate number of pieces and to determine break points, several evaluation criterions can be used. However, before applying these procedures, it is more crucial to carefully investigate dataset to find meaningful numbers of pieces and break points related to the research.

4. Data Analysis

In this section, we apply the suggested method to data from a NTP(National Toxicology Program), 2 year rodent carcinogenicity study of chloroprene (Dunson and Dinse, 2002). The occurrence of tumors found in several sites was investigated at the death time or sacrifice time. The dataset composed of 100 mice. A total of 50 mice were in the control group and 50 were exposed to chloroprene at concentration of 80 ppm by inhalation 6 hours per day. The main interest is to assess the effect of treatment on tumor incidence and death. Furthermore, tumor incidences at two difference sites may have inter-correlation. Table 4.1 shows the distribution of tumors by treatment. The first death with tumor occurs at 20 week and there is only one sacrifice time at 25 week. Denote $z_i = 1$ as a high dose group and 0 otherwise. Table 4.2 shows the estimated coefficients and standard errors obtained from the suggested method with several different break

Table 4.2. Analysis of NTP data with three different numbers of pieces

	$J = 3$	$J = 4$	$J = 5$
Tumor incidence			
Adrenal	0.723(0.234)	0.800(0.236)	0.768(0.236)
Lung	-0.356(0.408)	-0.443(0.408)	-0.404(0.407)
Death			
Adrenal	1.105(0.267)	0.894(0.267)	0.879(0.268)
Lung	1.095(0.501)	0.944(0.500)	1.008(0.500)
No tumor	0.063(0.178)	0.067(0.179)	0.079(0.180)
Random effect = $(\sigma_1^2, \sigma_{12}, \sigma_2^2)$			
	(0.75, 0.52, 0.42)	(0.53, 0.39, 0.35)	(0.16, 0.09, 0.09)
log-likelihood	-311.052	-278.021	-301.677

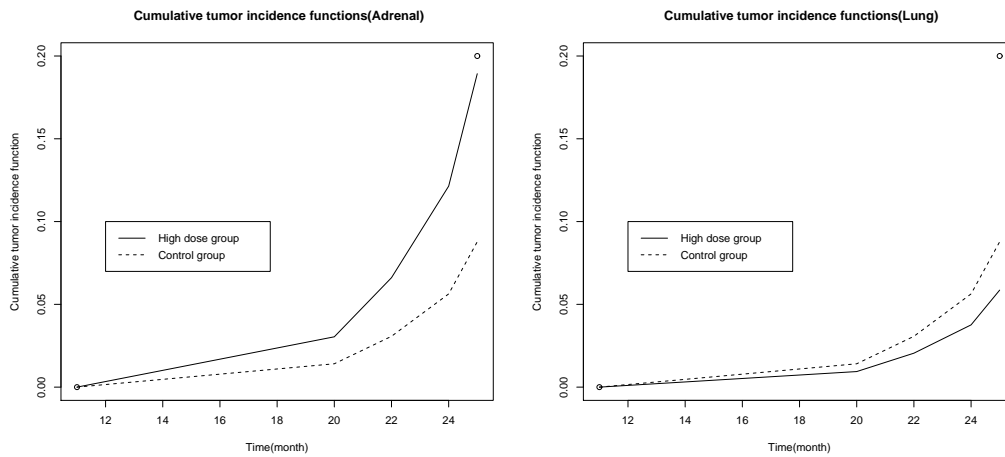


Figure 4.1. Estimated baseline tumor hazard functions

points ($J = 3, 4, 5$). Based on log-likelihood values, $J = 4$ seems to be appropriate. With four time points ($J = 4, s_1 = 20 < 22 < 24 < 25 = s_4$) for piecewise baseline hazards, several covariate effects and frailty effects are estimated. $\hat{\beta}_1 = 0.800$ (p -value = 0.0006) indicates that a treatment has a significant effect on adrenal tumor incidence while $\hat{\beta}_2 = -0.443$ (p -value = 0.278) shows a treatment has no significant effect on lung tumor incidence. From the estimated $\hat{\gamma}$ s values related with death, treatment has significant effects for deaths of patients with tumors ($\hat{\gamma}_1 = 0.894, \hat{\gamma}_2 = 0.944$). That is, mice with high dose tend to get a higher risk on death after the tumor onsets. However, treatment has no significant effect on death without tumors ($\hat{\gamma}_3 = 0.067$). Figure 4.1 and Figure 4.2 show the estimated baseline tumor hazards and death hazards with $J = 4$. Also the estimated covariance $\hat{\sigma}_{12} = 0.39$ shows a positive correlation between Adrenal and lung tumors.

5. Conclusion

In this study, an EM algorithm is applied to investigate covariate effects at bivariate current status with informative censoring. Using a bivariate frailty, the associations between two tumor onsets as

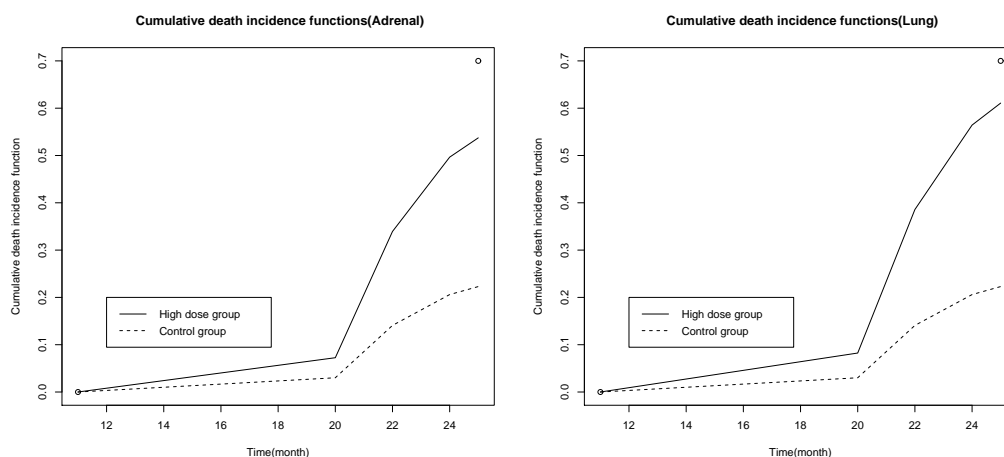


Figure 4.2. Estimated baseline death hazard functions

well as between tumor onset and death are considered. However, the present method still needs future research. One future work is to model transition between tumors. We assume that two tumors occur independently given a bivariate frailty. Bivariate current status data is unavailable for the order of occurrence as well as the times. In a paper with similar interest, Cook *et al.* (2008) considered bivariate interval-censored data and applied a Markov process in order to connect two processes. In their work, overlapping intervals of two interval-censored data complicate the estimation procedure. They assigned appropriate weights on regions composed of interval censored times in order to derive estimating equations. However, there are two problems to apply their methods to our dataset; less information owing to a common censoring time and the possibility of informative censoring. Therefore, a more complicated estimation procedure is required for appropriate estimates. In future work, more extended multi-state model would be incorporated to model two events simultaneously and connect them. As the referees pointed out, future work would develop a simulation scheme and check the performance of the suggested method under several conditions. In particular, these enable us to give guidelines to decide the suitable number of pieces. We will discuss this problem in future work.

References

- Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions*, Dover, New York.
- Chen, M., Tong, X. and Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data, *Statistics in Medicine*, **28**, 3424–3436.
- Cook, R., Zeng, L. and Lee, K. (2008). A multistate model for bivariate interval-censored failure time data, *Biometrics*, **64**, 1100–1109.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of Royal Statistical Society, Series B*, **39**, 1–38.
- Dunson, D. and Dinse, G. E. (2002). Bayesian models for multivariate current status data with informative censoring, *Biometrics*, **58**, 79–88.
- French, J. L. and Ibrahim, J. G. (2002). Bayesian methods for a three-state model for rodent carcinogenicity studies, *Biometrics*, **58**, 906–916.

- Hens, N., Wienke, A., Aerts, M. and Molenberghs, G. (2009). The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data, *Statistics in Medicine*, **28**, 2785–2800.
- Kim, Y. J., Nam, J. M., Kim, Y. N., Choi, E. H. and Kim, J. H. (2011). The analysis of tumorigenicity data using a frailty effect, *Journal of Korean Statistical Society*, **40**, 281–290.
- Lindsey, J. C. and Ryan, L. M. (1994). A comparison of continuous- and discrete time three state models for rodent tumorigenicity experiments, *Environmental Health Perspective Supplements*, **102**, 9–17.
- Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*, Springer-Verlag, New-York.