The Analysis of Tumorigenicity Data using a Normal Frailty Effect

Jinheum Kim[1]

[1]Department of Applied Statistics, University of Suwon

May 21, 2010

Outline

- Interval censoring
- Issues on the current status data
- Three-state model with normal frailty
- Likelihood
- EM estimation
- Simulation results
- Discussion

Interval censoring

- Four types of interval-censored data
  - Case I interval-censored data, called by current status data
  - Case II interval-censored data, in short interval-censored data
  - Doubly interval-censored data (*cf* Double censoring)
  - Panel count data (*cf* Recurrent event data)
-
- *Reference:* *Sun(2006, The Statistical Analysis of Interval-censored Failure Time Data)*

Current status data

- $T_i, i = 1, \ldots, n$: Survival time
- Data: $\{(C_i, I(T_i \leq C_i), \mathbf{Z}_i); i = 1, \ldots, n\}$
  - C_i : Examination (Observation) time such as death time or sacrifice time
  - Only left-censored, $T_i \in (0, C_i)$, or right-censored, $T_i \in (C_i, \infty)$
  - Assume T_i and C_i are independent given covariate $\mathbf{Z}_i$
- Eg., Tumorigenicity experiments

Issues

- When the examination time would be predetermined, the independent assumption is appropriate
- BUT, in tumorigenicity study, the censoring happens either at death or at sacrifice
- When examination is made with naturally dead animal, this death may be related with tumor onset and the independence assumption is not valid any more

Background

- With non-lethal tumor, death would be independent with tumor onset
- For lethal tumor, two cases can be considered:
  - With a rapidly lethal tumor, the death time follows the onset of tumor and log-rank test using death times is used to compare treatments on tumor onset time
  - With an intermediate lethal tumor, the correlation between the tumor onset time and death time is not known
-
- *Reference: Lagakos and Louis (1988, Applied Statistics)*

Three-state model

- Three states: Health, tumor, and death
- The lethality was measured by the ratio of two death hazards, i.e., death with tumor and death without tumor, and was composed of baseline lethality and treatment lethality.
-
- *Reference: Lindsey and Ryan (1993, Applied Statistics; 1994, Environmental Health Perspective Supplements) and French and Ibrahim (2002, Biometrics)*

Schematic diagram

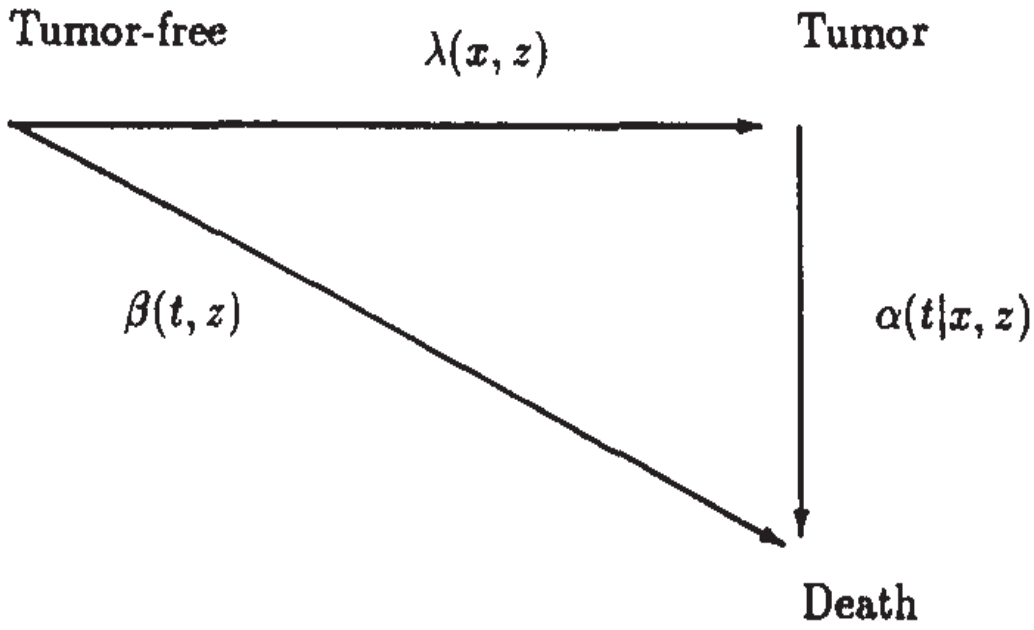

Figure: Three-state illness model

Our way

- Estimate the effect of treatment on the tumor onset time and the death time considering the possible correlation between tumor onset and death
- Introduce a normal frailty to incorporate the dependency

Data

- X : Tumor onset time
  - NOT directly observed, instead, only know whether or not tumor onset occurred before an examination time (T)
  - $\Delta = 1$ if $X < T$ and $\Delta = 0$ otherwise
- For the examination time, there are two possible cases: natural death time (T_1) and sacrifice time (T_2)
  - $T = \min(T_1, T_2)$: Observable examination time
  - $D = I(T_1 < T_2)$
  - Assume that T_2 is independent of X and T_1
- Observable data: $\{(t_i, \delta_i, d_i, \mathbf{z}_i); i = 1, \ldots, n\}$

Three models

- Hazard rate of tumor onset:

$$\alpha_i(x|\mathbf{z}_i, r_i) = \alpha_0(x)\exp(\boldsymbol{\beta}'\mathbf{z}_i + r_i), r_i \sim N(0, \sigma^2) \quad (1)$$

- Hazard rate of death without tumor onset:

$$\tilde{\lambda}_i(t|\mathbf{z}_i) = \lambda_0(t)\exp(\boldsymbol{\psi}'\mathbf{z}_i) \quad (2)$$

- Hazard rate of death with tumor onset:

$$\lambda_i(t|\mathbf{z}_i, r_i) = \lambda_0(t)\exp\{(\boldsymbol{\psi} + \boldsymbol{\gamma})'\mathbf{z}_i + \tau r_i\} \quad (3)$$

  - τ : A role to connect tumor onset and death due to frailty
    - If $\tau > 0$, an animal with higher frailty on tumor onset will result in earlier death
  - Unlike Lindsey and Ryan (1993, 1994) and French and Ibrahim (2002), our model assumes that the lethality due to tumor onset is animal-specific rather than same for all animals

Preliminary results

Table: Estimated mean, bias, empirical standard error (SE), mean of the 1,000 estimated standard errors (SEM), and 95% coverage rate (CR) for the parameters β, ψ, γ, and η ($\exp(\eta)$: lethality parameter) in LR model (Lindsey and Ryan, 1993) with sacrifice times of 12, 18, and 33

Parameter	True	Mean	Bias	SE	SEM	CR
β	3.2463	3.4574	0.0650	0.6666	0.5364	92.3
ψ	-0.1493	-0.1467	0.0173	0.5483	0.5215	95.7
γ	-3.4780	-2.5003	0.2811	1.3463	0.7943	65.6
η	4.7831	3.7605	0.2138	1.2167	0.5641	49.6

- The coverage rates of γ and η among the regression parameters in LR model are fairly lower than a nominal level due to ignoring the dependency

Four possible factors: With no tumor cases

- For sacrifice with no tumor (SNT), $(\delta_i, d_i) = (0, 0)$,

$$L_{i1} = \exp\{-\int_0^{t_i} \{\alpha_i(u|\mathbf{z}_i, r_i) + \tilde{\lambda}_i(u|\mathbf{z}_i)\} du\};$$

- For death with no tumor (DNT), $(\delta_i, d_i) = (0, 1)$,

$$L_{i2} = \tilde{\lambda}_i(t_i|\mathbf{z}_i) \exp\{-\int_0^{t_i} \{\alpha_i(u|\mathbf{z}_i, r_i) + \tilde{\lambda}_i(u|\mathbf{z}_i)\} du\};$$

Four possible factors: With tumor cases

- For sacrifice with tumor (SWT), $(\delta_i, d_i) = (1, 0)$,

$$L_{i3} = \int_0^{t_i} f_X(u|\mathbf{z}_i, r_i)\tilde{S}_T(u|\mathbf{z}_i)\frac{S_T(t_i|\mathbf{z}_i, r_i)}{S_T(u|\mathbf{z}_i, r_i)}du;$$

  - $S_X, \tilde{S}_T$ and S_T : Survival functions corresponding to the three models respectively
  - $f_X(x|\mathbf{z}_i, r_i) = \alpha_i(x|\mathbf{z}_i, r_i)S_X(x|\mathbf{z}_i, r_i)$

- For death with tumor (DWT), $(\delta_i, d_i) = (1, 1)$,

$$L_{i4} = \lambda_i(t_i|z_i, r_i)\int_0^{t_i} f_X(u|\mathbf{z}_i, r_i)\tilde{S}_T(u|\mathbf{z}_i)\frac{S_T(t_i|\mathbf{z}_i, r_i)}{S_T(u|\mathbf{z}_i, r_i)}du$$

Complete data

- Complete data: $\{(x_i, t_i, \delta_i, d_i, \mathbf{z}_i); i = 1, \ldots, n\}$
  - x_i : Tumor onset time
- Based on the complete data, L_{i3} and L_{i4} can be represented as, respectively,

$$\tilde{L}_{i3} = \alpha_i(x_i|\mathbf{z}_i, r_i) \exp\{-\int_0^{x_i} \{\alpha_i(u|\mathbf{z}_i, r_i) + \tilde{\lambda}_i(u|\mathbf{z}_i)\} du$$

$$- \int_{x_i}^{t_i} \lambda_i(u|\mathbf{z}_i, r_i) du\};$$

$$\tilde{L}_{i4} = \lambda_i(t_i|\mathbf{z}_i, r_i)\alpha_i(x_i|\mathbf{z}_i, r_i) \exp\{-\int_0^{x_i} \{\alpha_i(u|\mathbf{z}_i, r_i) + \tilde{\lambda}_i(u|\mathbf{z}_i)\} du$$

$$- \int_{x_i}^{t_i} \lambda_i(u|\mathbf{z}_i, r_i) du\}$$

Likelihood

- Complete-data-based likelihood:

$$L_c = \prod_{i=1}^{n} L_{i1}^{(1-\delta_i)(1-d_i)} L_{i2}^{(1-\delta_i)d_i} \tilde{L}_{i3}^{\delta_i(1-d_i)} \tilde{L}_{i4}^{\delta_i d_i} \phi(r_i|\sigma^2)$$

  - $\phi(r|\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-(2\sigma^2)^{-1} r^2\}$

Piecewise exponential model

- Following Lindsey and Ryan (1993, 1994) and French and Ibrahim (2002), suppose that there are J time points, $s_1, \ldots, s_J$, and that

$$\alpha_0(t) = \alpha_{0j}, \quad t \in I_j = (s_{j-1}, s_j], j = 1, \ldots, J;$$

$$\lambda_0(t) = \lambda_{0j}, \quad t \in I_j$$

Likelihood

- Complete-data-based log-likelihood: Letting $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\lambda}', \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\gamma}, \tau, \sigma^2)'$, $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}', \boldsymbol{\beta})'$, $\boldsymbol{\theta}_2 = (\boldsymbol{\lambda}', \boldsymbol{\psi}, \boldsymbol{\gamma}, \tau)'$, and $\theta_3 = \sigma^2$,

$$l_c = \log(L_c) = l_{\theta_1} + l_{\theta_2} + l_{\theta_3}$$

  - $l_{\theta_1} = \sum_{j=1}^{J} \{N_j^T \log \alpha_j - \alpha_j \sum_{i=1}^{n} \exp(\boldsymbol{\beta}' \mathbf{z}_i + r_i) T_{ij}^{NT}\} + \sum_{i=1}^{n} \mathbf{z}_i' (\delta_i \boldsymbol{\beta})$;
  - $l_{\theta_2} = \sum_{j=1}^{J} \{(a_j + b_j) \log \lambda_j - \lambda_j \sum_{i=1}^{n} \{\exp(\boldsymbol{\psi}' \mathbf{z}_i) T_{ij}^{NT} + \exp((\boldsymbol{\psi} + \boldsymbol{\gamma})' \mathbf{z}_i + \tau r_i) T_{ij}^{T}\}\} + \sum_{i=1}^{n} \{\mathbf{z}_i' (d_i \boldsymbol{\psi} + \delta_i d_i \boldsymbol{\gamma}) + (\delta_i + \tau d_i \delta_i) r_i\}$;
  - $l_{\theta_3} = \sum_{i=1}^{n} \log \phi(r_i | \sigma^2)$

Two kinds of missing terms: Unobservable tumor onset time

- $a_j = \sum_{i=1}^{n}(1-\delta_i)d_{ij}d_i$: Number of death without tumor in I_j
  - $d_{ij} = I(t_i \in I_j)$
- $b_j = \sum_{i=1}^{n}\delta_i d_{ij} d_i$: Number of death with tumor in I_j
- a_j and b_j : Known, BUT
- N_j^T : Number of subjects with tumor in interval I_j
- $T_{ij}^{NT} and T_{ij}^{T}$: Time the ith subject spends with no tumor and with tumor in interval I_j
- N_j^T, T_{ij}^T, and T_{ij}^{NT} : NOT known
- According to Lindsey and Ryan (1993, 1994), calculate $\mathrm{E}(N_j^T|\mathcal{O},\boldsymbol{\theta})$, $\mathrm{E}(T_{ij}^{NT}|\mathcal{O},\boldsymbol{\theta})$ and $\mathrm{E}(T_{ij}^{T}|\mathcal{O},\boldsymbol{\theta})$

Two kinds of missing terms: Frailty

- Conditional on observed data, $\mathcal{O} = \{O_i = (t_i, \delta_i, d_i, \mathbf{z}_i), i = 1, \ldots, n\}$,

$$f_{R_i|\mathcal{O}}(r|\boldsymbol{\theta}) = \frac{L_i(r;\boldsymbol{\theta})}{\int_{-\infty}^{\infty} L_i(u;\boldsymbol{\theta})du}$$

  - $L_i(r;\boldsymbol{\theta}) = L_{i1}^{(1-\delta_i)(1-d_i)} L_{i2}^{(1-\delta_i)d_i} L_{i3}^{\delta_i(1-d_i)} \tilde{L}_{i4}^{\delta_i d_i} \phi(r|\sigma^2)$
- Employ the Gauss-Hermite method for calculating the functionals of frailty such as $\mathrm{E}(R_i|\mathcal{O},\boldsymbol{\theta})$, $\mathrm{E}(R_i^2|\mathcal{O},\boldsymbol{\theta})$, $\mathrm{E}\{\exp(R_i)|\mathcal{O},\boldsymbol{\theta}\}$, $\mathrm{E}\{\exp(\tau R_i)|\mathcal{O},\boldsymbol{\theta}\}$ etc

EM Procedure

- Start with $\boldsymbol{\theta}_j^{(0)}, j = 1, 2, 3$
- For $k = 0, 1, \ldots,$ update $\boldsymbol{\theta}_j^{(k)}$ through

$$\boldsymbol{\theta}_j^{(k+1)} = \boldsymbol{\theta}_j^{(k)} - \{\frac{\partial^2 Q_j(\boldsymbol{\theta}_j|\boldsymbol{\theta}_j^{(k)})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_j'}\}^{-1}_{\theta_j=\theta_j^{(k)}}\{\frac{\partial Q_j(\boldsymbol{\theta}_j|\boldsymbol{\theta}_j^{(k)})}{\partial\boldsymbol{\theta}_j}\}_{\theta_j=\theta_j^{(k)}}, j = 1, 2, 3$$

  - $Q_j(\boldsymbol{\theta}_j|\boldsymbol{\theta}_j^{(k)}) = \mathrm{E}_{\theta_j^{(k)}}(l_{\theta_j}|\mathcal{O}), j = 1, 2, 3$
- Iterate E-M steps until $||\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}||_\infty < \epsilon$
  - $\boldsymbol{\theta}^{(r)} = (\boldsymbol{\theta}_1^{(r)'}, \boldsymbol{\theta}_2^{(r)'}, \theta_3)'$
  - $||\cdot||_\infty$: Maximum norm

Variance estimation

- Define negative Hessian matrix as a block diagonal matrix, Q, with elements,

$$-\frac{\partial^2 Q_j(\boldsymbol{\theta}_j|\boldsymbol{\theta}_j^{(s)})}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_j'}|_{\theta_j=\theta_j^{(s)}}, j = 1, 2, 3$$

- SE of the respective estimate is given by the square root of corresponding diagonal element of inverse of Q
- Another estimator: Louis' method, BUT analytically cumbersome as pointed out by Lindsey and Ryan (1993)

Simulation setup: Design parameters

- Binary covariate
  - Low-dose animals: $z_i = \mathbf{0}, i = 1, \ldots, 130$;
  - High-dose animals: $z_i = \mathbf{1}, i = 131, \ldots, 225$
- Frailty: $r_i \sim N(0, 0.3666^2), i = 1, \ldots, 225$
- Baseline hazard rates
  - $\alpha_{01} = 0.0006$ and $\lambda_{01} = 0.0017$ on $I_1 = (0, 13]$;
  - $\alpha_{02} = 0.0018$ and $\lambda_{02} = 0.0069$ on $I_2 = (13, 19]$;
  - $\alpha_{03} = 0.0243$ and $\lambda_{03} = 0.0721$ on $I_3 = (19, 25]$;
  - $\alpha_{04} = 0.0001$ and $\lambda_{04} = 0.2734$ on $I_4 = (25, 33]$
- Set $\beta = 3.4185$, $\psi = 0.1689$, $\gamma = 0.6946$, and $\tau = 3.8467$
- Sacrifice times
  - $T_{2i} = \mathbf{13}, i = 1, \ldots, 16$; $\mathbf{19}$, $i = 17, \ldots, 105$; $\mathbf{33}$, $i = 106, \ldots, 130$;
  - $T_{2i} = \mathbf{13}, i = 131, \ldots, 145$; $\mathbf{19}$, $i = 146, \ldots, 207$; $\mathbf{33}$, $i = 208, \ldots, 225$

Simulation setup: Tumor onset and death times

- Step 1: Generate X_{oi}, from model (1) with $\alpha_0(t) = \alpha_{0j}$ on $I_j, j = 1, \ldots, 4$
- Step 2: Generate $\tilde{T}_{1i}$, from model (2) with $\lambda_0(t) = \lambda_{0j}$ on $I_j, j = 1, \ldots, 4$
- Step 3: If $\tilde{T}_{1i} < X_{oi}$,
  - Go to Step 4;
  - Otherwise, go to Step 5
- Step 4: If $\tilde{T}_{1i} \leq T_{2i}$,
  - $T_i = \tilde{T}_{1i}, \delta_i = 0, d_i = 1$ (**DNT**);
  - Otherwise, $T_i = T_{2i}, \delta_i = 0, d_i = 0$ (**SNT**)
- Step 5: Conditional on X_{oi}, generate T_{1i} from piecewise exponential distribution with $\lambda_0(t) = \lambda_{0j}$ on $I_j, j = 1, \ldots, 4$
  - If $T_{1i} \leq T_{2i}$, $T_i = T_{1i}, \delta_i = 1, d_i = 1$ (**DWT**);
  - Otherwise, $T_i = T_{2i}, \delta_i = 1, d_i = 0$ (**SWT**) or $T_i = T_{2i}, \delta_i = 0, d_i = 0$ (**SNT**) depending on whether or not $X_{oi} \leq T_{2i}$

Comparison

- Apply the simulated data to both Lindsey and Ryan model (Lindsey and Ryan, 1993) and the proposed model
- Summary statistics
  - Estimated mean
  - Bias
  - Empirical standard error (SE)
  - Mean of the 1,000 estimated standard errors (SEM)
  - 95% coverage rate (CR) of the regression parameters, β, ψ, and γ, commonly included in both the Lindsey and Ryan's model (LR) and the proposed model (Proposed).

Simulation results

Table: Estimated mean, bias, empirical standard error (SE), mean of the 1,000 estimated standard errors (SEM), and 95% coverage rate (CR) for the common parameters, β, ψ, and γ, in LR and proposed models

Parameter	True	Mean	Bias	SE	SEM	CR
			LR model			
β	3.4185	3.4746	0.0561	0.6781	0.5369	**89.8**
ψ	0.1689	-0.2360	-0.4049	0.6206	0.5036	**85.2**
γ	0.6946	0.0179	-0.6767	1.4579	0.8445	**67.5**
			Proposed model			
β	3.4185	3.5991	0.1806	0.5866	0.5329	**95.7**
ψ	0.1689	0.0937	-0.0751	0.5781	0.5207	**94.3**
γ	0.6946	0.7850	0.0904	0.6909	0.5424	**90.2**

Discussion

- Propose a three-state model with normal frailty to incorporate the dependency of tumor onset and death for the current status data
- Based on a simulation study, our model was better than LR model in terms of summary statistics and the coverage rate
- According to the results not reported in Table 2, the coverage rate of the lethality parameter in LR model was 53.1% assuming the true value of 1 and that of τ were 81.4%. This low coverage rate of τ may be caused by lack of distinct information of DNT or DWT animals
- In E-step, the Gauss-Hermite algorithm was used to approximate the functionals of the random frailty, but MCMC sampling technique can be applied

Thanks to:

Chung Mo Nam (Yonsei University College of Medicine),
Yang-Jin Kim (Sookmyung Women's University),
Youn Nam Kim (Yonsei University Graduate School of Public Health),
Eun Hee Choi (Yonsei University College of Medicine)