Regression analysis of clustered interval-censored data with informative cluster size

Yang-Jin Kim∗†

Interval-censored data are commonly found in studies of diseases that progress without symptoms, which require clinical evaluation for detection. Several techniques have been suggested with independent assumption. However, the assumption will not be valid if observations come from clusters. Furthermore, when the cluster size relates to response variables, commonly used methods can bring biased results. For example, in a study on lymphatic filariasis, a parasitic disease where worms make several nests in the infected person’s lymphatic vessels and reside until adulthood, the response variable of interest is the nest-extinction times. As the extinction times of nests are checked by repeated ultrasound examinations, exact extinction times are not observed. Instead, data are composed of two examination points: the last examination time with living worms and the first examination time with dead worms. Furthermore, as Williamson et al. (Statistics in Medicine 2008; 27:543–555) pointed out, larger nests show a tendency for low clearance rates. This association has been denoted as an informative cluster size.

To analyze the relationship between the numbers of nests and interval-censored nest-extinction times, this study proposes a joint model for the relationship between cluster size and clustered interval-censored failure data. A proportional hazard model with random effect and a mixed ordinal regression model are applied to failure times and cluster size, respectively. The joint model approach addresses both the association among failure times from the same cluster and the dependency of failure times on cluster size. Simulation studies are performed to assess the finite sample properties of the estimators and lymphatic filariasis data are analyzed as an illustration. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: clustered failure time data; informative cluster size; interval-censored data; multiple imputations; random effect

1. Introduction

For infectious diseases, the occurrence of infection is checked during the periodic clinic visits. In such cases, the exact infection time is not observed, instead, only two examination times that bracket the infection time are known. These types of data are denoted as interval-censored data and commonly observed in studies of diseases that progress without symptoms requiring clinical evaluations for detection. Several techniques have been proposed under an independent assumption among observations [1]. However, the assumption will not be valid if the interval-censored data come from same clusters. To analyze clustered interval-censored data, Bellamy et al. [2] used a normally distributed frailty effect when the interval-censored failure times follow a Weibull distribution, Wong [3] employed a Bayesian approach for multilevel clustered interval-censored data and analyzed dental data, and Goethals et al. [4] applied a shared gamma frailty effect to analyze cow udder quarter infection times. Furthermore, if the cluster size includes information about failure times, then this relationship should be considered in the analysis. For example, in a toxicity experiment, mouse fertility was found to be related to offspring size as well as their survival times when mother mouse underwent treatment. That is, smaller size offspring tend to have shorter survival times. As another example, in a study on the relationship between the failure times of teeth of patients with dental disease and their number of teeth, Spiekerman and Lin [5] found that patients with poorer dental conditions had smaller numbers of teeth and shorter failure times. Williamson et al. [6] also analyzed the relationship between number of nests and their extinction times in a parasitic disease and showed that smaller numbers of nests showed higher clearance rates. This association has been denoted as an informative cluster.
size. To reflect informative cluster size, several approaches were suggested in the absence of censoring. Catalano and Ryan [7] incorporated cluster size as a covariate in a model for the response variable. Hoffman et al. [8] also suggested a within cluster resampling (WCR) technique that assigns equal sampling probabilities for all subunits. Williamson et al. [9] applied this technique to the clustered data with a GEE approach and again extended to the clustered survival data in Williamson et al. [6]. Cong et al. [10] also considered a similar situation and compared the WCR and the weighted score function (WSF) approach using cluster size as weight. Another related approach is the application of a joint model. Dunson et al. [11] suggested a Bayesian approach, a joint model for cluster size and clustered outcome. The purpose of our study is to develop new methodology to analyze clustered interval-censored data with informative cluster size. There are two main characteristics to consider simultaneously. The first is the relationship between cluster size and clustered failure times and the second is the incompleteness of interval-censored failure times. In this regard, we propose a joint model to examine the relationship between clustered interval-censored failure data and cluster size. A proportional hazard model with a random effect and a mixed ordinal regression model are applied to failure times and cluster size, respectively. Then, a joint model approach addresses the association among failure times generated from the same cluster as well as the relationship between failure times and the cluster size. Numerous previous studies related to joint models have been proposed to examine the relationship between failure times and longitudinal measurements (e.g. Wulfsohn and Tsiatis [12]; Henderson et al. [13]; Huang and Liu [14]). Section 2 describes model specifications and inference procedures and Section 3 shows the simulation results to validate the properties of the estimates based on the proposed method. Next, parasitic disease data are analyzed as an illustrative example in Section 4. Finally, Section 5 discusses some further comments.

2. Statistical model

Consider a survival study that involves n independent clusters and each cluster includes several subunits. Denote $T_{ij}$ as the failure time of the jth subunit ($j = 1, \ldots, m_i$) of the ith cluster ($i = 1, \ldots, n$) which cannot be exactly observed, instead, two time points including $T_{ij}$ are available and denoted as $(L_{ij}, U_{ij})$. Define the censoring indicator as $\delta_{ij} = I(U_{ij} < \infty)$, for the $L_{ij} \leq T_{ij} \leq U_{ij}$. We assume that the observation process generating $(L_{ij}, U_{ij})$ is independent of failure time process. Let $X_i$ be a q-vector of cluster-specific covariates and $Z_{ij}$ a p-vector of subunit-specific covariates. In this study, our interest is to estimate covariate effect on the failure time when the cluster size is informative for failure times. There are three characteristics that require extension of techniques applied to analyze the ordinary survival data.

First, for the model of cluster size, we adopt an ordinal regression model regarding $m_1, \ldots, m_n$ as K category ordinal responses with $1 \leq m_i \leq K$. Ordinal variables can be explained as the discretized values of latent variable with certain continuous distribution. In mouse fertility data, unmeasurable latent fertility is assumed to be realized as offspring size. Furthermore, ordinal variables is a reasonable range, for example, the number of mice offspring is known to be less than 10.

Given a random effect $u_i$ and cluster-specific covariate vector $X_i = x_i$, the following mixed effect cumulative logit model is used for the cluster size $m_i$:

$$\log \frac{\Pr(m_i \leq k|x_i, u_i)}{1 - \Pr(m_i \leq k|x_i, u_i)} = \log \frac{P_{ik}}{1 - P_{ik}} = \alpha_k + x_i'\gamma + u_i$$

(1)

where $\gamma$ is the regression parameter vector and $\alpha = (\alpha_1, \ldots, \alpha_{K-1})$ are ordered, category-specific cutpoints or thresholds. The random effect $u_i$ reflects the heterogeneity of each cluster and is assumed to follow a normal distribution with mean zero and variance $\sigma_u^2$. Based on Model (1), clusters with larger random effect tend to have smaller cluster sizes. While the fertility can result in offspring size, it also affect their failure times, $T_{i1}, \ldots, T_{im_i}$.

Second, in terms of failure times of subunits, another random effect $v_i$ is incorporated to reflect the correlation among failure times of subunits within the same cluster. It is also assume to follow a normal distribution with mean zero and variance $\sigma_v^2$. Given covariate $z_{ij}$ and random effects $u_i$ and $v_i$, the hazard function of failure time is defined by the following proportional hazard frailty model:

$$\lambda_{ij}(t|z_{ij}, u_i, v_i) = \lambda_0(t)\exp(\beta'z_{ij} + \eta u_i + v_i)$$

(2)

where $\lambda_0$ is an unspecified baseline hazard function and $\beta$ is a $p \times 1$ vector of regression coefficient vector. In survival data analysis, frailty has been applied with the understanding that the relatively frail subjects will experience an event earlier than others. In addition, frailty effect has been applied as medium at the joint models in which longitudinal markers and failure times are connected. By extending this joint model approach to informative cluster size, $\eta$ will measure the degree of relationship between failure times and the cluster size. For example, $\eta > 0$ would indicate a positive relationship between a small cluster size and early failure times. Likewise, a negative $\eta$ would show that a larger cluster size results in earlier failure times of subunits.
The third characteristic is related to interval-censored data. For the derivation of suitable likelihood, the approximate likelihood approach suggested by Goetghebeur and Ryan [15] is applied. A piecewise exponential model is defined with hazard rate $\lambda_l$ for the $l(=1, \ldots, q)$ th time interval. Here, the time unit is assumed to choose fine enough that tied observation times are unlikely to occur. Then $Y_{ijl}$ is an a risk indicator showing the $j$ th subunit of the $i$ th cluster is at risk at the $l$ th interval. Let $N_{ijl}(t)$ count the events occurred and then $dN_{ijl}$ has a value one when the $j$ th subunit of the $i$ th cluster has an event at the $l$ th interval. Given two frailty effects $u_i$ and $v_i$, clustered interval-censored failure times have the following approximate likelihood:

$$
\prod_{i=1}^{n} \prod_{j=1}^{m_i} \prod_{l=1}^{q} \left( \lambda_l \exp(\beta' z_{ijl} + \eta u_i + v_i) \right) dN_{ijl} \exp[-Y_{ijl} \lambda_l \exp(\beta' z_{ijl} + \eta u_i + v_i)] = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \prod_{l=1}^{q} L_{ijl}(\beta, \eta, \lambda|u, v) \tag{3}
$$

### 3. Estimation

The EM algorithm provides a popular method to estimate parameters when the likelihood function includes incomplete data. In our study, the parameters of interest are denoted as $\theta = (\beta, \eta, \lambda_1, \ldots, \lambda_q, \alpha, \gamma, \sigma_u^2, \sigma_v^2) = (\theta_1, \theta_2, \theta_3)$, where $\theta_1 = (\beta, \eta, \lambda_1, \ldots, \lambda_q)$, $\theta_2 = (\alpha, \gamma)$ and $\theta_3 = (\sigma_u^2, \sigma_v^2)$. Observed data for cluster $i(i = 1, \ldots, n)$ are composed of $O_i = ((L_{ijl}, U_{ijl}, d_{ijl}, z_{ijl}) | y_{ijl}, x_{ijl})$ where $r_i = (r_{i1}, \ldots, r_{ik-1})$ with $r_{ij} = 1$ when $m_i = j$ and $r_{ij} = 0$ for otherwise. Missing data are related to both interval-censored data and random effects $M_i = (d_{ijl}, Y_{ijl}, u_i, v_i; l = 1, \ldots, q, j = 1, \ldots, m_i)$. Following the notations set by McCullagh [16], the complete data log-likelihood is written as

$$
\log L(\theta|\mathbf{b}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{l=1}^{q} \left[ dN_{ijl} \log \lambda_l + \beta' z_{ijl} + \eta u_i + v_i \right] - Y_{ijl} \lambda_l \exp(\beta' z_{ijl} + \eta u_i + v_i)
$$

$$
+ \sum_{i=1}^{n} \sum_{k=1}^{K} \left[ R_{ik} \phi_{ik}(\theta_2) - R_{ik+1} g(\phi_{ik}(\theta_2)) \right] \tag{4}
$$

where $R_{ik} = r_{i1} + \cdots + r_{ik}$, $\phi_{ik}(\theta_2) = \log(P_{ik}/(P_{ik+1} - P_{ik}))$ and $g(a) = \log(1+\exp(a))$. At the E-step, the conditional expectations of these incomplete data are calculated using a two-stage algorithm.

#### 3.1. E-step

At the E-step, we calculate the expectation of the complete data log-likelihood conditional on the observed data and the current parameters, $\hat{\theta}$. Denote the conditional expectations $E(\cdot|O, \hat{\theta})$ as $E^*(\cdot)$. Then

$$
Q(\theta) = E[I(\theta)|O, \hat{\theta}] = Q_1(\theta_1) + Q_2(\theta_2) + Q_3(\theta_3),
$$

where

$$
Q_1(\theta_1) = \sum_{ijl} E^*[dN_{ijl}] \left[ \log \lambda_l + \beta' z_{ijl} + \eta E^*[u_i] + E^*[v_i] \right] - E^*[Y_{ijl}] \lambda_l e^{\beta' z_{ijl}} E^*(e^{\eta u_i}) E^*(e^{v_i}),
$$

$$
Q_2(\theta_2) = \sum_{i=1}^{n} \sum_{k=1}^{K} \left[ R_{ik} E^*[\phi_{ik}(\theta_2)] - R_{ik+1} E^*(g[\phi_{ik}(\theta_2)]) \right],
$$

$$
Q_3(\theta_3) = -\frac{1}{2} \left( n \log \sigma_u^2 + \log \sigma_v^2 + \frac{1}{\sigma_u^2} \sum_{i=1}^{n} E^*[u_i^2] + \frac{1}{\sigma_v^2} \sum_{i=1}^{n} E^*[v_i^2] \right)
$$

Following, two-stage procedure is applied to calculate the expectations in $Q_1$, $Q_2$ and $Q_3$.

**Stage 1.** Assume that $dN_{ijl}$ and $Y_{ijl}$ are known. Then, the conditional expectations of the functions of frailties, $E^*(h(u_i))$ and $E^*(g(v_i))$ are calculated as follows:

$$
E^*(h(u_i)) = \frac{\int_{-\infty}^{\infty} h(u_i) L_c(t) d\Phi_u(u_i)}{\int_{-\infty}^{\infty} L_c(t) d\Phi_u(u_i)} , \quad E^*(g(v_i)) = \frac{\int_{-\infty}^{\infty} g(v_i) L_c(t) d\Phi_v(v_i)}{\int_{-\infty}^{\infty} L_c(t) d\Phi_v(v_i)}
$$

where $\Phi_u, \Phi_v$ are the normal distributions of mean zero and variances $\sigma_u^2$ and $\sigma_v^2$, respectively. Because these integrations have no closed form, a 15-point Gauss-Hermite integration is employed.
Stage 2. Using the frailty functions estimated in Stage 1, the conditional expectations of \( dN_{ijt} \) and \( Y_{ijt} \) are calculated as follows:

\[
E^*[dN_{ijt}] = E[dN_{ijt}|O_t, \theta] = \frac{p_{ij}}{\sum_{L_{ij} \leq t \leq R_{ij}} p_{ij}} \quad \text{if} \ t_1 \in [L_{ij}, U_{ij}],
\]

where

\[
p_{ij} = E^{*}\left( \lambda_{ij} e^{(\beta z_{ij} + \eta u_i + v_j) + \sum_{l=1}^{I_l} \lambda_{il}} \right)
\]

Also, with a censoring indicator, \( \delta_{ijt} \)

\[
E^*(Y_{ijt}) = E(Y_{ijt}|O_t, \theta) = \sum_{q \geq t} E^*(dN_{ijt})\delta_{ijt} + I(t_1 \leq L_{ij})(1 - \delta_{ijt})
\]
is calculated.

3.2. M-step

At the M-step, by taking derivatives of the log-likelihood function with respect to unknown parameters, the MLE of \( \lambda \) is defined as

\[
\hat{\lambda}_l = \frac{\sum_{ij} E^*[dN_{ijt}]}{\sum_{ij} E^*[Y_{ijt}]\exp(\beta z_{ij})E^*[e^{\eta u_i}]E^*[e^{v_j}]}, \quad l = 1, \ldots, q
\]

By applying (5) into the likelihood, the profile likelihood score equations for \( \beta \) and \( \eta \) are defined as

\[
\sum_{ij} \left( z_{ij} - \frac{\sum_{i', j'} \exp(\beta z_{ij'})E^*[e^{\eta u_i}]E^*[e^{v_j}]E^*[Y_{ij'}]}{\sum_{i', j'} \exp(\beta z_{ij'})E^*[e^{\eta u_i}]E^*[e^{v_j}]E^*[Y_{ij'}]} \right) E^*[dN_{ijt}],
\]

\[
\sum_{ij} \left( E^*(u_i) - \frac{\sum_{i', j'} \exp(\beta z_{ij'})E^*[e^{\eta u_i}]E^*[e^{v_j}]E^*[Y_{ij'}]}{\sum_{i', j'} \exp(\beta z_{ij'})E^*[e^{\eta u_i}]E^*[e^{v_j}]E^*[Y_{ij'}]} \right) E^*[dN_{ijt}].
\]

A Fisher’s scoring method is applied for \( \theta_2 = (x_1, \ldots, x_{K-1}, \gamma) \) and the estimates of \( \sigma_u^2 \) and \( \sigma_v^2 \) are

\[
\hat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^{n} E^*(u_i^2), \quad \hat{\sigma}_v^2 = \frac{1}{n} \sum_{i=1}^{n} E^*(v_i^2)
\]

The complete estimation procedure can be summarized as follows:

(i) For obtaining initial parameters, estimate \((\beta_0, \lambda_0, x_0, \gamma_0)\) using midpoints of interval with \( u_i = 0 \) and \( v_i = 0 \).
(ii) Calculate the conditional expectations for \( dN_{ijt}, Y_{ijt}, h(u_i) \) and \( g(v_i) \) at the E-step.
(iii) At the M-step, estimate \( \theta = (\beta, \eta, \lambda, x, \gamma) \) using the conditional expectations calculated at (ii).
(iv) Iterate (ii) and (iii) until the condition of convergence is satisfied.

To estimate the covariance of \( \beta, \eta, x \) and \( \gamma \), multiple imputation approach (Rubin, [17]) is applied. Let \( F_{ij}(t) \) denote the cumulative distribution function of failure times and

\[
\hat{F}_{ij}(t) = 1 - E^*\left\{ \exp\left( -\exp\left( \sum_{t_k \leq t} \hat{\lambda}(t_k)\hat{\beta} z_{ij} + \hat{\eta} u_i + v_j \right) \right) \right\}
\]

provides an estimate of \( F(t) \) corresponding to the \( i \)th subunit of the \( i \)th cluster. Let \( G \) be an integer. For each \( g = 1, \ldots, G \), \( T_{g,ij} \) is generated independently from \( \hat{F}_{ij} \) satisfying \( L_{ij} \leq T_{g,ij} \leq U_{ij} \). Given \( u_i \) and \( v_i \), the proportional hazards frailty model is fitted and the new imputed parameters, \( \hat{\beta}_{g}^{(I)} \) and variance of \( \hat{\beta}_{g}^{(I)} \), denoted by \( V_{g}^{(I)} \), can be obtained by iteration. The variance of \( \hat{\beta} \) can then be estimated by

\[
\frac{\sum_{g=1}^{G} V_{g}^{(I)}}{G} + \left( 1 + \frac{1}{G} \right) \frac{\sum_{g=1}^{G} (\hat{\beta}_{g}^{(I)} - \hat{\beta}) (\hat{\beta}_{g}^{(I)} - \hat{\beta})}{G}
\]

where \( \hat{\beta}^{(I)} \) is the average of imputed estimates, \( \hat{\beta}^{(I)}_1, \ldots, \hat{\beta}^{(I)}_G \). This study uses \( G = 20 \). For the variances of \( \eta, x \) and \( \gamma \), similar procedures are applied.
The proposed method is applied to the lymphatic filariasis data previously analyzed by Williamson et al. 

<table>
<thead>
<tr>
<th>Simulation results at two different associations.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(i)</em> Positive association, ( \eta &gt; 0 )</td>
</tr>
<tr>
<td>( \beta(0.3) )</td>
</tr>
<tr>
<td>( \eta(0.3) )</td>
</tr>
<tr>
<td>( z_1(0.6) )</td>
</tr>
<tr>
<td>( z_2(0.8) )</td>
</tr>
<tr>
<td>( \gamma(0.5) )</td>
</tr>
</tbody>
</table>

*(ii)* Negative association, \( \eta < 0 \)
| \( \beta(-0.3) \) | -0.300 | 0.180 | 0.176 | 0.952 | 0.248 | 0.085 | 0.088 | 0.876 |
| \( \eta(-0.3) \) | -0.300 | 0.263 | 0.250 | 0.955 | 0.294 | 0.211 | 0.220 | 0.920 |
| \( z_1(0.6) \) | 0.615 | 0.268 | 0.248 | 0.962 | 0.613 | 0.227 | 0.232 | 0.936 |
| \( z_2(0.8) \) | 0.824 | 0.273 | 0.266 | 0.948 | 0.838 | 0.234 | 0.246 | 0.926 |
| \( \gamma(0.5) \) | 0.514 | 0.271 | 0.286 | 0.948 | 0.503 | 0.092 | 0.094 | 0.946 |

4. Simulation

We conducted a simulation study to evaluate the performance of the proposed method. For the cluster \( i = 1, \ldots, 100 \), frailty effects, \( u_i \) and \( v_i \) are generated from \( N(0,0.5^2) \) and \( N(0,0.5^2) \), respectively. Continuous covariate is considered, \( x_i \sim N(0,1) \). With \( x_i \) and \( u_i \), cluster sizes are generated using the following parameter values: \( (x_1, x_2, x_3, \gamma) = (0.3, 0.6, 0.8, 0.5) \) and the resulting range of cluster size is \( 1 \leq m_i \leq 4 \). Once cluster size \( m_i \) is determined, interval-censored failure times \( \{L_{ij}, U_{ij}, \delta_{ij}\} \) are generated for cluster \( i \). For \( j \)th subunit in the \( i \)th cluster, survival time \( T_{ij} \) is generated from the exponential distribution with model (2) and a corresponding censoring variable \( C_{ij} \) from the exponential distribution with mean 10.0. Then, the censoring indicator \( \delta_{ij} \) is defined as \( \delta_{ij} = I(T_{ij} < C_{ij}). For \delta_{ij} = 1, \) interval-censored failure times including \( T_{ij} \) are generated following Pan’s [18] approach: Suppose that \( \text{len} \sim U(0.1, 0.2) \) and \( U \sim U(0, 1) \). By identifying \( U + k \cdot \text{len}, U + (k + 1) \cdot \text{len} \) as \( (L_{ij}, R_{ij}) \) when \( U + k \cdot \text{len} < T_{ij} \leq U + (k + 1) \cdot \text{len} \), \( k = 1, \ldots \). For \( \delta_{ij} = 0, L_{ij} = C_{ij} \) and \( R_{ij} = \infty \) are defined. Table I presents the results from 500 replications of 100 clusters at the following two cases; (i) positive and (ii) negative relationship between the cluster size and failure times. SE is the empirical standard error calculated from 500 parameter estimates, SEM is the mean of 500 estimated standard errors of the parameters and CP is the empirical coverage probability of the estimated 95 per cent confidence interval. The proposed procedure yields satisfactory results in both cases. The biases of estimates were negligible and the coverage probabilities were close to the nominal level at 0.95. In addition, the estimated standard errors showed similar values with the empirical standard errors. For comparison purposes, reduced models ignoring dependence, \( \eta = 0 \), were considered and provide biased estimates for \( \beta \). 

5. Data analysis

The proposed method is applied to the lymphatic filariasis data previously analyzed by Williamson et al. [6]. Infected mosquitos transmitted infective \( W. bancrofti \) larvae into the lymphatic vessels of people, which developed into adult worms that live and nest in the vessels. Ultrasound allows for the visualization of the adult worms and detection of their deaths. In this study, a patient represents the cluster, and subunits of cluster are adult worms. The main purpose of this study is to compare the effectiveness of two treatments, DEC/ALB and DEC alone, on the eradication of adult worm nests. Forty-seven patients participated in the study, 25 underwent DEC treatment and the remaining 22 underwent DEC/ALB treatment. Ultrasound examinations were done at 7, 14, 30, 45, 60, 90, 180, 270 and 360 days after treatment. In each examination, adult worms’ deaths were detected. A total of 78 adult worm nests were found in patients; there was a range of 1–5 nests per patient. Just as Williamson et al. [6] found, as cluster size (i.e. the number of nests) increased, the proportion of nest clearance decreased. For example, out of the 22 patients hosing one nest, there was an 82 per cent clearance while patients with 4–5 nests only had a 33 per cent clearance proportion. That is, worms living in larger communities of nests seem to have higher survival rates. Therefore, the number of nests is related to survival time and this situation can be denoted as an informative cluster size. The response variable of interest is the extinction time of nests, which cannot be exactly observed. Instead, the available data are the two
Table II. The comparisons of two models for lymphatic filariasis data.

<table>
<thead>
<tr>
<th></th>
<th>Joint</th>
<th>Reduced model</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH model</td>
<td>$\hat{\beta}$</td>
<td>$\text{se}(\hat{\beta})$</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.887</td>
<td>0.499</td>
</tr>
<tr>
<td>Age</td>
<td>0.069</td>
<td>0.193</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.853</td>
<td>0.393</td>
</tr>
<tr>
<td>Ordinal regression</td>
<td>$\hat{b}$</td>
<td>$\text{se}(\hat{b})$</td>
</tr>
<tr>
<td>$z_1$</td>
<td>-0.005</td>
<td>0.303</td>
</tr>
<tr>
<td>$z_2$</td>
<td>2.863</td>
<td>0.561</td>
</tr>
<tr>
<td>$z_3$</td>
<td>3.649</td>
<td>0.769</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.942</td>
<td>0.646</td>
</tr>
<tr>
<td>Age</td>
<td>0.073</td>
<td>0.334</td>
</tr>
<tr>
<td>$(\sigma_1, \sigma_2)$</td>
<td>(0.882, 0.432)</td>
<td>(0.600, 0.432)</td>
</tr>
</tbody>
</table>

Figure 1. Survival curves by the number of nests.

examination times: the last examination time with living worms and the first examination time with dead worms. The two covariates are the treatment groups (DEC/ALB =0, DEC =1) and the patients’ ages in years, centering on the mean age, 21.5 and dividing by 10. Table II shows the results of the joint and reduced models. In both cases, treatment has significant effect. Thus, nests with treatment DEC = 1 have faster failures. However, the influence of age is insignificant on the failure times. For the relationship between cluster size and failure time, $\eta = 0.853$ indicates that clusters with small nests have faster failure times, as expected. Recently, Zhang and Sun [19] extended Williamson et al.’s [6] approach to interval-censored data and their results were close to Williamson’s ones. The difference between results based on their approach and those given here is due to the method of modeling and the assumption made. In their approaches, the cluster size is regarded as weights and a weighted estimating procedures are used to estimate parameters. However, in our case, random effect plays medium role to connect failure time and cluster size. Figure 1 shows the estimated survival curves according to cluster size using the self-consistency algorithm suggested by Turnbull [20]. These survival curves indicate that four groups (number of nests $\geq 1, 2, 3$ and $\geq 4$) have very different survival curves.
6. Discussion

This paper proposes a joint model for clustered interval-censored failure times when cluster size is informative for failure times. Random effects were used to investigate the relationship between cluster size and failure times and an approximate likelihood was applied to derive a suitable likelihood for interval-censored data. While most approaches for the analysis of informative cluster size derive weights to adjust for such informativeness, this study used an approach similar to that of Catalano and Ryan [7] in which cluster size is included as a covariate in the response model. However, these are strong assumptions. Instead, subject-specific random effect is incorporated. There will be several related works for modeling informativeness at clustered interval-censored data. In particular, work should center on the extension of weight approaches on the interval-censored data. By deriving an appropriate estimating equation and defining weights, several approaches previously proposed for right-censored data can be applied to interval-censored data. Another possible area will be consideration of the informative censoring process in the context of informative cluster size (Liu et al. [21]; Finkelstein et al. [22]). Related works could be considered with more complex models for failure times process and censoring process and presented in another paper.

Acknowledgements

The author thank Editor and referees for their helpful comments that greatly improved this paper. This research was supported by the Sookmyung Women’s University Research Grants 2010.

References

学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，
提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。
图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具