

Mixture cure model with random effects for the analysis of a multi-center tonsil cancer study

Yingwei Peng^{a,b,c*†} and Jeremy M. G. Taylor^d

Cure models for clustered survival data have the potential for broad applicability. In this paper, we consider the mixture cure model with random effects and propose several estimation methods based on Gaussian quadrature, rejection sampling, and importance sampling to obtain the maximum likelihood estimates of the model for clustered survival data with a cure fraction. The methods are flexible to accommodate various correlation structures. A simulation study demonstrates that the maximum likelihood estimates of parameters in the model tend to have smaller biases and variances than the estimates obtained from the existing methods. We apply the model to a study of tonsil cancer patients clustered by treatment centers to investigate the effect of covariates on the cure rate and on the failure time distribution of the uncured patients. The maximum likelihood estimates of the parameters demonstrate strong correlation among the failure times of the uncured patients and weak correlation among cure statuses in the same center. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: EM algorithm; Gaussian quadrature; importance sampling; maximum likelihood estimation; rejection sampling

1. Introduction

Modeling survival data with a surviving or cure fraction is an important issue since there exist many examples in cancer clinical studies, such as breast cancer, lymphoma, and head and neck cancer, in which a significant proportion of patients are cured. The issue becomes even more complicated when patients in a study form clusters and the independence of the patients within one cluster cannot be assumed or justified.

A motivating example of the study in this paper is a multi-institutional study of local recurrences of tonsil cancer patients [1]. This retrospective study consists of 672 patients with squamous cell carcinoma of the tonsil who were treated with radiation between 1976 and 1985 in nine centers in North America and in the United Kingdom. The failure time of interest is the time from initial treatment to local recurrence. Covariates in this study include age, sex, T stages, total dose, treatment duration, and node status. The follow-up period of this study is about 14 years, and most of the local recurrences occurred before 3 years. The objective of this study is to investigate the effects of some modifiable factors of the radiotherapy, including dose per fraction, overall treatment duration, and total dose, on the outcome of the radiation treatment of tonsil cancer patients. It is generally accepted that tonsil cancer is curable. The existence of cured patients in this particular study can be seen in Kaplan–Meier survival curves of the failure times based on node status, which is given in Figure 1. The curves level off beyond 3 years, indicating that patients not experiencing any recurrences in 3 years may likely be cured. Thus, it will be interesting to examine the effects of the covariates on both the time to recurrence of uncured patients and the probability of being cured.

This issue, however, is entangled by the fact that the time to recurrence or the probability of being cured for the patients from the same institution may be correlated due to a shared environment. The nine participating institutions are located in very different regions of North America and the United Kingdom. The shared (but unobservable or

^aDepartment of Community Health and Epidemiology, Queen's University, Kingston, ON, Canada K7L 3N6

^bDepartment of Mathematics and Statistics, Queen's University, Kingston, ON, Canada K7L 3N6

^cCancer Care and Epidemiology, Queen's Cancer Research Institute, Kingston, ON, Canada K7L 3N6

^dDepartment of Biostatistics, University of Michigan, 1420 Washington, Heights, Ann Arbor, MI 48109-2029, U.S.A.

*Correspondence to: Yingwei Peng, Department of Community Health and Epidemiology, Queen's University, Kingston, ON, Canada K7L 3N6.

†E-mail: pengp@queensu.ca

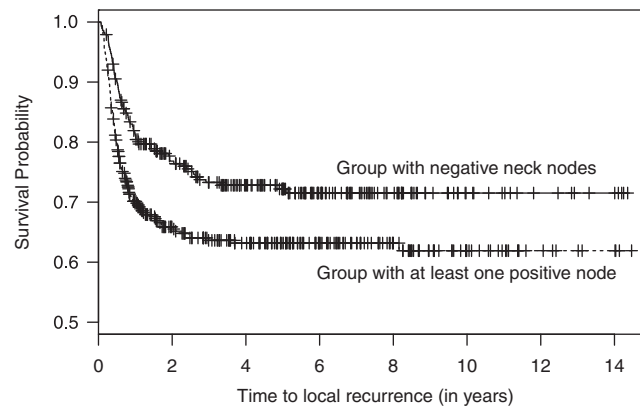


Figure 1. Kaplan–Meier survival curves of patients in the two node status groups.

not quantifiable) environment, treatment facilities, and other factors in one institution can be very different from other institutions. The absence of information about the factors will lead to positively correlated time to recurrence or cure status among patients in the same institution. When the correlation is strong, ignoring the correlation in an analysis could lead to biased estimates of the covariate effects. Our goal is to understand the effects of total dose and treatment duration as well as other covariates on both the time to recurrence for uncured tonsil cancer patients and on the probability of being cured when the potential correlation among patients in the same institution is properly handled.

There is a great deal of recent interest in cure models for clustered or correlated data in the literature. Peng *et al.* [2] proposed a semiparametric marginal cure model for clustered data. The model does not specify any correlation structure for the data and thus is robust to misspecification of the correlation structure. However, it may not be useful if the correlation is of interest. To model the correlation explicitly, random effects/frailty models are often considered. For example, Chen *et al.* [3] extended the frailty model [4] for bivariate survival data with a cure fraction. The frailty term is assumed to follow a stable distribution and the baseline distribution is assumed to be an improper parametric distribution to accommodate cured patients. Yin [5] extended this approach to a general cluster setting with a semiparametric baseline assumption. The authors considered Bayesian approaches to estimate parameters in the models. Chatterjee and Shih [6] and Wienke *et al.* [7] extended the mixture cure model to bivariate survival data with a cure fraction by modeling the distribution of the uncured patients with copulas, which are also frailty models. They proposed methods to obtain the maximum likelihood estimates of the parameters. However, they do not consider covariate effects in their models and the estimation methods become infeasible when cluster size is large. Yau and Ng [8] considered an extension of the popular Weibull mixture cure model by adding random effects, respectively, to both parts of the model for the cure proportion and for the failure time distribution of uncured patients. The authors assumed normal distributions for the random effects and adapted the best linear unbiased predictor (BLUP) method and the residual maximum likelihood estimator (REML) to estimate the parameters in the fully parametric cure model. Lai and Yau [9] extended this method to correlated random effects and non-parametric baseline specification.

In this paper, we consider a generalization of the univariate semiparametric mixture cure model [10, 11] to clustered failure time data with a surviving fraction. Two random effects are introduced to model the correlation between the cure status and between failure times, respectively, in the same cluster. The model is quite general and includes the models considered in Yau and Ng [8] and Lai and Yau [9] as special cases. We will consider maximum likelihood estimates of the parameters in the model, and propose several estimation methods to obtain the estimates. Their performances will be compared with the existing methods.

The paper is organized as follows. In Section 2, we outline the proportional hazards mixture cure model with random effects. Two estimation methods, one based on the Gaussian quadrature method and the other based on the Monte Carlo EM algorithm, are proposed in Section 3 to estimate the parameters in the model. Section 4 presents simulation studies to investigate the small sample properties of the proposed estimation methods. We apply the model to the tonsil cancer data in Section 5. Finally, we provide conclusions and some remarks on the proposed model and estimation methods in Section 6.

2. The model

Let $t_{ij} = \min(t_{ij}^*, c_{ij})$, $\delta_{ij} = 1(t_{ij}^* \leq c_{ij})$, $\mathbf{x}_{ij}, \mathbf{z}_{ij}$, $i = 1, \dots, n$, $j = 1, \dots, n_i$, be the observed failure time, censoring indicator, and the covariates of the j th individual in the i th cluster that may affect the failure time distribution of uncured individuals

and the cure proportion, where $1(A)$ is the indicator function of the event in the brackets and t_{ij}^* is the value of the underlying failure time T_{ij} that may be subject to censoring with the censoring time c_{ij} . Note that z_{ij} and x_{ij} may include the same covariates. We assume that given covariates, the censoring time c_{ij} is independent of the failure time t_{ij}^* and the cure status y_{ij} .

Let y_{ij} be the cure status of individual j in cluster i , that is, $y_{ij}=0$ if this individual is cured and 1 otherwise. It is obvious that y_{ij} is a partially observed latent variable. If $\delta_{ij}=1$, the individual experienced an event and $y_{ij}=1$. However, if $\delta_{ij}=0$, the individual's failure time is censored and the cure status of this individual is unknown. Therefore, y_{ij} may be 1 or 0.

We assume that the failure times t_{ij}^* 's and y_{ij} 's from individuals in the same cluster tend to be correlated. The correlation may be induced by, for example, shared environments for the members in the same cluster. For cluster i , we further assume that the correlations between the t_{ij}^* 's and between y_{ij} 's can be modeled by random effects $\mathbf{u}'=(u_1, u_2, \dots, u_l)$ and $\mathbf{v}'=(v_1, v_2, \dots, v_l)$, respectively. Given \mathbf{u} and \mathbf{v} , t_{ij}^* 's and y_{ij} 's are assumed to be independent and they can be modeled by the mixture cure model

$$P(T_{ij}>t_{ij}^*|\mathbf{u}, \mathbf{v})=S(t_{ij}^*|\mathbf{u}, \mathbf{v})=S_u(t_{ij}^*|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v})+1-\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v}) \quad (1)$$

where

$$S_u(t_{ij}^*|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})=P(T_{ij}>t_{ij}^*|\mathbf{u}, y_{ij}=1)=S_{u0}(t_{ij}^*)^{\exp(\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})} \quad (2)$$

$$\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v})=P(y_{ij}=1|\mathbf{v})=\frac{e^{z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v}}}{1+e^{z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v}}}, \quad (3)$$

$S_{u0}(\cdot)$ is an arbitrary baseline survival function for uncured individuals, $\boldsymbol{\beta}$ and \mathbf{x}_{ij}^* are the fixed effects and coefficients for the random effect \mathbf{u} in the proportional hazards model for $S_u(t)$, $\boldsymbol{\gamma}$ and \mathbf{z}_{ij}^* are the fixed effects and coefficients for the random effect \mathbf{v} in the logistic model for π . We usually require that $E(\mathbf{u})=0$ and $E(\mathbf{v})=0$. Let the density function of the joint distribution of \mathbf{u} and \mathbf{v} be $\phi[(\mathbf{u}', \mathbf{v}')', \mathbf{D}]$, where $\mathbf{D}=\mathbf{D}(\boldsymbol{\sigma})$ with $\boldsymbol{\sigma}$ denoting a vector of unknown parameters. The marginal likelihood function is

$$L(\boldsymbol{\theta})=\int \prod_{i=1}^n \prod_{j=1}^{n_i} [\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v})f_u(t_{ij}|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})]^{\delta_{ij}} \times [1-\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v})+\pi(z_{ij}\boldsymbol{\gamma}+\mathbf{z}_{ij}^*\mathbf{v})S_u(t_{ij}|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})]^{1-\delta_{ij}} \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}]d(\mathbf{u}, \mathbf{v}) \quad (4)$$

where $f_u(t_i|\mathbf{x}_i\boldsymbol{\beta}+\mathbf{u}_i)$ is the corresponding density function of (2), and $\boldsymbol{\theta}=(\boldsymbol{\beta}, \boldsymbol{\gamma}, S_{u0}, \boldsymbol{\sigma})$. The integral in (4) is usually intractable, which makes a direct maximization of the likelihood function difficult.

A simple example of the above general mixture cure model with random effects is the model with simple shared random effects where $l=n$, and \mathbf{x}_{ij}^* and \mathbf{z}_{ij}^* are vectors with 1 at the i th element and 0 at the other elements so that $\mathbf{x}_{ij}^*\mathbf{u}=\mathbf{u}_i$, $\mathbf{z}_{ij}^*\mathbf{v}=\mathbf{v}_i$. It represents a shared environment situation. Yau and Ng [8] and Lai and Yau [9] considered this special case of (1), (2), and (3) and assumed that (u_i, v_i) follows a bivariate normal distribution with zero means. The estimation method for parameters in their model is based on a combination of BLUP and REML methods. In this paper, we consider maximum likelihood estimates of the parameters in the general model (1), (2), and (3) via maximizing the likelihood function (4).

3. Estimation methods

3.1. Gaussian quadrature method

To maximize $\log L(\boldsymbol{\theta})$, one can employ numerical integration methods to approximate the integrals in (4). The adaptive Gaussian quadrature method as described in Pinheiro and Bates [12] is often used to approximate integrals in random effects models and it can be used to approximate the integrals in (4). This approach is conceptually simple. It allows non-normal random effects. The proportional hazard assumption (2) is not required. Other models, such as the accelerated failure time model $S_u(t_{ij}^*|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})=S_{u0}(t_{ij}^*)e^{\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u}}$ and the proportional odds model $S_u(t_{ij}^*|\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u})=S_{u0}(t_{ij}^*)e^{\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u}}/[1+S_{u0}(t_{ij}^*)+S_{u0}(t_{ij}^*)e^{\mathbf{x}_{ij}\boldsymbol{\beta}+\mathbf{x}_{ij}^*\mathbf{u}}]$ are possible. However, the baseline survival function $S_{u0}(t_{ij}^*)$ in these models has to be fully specified. In this paper, we will consider the Weibull and the piecewise constant hazard distributions, and the method discussed can be easily extended to other distributions. Under the Weibull baseline assumption,

$S_{u0}(t) = \exp[-(t/\lambda)^\kappa]$. It reduces to the exponential baseline distribution when $\kappa = 1$. The piecewise constant hazard distribution is defined as follows: Let $\tau_0 = 0 < \tau_1 < \tau_2 < \dots < \tau_K = \infty$ and $h_{u0}(t) = \exp(\alpha_j)$ if $\tau_{j-1} \leq t < \tau_j$, where $j = 1, \dots, K$ and $K \geq 1$. Under this definition, the baseline survival function is

$$S_{u0}(t) = \exp \left\{ \sum_{j: \tau_j < t} \exp(\alpha_{j+1}) [\min(t, \tau_{j+1}) - \tau_j] \right\}. \quad (5)$$

When $K = 1$, this distribution also reduces to the exponential distribution. We usually require that $\tau_1 > \min(t_{ij})$ and $\tau_{K-1} < \max(t_{ij})$. By choosing a large K , the piecewise baseline distribution can increase the flexibility of the parametric approach. The values $\tau_1, \dots, \tau_{K-1}$ in this method are often chosen to make equal-spaced intervals. Since the uncensored times contribute essential information to the hazard estimation, we suggest that τ_j be the $100j/K$ th percentile of the uncensored times so that each interval has events to estimate the hazard accurately.

The details of using the Gaussian quadrature approach to approximate the integrals can be found in the literature and thus are omitted. It is also used to approximate the negative second derivative of $\log L(\theta)$ with respect to θ to estimate the standard errors of the estimated parameters.

3.2. EM algorithm

The EM algorithm can be used to estimate the maximum likelihood estimate of θ in (4). The random effects \mathbf{u} and \mathbf{v} are usually treated as missing variables in the EM algorithm. Given the values of \mathbf{u} and \mathbf{v} , the complete data log-likelihood function for model (4) is

$$\begin{aligned} \tilde{\ell}(\theta | \mathbf{t}, \mathbf{u}, \mathbf{v}) = & \log \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}(\sigma)] + \sum_{i=1}^n \sum_{j=1}^{n_i} \{ \delta_{ij} \log \pi(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v}) + \delta_{ij} \log f_u(t_{ij} | \mathbf{x}_{ij} \boldsymbol{\beta} + \mathbf{x}_{ij}^* \mathbf{u}) \\ & + (1 - \delta_{ij}) \log [1 - \pi(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v})] + \pi(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v}) S_u(t_{ij} | \mathbf{x}_{ij} \boldsymbol{\beta} + \mathbf{x}_{ij}^* \mathbf{u}) \} \end{aligned} \quad (6)$$

The E-step of the EM algorithm computes the posterior expectation of $\tilde{\ell}(\theta | \mathbf{t}, \mathbf{u}, \mathbf{v})$ with respect to \mathbf{u} and \mathbf{v} , and the M-step maximizes the posterior expectation with respect to $\boldsymbol{\beta}$, γ , and h_{u0} to update the estimates. However, we will also consider \mathbf{y} as the latent variable in the EM algorithm. Given the values of \mathbf{y} , \mathbf{u} , and \mathbf{v} , the complete log-likelihood function of the mixture cure model is

$$\begin{aligned} \ell(\theta | \mathbf{t}, \mathbf{y}, \mathbf{u}, \mathbf{v}) = & \sum_{i=1}^n \sum_{j=1}^{n_i} \{ \delta_{ij} \log h_u(t_{ij} | \mathbf{x}_{ij} \boldsymbol{\beta} + \mathbf{x}_{ij}^* \mathbf{u}) + y_{ij} \log S_u(t_{ij} | \mathbf{x}_{ij} \boldsymbol{\beta} + \mathbf{x}_{ij}^* \mathbf{u}) \} \\ & + \sum_{i=1}^n \sum_{j=1}^{n_i} \{ y_{ij} \log \pi(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v}) + (1 - y_{ij}) \log [1 - \pi(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v})] \} + \sum_{i=1}^n \log \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}(\sigma)] \end{aligned} \quad (7)$$

The E-step of the EM algorithm computes the posterior expectation of $\ell(\theta | \mathbf{t}, \mathbf{y}, \mathbf{u}, \mathbf{v})$ with respect to \mathbf{y} , \mathbf{u} , and \mathbf{v} given the current estimates of the model parameters. Let the current estimate of θ be $\theta^{(r)}$, and the posterior density of $(\mathbf{y}, \mathbf{u}, \mathbf{v})$ given $\theta^{(r)}$ be $p(\mathbf{y}, \mathbf{u}, \mathbf{v} | \theta^{(r)})$. The E-step in the r th iteration calculates

$$\begin{aligned} Q(\theta | \theta^{(r)}) = & \sum_{i=1}^n \sum_{j=1}^{n_i} \{ \delta_{ij} [\log h_{u0}(t_{ij}) + \mathbf{x}_{ij} \boldsymbol{\beta} + E(\mathbf{x}_{ij}^* \mathbf{u} | \theta^{(r)})] - H_{u0}(t_{ij}) e^{\mathbf{x}_{ij} \boldsymbol{\beta} + \log E[y_{ij} e^{\mathbf{x}_{ij}^* \mathbf{u}} | \theta^{(r)}]} \} \\ & + \sum_{i=1}^n \sum_{j=1}^{n_i} \{ E(y_{ij} | \theta^{(r)}) \mathbf{z}_{ij} \gamma + E[y_{ij} \mathbf{z}_{ij}^* \mathbf{v} | \theta^{(r)}] - E(\log [1 + e^{\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v}}] | \theta^{(r)}) \} + \log E \{ \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}(\sigma)] | \theta^{(r)} \} \\ = & Q_1(\boldsymbol{\beta}, h_{u0} | \theta^{(r)}) + Q_2(\gamma | \theta^{(r)}) + Q_3(\sigma | \theta^{(r)}) \end{aligned} \quad (8)$$

where the expectation is taken with respect to $p(\mathbf{y}, \mathbf{u}, \mathbf{v} | \theta^{(r)})$.

3.2.1. Maximization step. The M-step updates θ via maximizing (8). Maximizing γ in $Q_2(\gamma | \theta^{(r)})$ has to be solved numerically. So does updating σ from $Q_3(\sigma | \theta^{(r)})$. Let $\omega_{ij}^{(r)} = \log E[y_{ij} e^{\mathbf{x}_{ij}^* \mathbf{u}} | \theta^{(r)}]$. It is easy to see that $Q_1(\boldsymbol{\beta}, h_{u0} | \theta^{(r)})$ is the log-likelihood function of the proportional hazards model, if the constant $E(u_i | \theta^{(r)})$ is replaced with the constant $\omega_{ij}^{(r)}$. A simple approach to update $\boldsymbol{\beta}$ and h_{u0} is to assume that $h_{u0}(t)$ is the hazard function of a parametric distribution,

and then β and $h_0(t)$ can be updated easily by the Newton–Raphson method. For instance, if $h_{u0}(t) = \lambda$ is the hazard function of the exponential distribution, then

$$Q_1(\beta, h_0(t)|\theta^{(r)}) = \log \lambda \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} + \sum_{i=1}^n \sum_{j=1}^{n_i} \{\delta_{ij} [\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}]\} - \lambda \sum_{i=1}^n \sum_{j=1}^{n_i} \{t_{ij} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}}\}$$

It is easy to see that $\hat{\lambda}^{(r)} = \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} / \sum_{i=1}^n \sum_{j=1}^{n_i} t_{ij} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}}$. Plugging in this estimate into the likelihood function, we obtain

$$Q_1(\beta|\theta^{(r)}) = \left\{ \log \left[\sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} \right] - \log \left[\sum_{i=1}^n \sum_{j=1}^{n_i} t_{ij} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} \right] \right\} \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} + \sum_{i=1}^n \sum_{j=1}^{n_i} \{\delta_{ij} [\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}]\} - \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij}$$

which can be maximized via the Newton–Raphson method. If a piecewise constant baseline distribution (5) is considered, then

$$Q_1(\beta, h_{u0}(t)|\theta^{(r)}) = \sum_{l=1}^K \alpha_l \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij} + \sum_{l=1}^K \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \{\delta_{ij} [\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}]\} - \sum_{l=1}^K e^{\alpha_l} \left\{ I(l < K) (\tau_l - \tau_{l-1}) \sum_{\tau_l \leq t_{ij}} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} + \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} (t_{ij} - \tau_{l-1}) e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} \right\}$$

It is easy to see that

$$e^{\alpha_l^{(r+1)}} = \frac{\sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij}}{I(l < K) (\tau_l - \tau_{l-1}) \sum_{\tau_l \leq t_{ij}} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} + \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} (t_{ij} - \tau_{l-1}) e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}}}$$

Plugging this in the Q_1 function, we obtain

$$Q_1(\beta|\theta^{(r)}) = \sum_{l=1}^K \left(\log \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij} \right) \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij} - \sum_{l=1}^K \left\{ \log \left[I(l < K) (\tau_l - \tau_{l-1}) \sum_{\tau_l \leq t_{ij}} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} + \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} (t_{ij} - \tau_{l-1}) e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} \right] \right\} \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij} + \sum_{l=1}^K \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \{\delta_{ij} [\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}]\} - \sum_{l=1}^K \sum_{\tau_{l-1} \leq t_{ij} < \tau_l} \delta_{ij}$$

A slightly different way of using the piecewise constant baseline hazard can lead to Cox’s partial likelihood. Let $\tau_1, \tau_2, \dots, \tau_K$ be the distinct uncensored times. Johansen [13] defined $h_{u0}(t) = \lambda_j$ for t in a small interval consisting of τ_j , and 0 otherwise. Let D_j be the set of d_j subjects with uncensored times equal to τ_j , and R_j be the set of subjects who are at risk prior to τ_j . The Q_1 likelihood function can be written as

$$Q_1(\beta, h_{u0}(t)|\theta^{(r)}) = \sum_{j=1}^K \left[d_j \log \lambda_j + \sum_{(i,j) \in D_j} \{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}\} - \lambda_j \sum_{(i,j) \in R_j} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}} \right]$$

It is easy to see that $\lambda_j^{(r+1)} = d_j / \sum_{(i,j) \in R_j} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}}$. Plugging this in the Q_1 function, we obtain $Q_1(\beta|\theta^{(r)}) \propto \sum_{j=1}^K [\sum_{(i,j) \in D_j} \{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}\} - d_j \log \sum_{(i,j) \in R_j} e^{\mathbf{x}_{ij} \beta + \omega_{ij}^{(r)}}]$. This is equivalent to using Cox’s partial likelihood method and the Nelson–Aalen baseline survival estimator to update β and $h_{u0}(t)$. Ripatti *et al.* [14] considered this approach for the frailty model.

3.2.2. *Expectation step.* Evaluating function (8) in E-step requires evaluating the following expectations ($E(y_{ij} v_i | \theta^{(r)})$ and $E(u_i | \theta^{(r)})$ are not involved in M-step, thus are not considered):

$$E(y_{ij} | \theta^{(r)}), \quad E[y_{ij} e^{\mathbf{x}_{ij}^* \mathbf{u}} | \theta^{(r)}] \tag{9}$$

$$E(\log[1 + \exp(\mathbf{z}_{ij} \gamma + \mathbf{z}_{ij}^* \mathbf{v})] | \theta^{(r)}), \quad E(\phi[\mathbf{u}, \mathbf{v}, \mathbf{D}(\sigma)] | \theta^{(r)}) \tag{10}$$

with respect to $p(\mathbf{y}, \mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)})$:

$$p(\mathbf{y}, \mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)}) \propto \prod_{i=1}^n \left\{ \prod_{j=1}^{n_i} e^{\delta_{ij} \mathbf{x}_{ij}^* \mathbf{u} - y_{ij} H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})} \prod_{j=1}^{n_i} \frac{e^{y_{ij} (\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v})}}{1 + e^{\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}}} \right\} \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}(\boldsymbol{\sigma}^{(r)})]$$

Given \mathbf{u} and \mathbf{v} , it is obvious that y_{ij} are independent, and $y_{ij} \equiv 1$ if $\delta_{ij} = 1$ and y_{ij} follows the following binomial distribution if $\delta_{ij} = 0$:

$$y_{ij} | \mathbf{u}, \mathbf{v} \sim \text{binomial} \left(1, p = \frac{\pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}}{1 - \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) + \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}} \right) \quad (11)$$

Thus, equations (9) can be further simplified into:

$$E(y_{ij} | \boldsymbol{\theta}^{(r)}) = \begin{cases} E \left[\frac{\pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}}{1 - \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) + \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}} \middle| \boldsymbol{\theta}^{(r)} \right] & \text{if } \delta_{ij} = 0 \\ 1 & \text{otherwise} \end{cases} \quad (9')$$

$$E[y_{ij} e^{\mathbf{x}_{ij}^* \mathbf{u}} | \boldsymbol{\theta}^{(r)}] = \begin{cases} E \left[\frac{e^{\mathbf{x}_{ij}^* \mathbf{u}} \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}}{1 - \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) + \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) e^{-H_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})}} \middle| \boldsymbol{\theta}^{(r)} \right] & \text{if } \delta_{ij} = 0 \\ E(e^{\mathbf{x}_{ij}^* \mathbf{u}} | \boldsymbol{\theta}^{(r)}) & \text{otherwise} \end{cases}$$

where the expectations are taken with respect to the distribution of (\mathbf{u}, \mathbf{v}) . The E-step requires to update the expectations above and those in (10) with respect to (\mathbf{u}, \mathbf{v}) . Following (6), the density function of the distribution is $\tilde{q}(\mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)}) \propto q(\mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)})$ and

$$q(\mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)}) = \phi[(\mathbf{u}, \mathbf{v}), \mathbf{D}(\boldsymbol{\sigma}^{(r)})] \prod_{i=1}^n \left\{ \prod_{j=1}^{n_i} [e^{\mathbf{x}_{ij}^* \mathbf{u}} \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) S_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})]^{\delta_{ij}} \times \prod_{j=1}^{n_i} [1 - \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) + \pi(\mathbf{z}_{ij} \boldsymbol{\gamma}^{(r)} + \mathbf{z}_{ij}^* \mathbf{v}) S_{u0}^{(r)}(t_{ij}) \exp(\mathbf{x}_{ij} \boldsymbol{\beta}^{(r)} + \mathbf{x}_{ij}^* \mathbf{u})]^{1 - \delta_{ij}} \right\} \quad (12)$$

Unfortunately, the expectations do not have closed form because of the complicated form of $\tilde{q}(\mathbf{u}, \mathbf{v}|\boldsymbol{\theta}^{(r)})$. We will rely on Monte Carlo methods to generate random numbers from the distribution and evaluate the expectations. The details of the Monte Carlo methods are provided in the Appendix.

3.2.3. Variance estimation in EM algorithm. The variances of the estimated parameters in the EM algorithm cannot be obtained directly from the algorithm. There are several approaches suggested in the literature to estimate the variances based on the EM algorithm [15–17], and they closely relate to Louis' formula $-\partial^2 \log L(\boldsymbol{\theta}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T) = -E(\partial^2 \ell(\boldsymbol{\theta} | \mathbf{t}, \mathbf{y}, \mathbf{u}, \mathbf{v}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T)) - \text{Var}(\partial \ell(\boldsymbol{\theta} | \mathbf{t}, \mathbf{y}, \mathbf{u}, \mathbf{v}) / \partial \boldsymbol{\theta})$, where the expectation and the variance are taken with respect to $(\mathbf{y}, \mathbf{u}, \mathbf{v})$ [18]. It is easy to see from (7) that both $\partial^2 \ell(\boldsymbol{\theta} | \mathbf{y}, \mathbf{u}, \mathbf{v}, \mathbf{t}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T)$ and $\partial \ell(\boldsymbol{\theta} | \mathbf{y}, \mathbf{u}, \mathbf{v}, \mathbf{t}) / \partial \boldsymbol{\theta}$ are linear functions of \mathbf{y} . Hence, the expectation and the variance can be obtained easily from the binomial distribution (11) and the random samples obtained in the EM algorithm.

4. Simulation

We conducted a simulation study to investigate the performance of the proposed method and to compare its performance with the existing methods in the literature. In the simulation study, we generate data from a model specified by (2) and (3). We assume that $\mathbf{x}_{ij}^* \mathbf{u} = u_i$, $\mathbf{z}_{ij}^* \mathbf{v} = v_i$, and $u_i \sim N(0, \sigma_u^2)$, $v_i \sim N(0, \sigma_v^2)$, the correlation coefficient of u_i and v_i is ρ (that is, $\boldsymbol{\sigma} = (\sigma_u, \sigma_v, \rho)$). We consider two distributions for the baseline: the Weibull distribution with $S_{u0}(t) = \exp[-(t/\lambda)^k]$ and the lognormal distribution with $f_{u0}(t) = 1/(\sqrt{2\pi} \sigma t) \exp\{-[\log t - \mu]^2 / (2\sigma^2)\}$. The corresponding parameter values are chosen so that the mean and the variance of the two different baseline distributions are comparable. Only a binary covariate is considered, indicating treatment and control (baseline) groups for example, in both model (2) and model (3).

Table I. Parameter settings for simulation study.

Case	n	n_i	β	γ_0	γ_1	σ_u^2	σ_v^2	Baseline
1	100	4	log(0.5)	2	-1	0.25	0.49	Weibull ($\kappa=2, \lambda=1$)
2	100	4	log(0.5)	2	-1	0.25	0.49	Lognormal ($\mu=-\log(\sqrt{2}), \sigma=\sqrt{\log 2}$)
3	10	40	log(0.5)	2	-1	0.25	0.49	Weibull ($\kappa=2, \lambda=1$)
4	10	40	log(0.5)	2	-1	0.25	0.49	Lognormal ($\mu=-\log(\sqrt{2}), \sigma=\sqrt{\log 2}$)

Table II. MSEs and biases of estimated parameters in simulation study.

Case		Independent		Weibull		Piecewise	
		MSE	Bias	MSE	Bias	MSE	Bias
1	β	0.027	0.098	0.017	0.010	0.016	-0.007
	γ_0	0.074	-0.149	0.054	-0.046	0.057	-0.072
	γ_1	0.085	0.058	0.079	0.067	0.087	0.122
	σ_u	—	—	0.011	-0.046	0.014	-0.085
	σ_v	—	—	0.041	-0.096	0.054	-0.129
2	β	0.034	0.116	0.018	-0.051	0.024	-0.084
	γ_0	0.088	-0.193	0.030	-0.024	0.043	-0.049
	γ_1	0.087	0.027	0.050	-0.068	0.059	0.045
	σ_u	—	—	0.048	0.208	0.024	0.102
	σ_v	—	—	0.014	0.109	0.025	0.075
3	β	0.028	0.090	0.014	0.004	0.013	0.030
	γ_0	0.113	-0.141	0.077	0.022	0.075	0.017
	γ_1	0.091	0.055	0.057	0.013	0.057	0.029
	σ_u	—	—	0.011	-0.014	0.009	-0.028
	σ_v	—	—	0.017	-0.001	0.016	0.006
4	β	0.039	0.138	0.025	-0.059	0.016	-0.009
	γ_0	0.138	-0.186	0.076	-0.011	0.078	-0.016
	γ_1	0.089	0.025	0.067	-0.057	0.064	-0.047
	σ_u	—	—	0.020	0.093	0.016	0.056
	σ_v	—	—	0.025	0.055	0.022	0.005

We first consider cases with $\rho=0$. The values of the other parameters in the model are given in Table I. The values imply that the hazard of uncured patients in the treatment group is only half of the hazard of uncured patients in the control group, and the cure rates in the treatment and control groups are 27 and 12 per cent, respectively. In each cluster, individuals are randomly assigned to one of the two groups. We consider two situations for cluster size and the number of clusters: $n=100$ and $n_i=4$ for small cluster size situation and $n=10$ and $n_i=40$ for large cluster size situation. The latter is similar to the tonsil data described in Section 1. The censoring times are independently generated from a uniform distribution, and the resulting censoring rate is about 43 per cent.

For each case, we generate 500 data sets and fit the mixture cure random effects model with the proposed methods to the data sets. The Weibull distribution and the piecewise constant hazard distribution with $K=4$ are used as the baseline distribution in the model. The model is fit to each data set in the simulation via the Gaussian quadrature method because it is faster than the EM algorithm. We also fit the semiparametric mixture cure model [10, 11] that ignores the correlation among individuals within clusters. The mean square errors and biases of the estimated $\beta, \gamma_0, \gamma_1, \sigma_u$, and σ_v are summarized in Table II.

It is clear that ignoring the correlation within clusters lead to large biases in β, γ_0 , and γ_1 . The proposed method with both baseline assumptions performs well. The method with the Weibull baseline distribution assumption is slightly better in cases 1 and 2 (larger number of clusters and smaller cluster size), whereas the method with the piecewise constant hazard baseline distribution assumption is slightly better in cases 3 and 4 (smaller number of clusters and larger cluster size).

To examine the standard error estimation in these methods, we compute the coverage proportions of 95 per cent confidence intervals of the five parameters based on normal approximation $\theta \pm 1.96 \times \text{s.e.}(\theta)$. The results are summarized in Table III. It is obvious that ignoring the correlation makes the independent model underestimate the standard errors of the parameters and the confidence intervals too short to cover the true parameter values with correct coverage proportions. The proposed method, particularly with the piecewise baseline, provides the closest coverage proportions to the nominal 95 per cent in estimating β, γ_0 , and γ_1 . However, the coverages of σ_u are sometimes smaller than the nominal 95 per cent.

Table III. Coverage of 95 per cent confidence intervals of the estimated parameters from the independent and random effects models in simulation study.

Case	Parameter	Independent	Weibull	Piecewise
1	β	0.87	0.96	0.96
	γ_0	0.87	0.95	0.95
	γ_1	0.95	0.95	0.96
	σ_u	—	0.96	0.96
	σ_v	—	0.93	0.94
2	β	0.87	0.90	0.93
	γ_0	0.80	0.96	0.96
	γ_1	0.94	0.93	0.95
	σ_u	—	0.64	0.98
	σ_v	—	0.95	0.96
3	β	0.88	0.95	0.95
	γ_0	0.80	0.96	0.96
	γ_1	0.94	0.95	0.95
	σ_u	—	0.86	0.83
	σ_v	—	0.96	0.96
4	β	0.84	0.86	0.93
	γ_0	0.73	0.93	0.93
	γ_1	0.94	0.95	0.95
	σ_u	—	0.92	0.84
	σ_v	—	0.92	0.94

Table IV. MSEs and biases of estimated parameters for data with correlated and uncorrelated random effects.

Case		Weibull baseline				Piecewise constant baseline			
		Correlated		Uncorrelated		Correlated		Uncorrelated	
		MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
1	β	0.016	0.029	0.024	0.030	0.014	0.001	0.022	0.019
	γ_0	0.051	-0.086	0.072	-0.048	0.049	-0.079	0.078	-0.076
	γ_1	0.081	0.125	0.115	0.084	0.089	0.161	0.140	0.143
	σ_u	0.004	-0.032	0.011	-0.030	0.003	-0.035	0.014	-0.073
	σ_v	0.047	-0.212	0.036	-0.077	0.046	-0.210	0.052	-0.117
	ρ	0.020	-0.058	—	—	0.017	-0.058	—	—
2	β	0.026	0.004	0.031	-0.009	0.033	-0.067	0.036	-0.036
	γ_0	0.070	-0.169	0.059	-0.069	0.076	-0.138	0.072	-0.020
	γ_1	0.061	-0.022	0.086	-0.100	0.090	0.120	0.119	0.023
	σ_u	0.028	0.132	0.048	0.197	0.013	0.057	0.027	0.097
	σ_v	0.022	-0.113	0.015	0.098	0.030	-0.139	0.029	0.075
	ρ	0.013	0.043	—	—	0.020	0.001	—	—
3	β	0.017	-0.003	0.017	0.002	0.016	0.026	0.016	0.026
	γ_0	0.082	-0.013	0.106	0.013	0.086	0.033	0.107	0.016
	γ_1	0.073	0.021	0.080	0.015	0.071	0.031	0.080	0.029
	σ_u	0.013	-0.010	0.016	-0.010	0.011	-0.024	0.015	-0.026
	σ_v	0.047	-0.083	0.032	-0.005	0.043	-0.079	0.029	-0.004
	ρ	0.029	0.038	—	—	0.024	0.047	—	—
4	β	0.031	0.006	0.035	0.007	0.027	0.021	0.027	0.010
	γ_0	0.110	-0.064	0.116	-0.012	0.111	0.106	0.130	0.048
	γ_1	0.091	-0.098	0.102	-0.108	0.094	-0.060	0.099	-0.049
	σ_u	0.020	0.059	0.025	0.059	0.021	0.051	0.033	0.080
	σ_v	0.040	-0.001	0.040	0.082	0.043	-0.018	0.033	0.038
	ρ	0.024	0.072	—	—	0.028	0.055	—	—

Finally, we consider cases with $\rho \neq 0$. The settings of the cases are similar to the four cases described in Table I except that ρ is set to 0.7 instead of 0. We generated 500 data sets from each case and fit the mixture cure random effects model with and without the assumption $\rho \equiv 0$ to the data sets. The Weibull distribution and the piecewise constant hazard distribution with $K=4$ are used as the baseline distribution in the model. The mean square errors and biases of the estimated β , γ_0 , γ_1 , σ_u , σ_v , and ρ are summarized in Table IV. The results show that the proposed

method provides reasonable estimates of ρ . The inclusion of correlated random effects largely improves the estimation of the parameters in the model. However, using the Weibull or the piecewise constant baseline distributions makes little difference in the estimation. Therefore, we conclude that when the random effects u and v are strongly correlated, a mixture cure model with correlated random effects provides more accurate estimation of fixed effects on both cure fraction and on the distribution of the uncured patients than the model with independent random effects.

5. Application to tonsil cancer data

We consider the tonsil cancer data set described in Section 1. The data were analyzed by Withers *et al.* [1] and Sy and Taylor [11] with the mixture cure model without assuming possible correlation among patients from the same center. Peng *et al.* [2] further investigated the data set using a marginal mixture cure model. This model assumes that there is a correlation among the survival times of patients from the same center, but the model does not specify it. Instead, the model adjusted the standard errors of the estimates from a marginal model where the correlation is ignored. The disadvantage of the marginal model is that the correlation cannot be estimated and thus it is difficult to measure the strength of the correlation. The model may not be fully efficient compared with a model that appropriately specifies the correlation structure explicitly. Since the proposed model in this paper specifies the correlation structure, it is of great interest to apply the proposed model and the estimation method to the data and to compare the results from the model with those from the marginal model and the model for independent data.

When analyzing the data with the proposed model, in addition to neck node status, we also consider age, sex, T stages (stages I, II, III, and IV), total dose, and treatment duration. We will investigate the effects of the covariates on the time to local recurrence as well as on the probability of being cured under the proposed model by taking the possible correlation within centers into account. The baseline distribution is assumed to be the Weibull distribution or the piecewise constant hazard distribution with either $K=2$ or $K=4$ where τ_j 's are chosen to be the 50th percentile of the uncensored times ($K=2$) or the 25th, 50th, and 75th percentiles of the uncensored times ($K=4$).

Since we are interested in the potential correlation induced by the nine centers, we will consider the models (1), (2), (3) with shared random effects $\mathbf{x}_{ij}^* \mathbf{u} = u_i$ and $\mathbf{z}_{ij}^* \mathbf{v} = v_i$, and assume that (u_i, v_i) follows the bivariate normal distribution with $\boldsymbol{\sigma} = (\sigma_u, \sigma_v, \rho)$. We observed that $\hat{\rho} = 0.0001$ is close to 0 and there is no evidence from the likelihood ratio test to support $\rho \neq 0$. Therefore, we only report details of the model under the assumption $\rho = 0$ (i.e. independent u_i and v_i) in this section. The results of the models with three baseline distributions are summarized in Table V.

The table includes the maximum log-likelihood and the Akaike information criterion ($\text{AIC} = -2 \times \text{maximum log-likelihood} + 2 \times \text{the number of parameters estimated}$) to demonstrate the goodness of fit of the models. It is obvious that the model with the piecewise four constant hazards provides the best fit to the data among the three models based on these values.

The proposed random effects model shows that except for sex, all fixed covariates effects are significant or marginally significant in at least one of the two parts of the model. For example, age does not have a significant effect on the probability of being cured, but among those uncured patients a higher age tends to imply a later local recurrence. Node status only has a weak effect on both parts: patients having at least one positive node tend to have smaller probability of being cured and shorter survival time if not cured than the patients with only negative nodes. The rest of the covariates have significant effects on the probability of being cured, but have little impact on the failure time of uncured patients: larger total dose, shorter treatment duration, and lower T stages imply better cure rates.

The estimated standard deviation of the two random effects are $\sigma_u = 0.246$ and $\sigma_v = 0.047$, respectively. It shows that the correlation among the failure times of uncured patients from the same center is stronger than the correlation among the cure statuses of the patients. In fact, the correlation among the cure statuses of the patients from the same center is not significant. This can be seen from the fact the maximum log-likelihood and parameter estimates of the model do not change very much (not shown in this paper) when σ_v is set to 0. However, the correlation among the failure times of uncured patients from the same center is significant, which is demonstrated by comparing the results with those from the model when both σ_u and σ_v are set to 0. Without the random effects, this model reduces to the proportional hazard mixture cure model [10, 11] with the piecewise constant hazards baseline distribution for independent data. The results from this model are also summarized in Table V. The likelihood ratio test for testing $H_0: \sigma_u = 0$ produces a p -value 0.03 under the chi-square distribution with one degree of freedom. Since H_0 is on the boundary of the parameter space, this test may be conservative. Therefore, we conclude that $\sigma_u \neq 0$ and the correlation among the failure times of uncured patients is significant.

Table V. Estimated parameters and model statistics from the proposed random effects model applying to the tonsil cancer data with four different baseline distributions.

Covariate	Random effects model								Independent	
	Piecewise (4)		Piecewise (2)		Weibull		Marginal		Piecewise (4)	
	$\hat{\theta}$	$\frac{\hat{\theta}}{s.e.(\hat{\theta})}$	$\hat{\theta}$	$\frac{\hat{\theta}}{s.e.(\hat{\theta})}$	$\hat{\theta}$	$\frac{\hat{\theta}}{s.e.(\hat{\theta})}$	$\hat{\theta}$	$\frac{\hat{\theta}}{s.e.(\hat{\theta})}$	$\hat{\theta}$	$\frac{\hat{\theta}}{s.e.(\hat{\theta})}$
<i>Fixed effects on the failure time of uncured patients</i>										
Node	0.308	1.696	0.302	1.714	0.348	1.998	0.339	2.25	0.299	1.662
Dose	0.009	0.520	0.012	0.719	0.018	1.008	-0.005	-0.33	-0.001	-0.039
Duration	-0.038	-0.462	-0.044	-0.529	-0.057	-0.681	-0.006	-0.09	-0.027	-0.354
Age	-0.261	-2.632	-0.242	-2.517	-0.261	-2.690	-0.308	-2.45	-0.283	-3.002
T stage II	-0.581	-1.631	-0.656	-1.918	-0.650	-1.882	-0.628	-1.41	-0.612	-1.718
T stage III	-0.122	-0.361	-0.261	-0.806	-0.251	-0.764	-0.110	-0.28	-0.157	-0.456
T stage IV	0.336	0.885	0.232	0.628	0.385	1.044	0.383	1.27	0.354	0.937
Sex: male	0.006	0.030	-0.073	-0.407	-0.075	-0.418	0.062	0.30	0.011	0.059
<i>Random effects on the failure time of uncured patients</i>										
σ_u	0.246	1.799	0.266	1.905	0.309	2.096	—	—	—	—
<i>Fixed effects on the cure probability</i>										
Intercept	0.075	0.074	-0.795	-0.564	-0.756	-0.570	-0.198	-0.29	-0.011	-0.011
Node	0.373	1.862	0.385	1.927	0.463	2.331	0.356	2.73	0.378	1.883
Dose	-0.079	-4.292	-0.067	-3.011	-0.069	-3.220	-0.077	-4.42	-0.080	-4.387
Duration	0.468	3.694	0.474	3.395	0.462	3.408	0.464	2.95	0.479	3.778
Age	0.103	1.093	0.123	1.307	0.124	1.338	0.136	2.03	0.122	1.292
T stage II	0.805	2.281	0.767	2.186	0.821	2.321	0.854	2.42	0.826	2.331
T stage III	1.658	4.843	1.644	4.823	1.703	4.941	1.656	14.33	1.660	4.837
T stage IV	2.200	5.119	2.199	5.091	2.143	5.000	2.200	5.93	2.210	5.139
Sex: male	0.122	0.578	0.153	0.727	0.147	0.702	0.118	0.55	0.136	0.644
<i>Random effects on the cure probability</i>										
σ_v	0.047	0.456	0.234	1.058	0.191	0.911	—	—	—	—
<i>Model statistics</i>										
Log likelihood	-480.90		-495.65		-490.35		—		-483.20	
AIC	1007.8		1033.3		1022.7		—		1008.4	

Comparing the estimated parameters between the random effects mixture cure model and the independent mixture cure model, we can see that the fixed effects in the two models do not have substantial differences. The main difference is on the presence of the significant random effect for the failure times of uncured patients. However, there is a noticeable difference between the random effects mixture cure model and the marginal mixture cure model (also shown in Table V): the effect of neck node status becomes significant in the marginal model. Although the two approaches may be equivalent in certain special cases, the random effects mixture cure model considered in this paper is not equivalent to the marginal model. Thus, it is not surprising to see the difference in the result.

When using the EM algorithm to estimate the parameters, it produces a set of draws from the posterior distribution of the random effects u and v as well as the estimates of the parameters. The draws are useful to examine how the random effects vary in different clusters. For the tonsil data, the box-plots of the draws of the random effects u and v are presented in Figure 2. Note that the widths of the box-plots are proportional to the sizes of the clusters.

It is clear that v_i 's do not vary significantly in different centers and that their variances are smaller. It indicates weak correlation in cure rates within a center, as observed from the estimate of σ_v . However, u_i 's from some centers, such as centers 3, 4, 7 and 9, are substantially different from 0 and their predicted random effects are larger. It implies that the shared environment in these centers has a substantial impact on time to recurrence of uncured patients in the centers. The survival times of uncured patients in centers 4 and 7 tend to be shorter than the fixed effects suggested and those in centers 3 and 9 tend to be longer.

One possible explanation of the differences in the failure time distributions between institutions is that exactly when a recurrence is detected is strongly dependent on the schedule, frequency, and intensity of the follow-up for each patient. The pattern of monitoring each patient after the initial treatment is likely to differ between institutions that are in different continents. In contrast, the recurrence of cancer for each patient will eventually manifest itself if they are followed for long enough and is thus not dependent on the follow-up procedures at each institution.

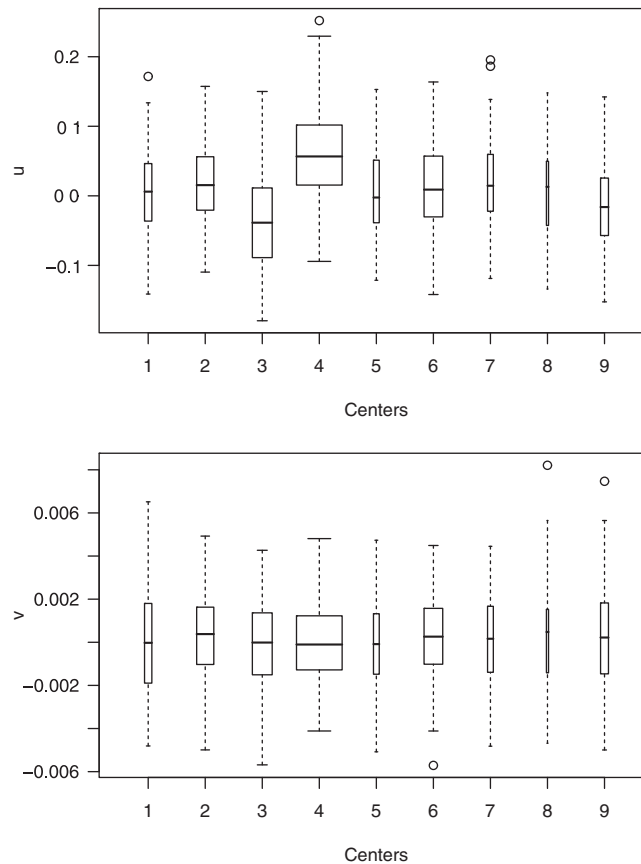


Figure 2. Box-plots of simulated values of the random effect u_i (top) and v_i (bottom) for nine centers.

6. Conclusions

In this study, we proposed a novel generalization of the semiparametric proportional hazards mixture cure model for clustered data. The extension allows different correlation structures for cure rates and for failure times of uncured patients, and the correlation structures are modeled by random effects. The proposed estimation method is based on the Gaussian quadrature approximation or the EM algorithm with the Monte-Carlo approximation, thus is computational intensive. However, the proposed method is very flexible and can be used for various random effect distributions and correlation structures. For example, we considered shared random effects in the numerical studies. We may also choose \mathbf{x}^* and \mathbf{z}^* such that the model is suitable for cluster data where the strength of correlation within a cluster can be different (see [14, 19], for more details). There are a number of distributions proposed in the literature for the random effects/frailties, such as normal, student's t , gamma/generalized gamma, positive stable, compound Poisson, etc. The proposed method can easily use such distributions for $\phi(\cdot)$ in (4) for the mixture cure model.

Compared with existing models and methods for clustered data with a cure fraction, the proposed estimation method has a number of advantages: It allows covariates for both failure time distribution of uncured patients and the cure proportion, can be semiparametric, and accommodates flexible correlation specification.

The model considered in Yau and Ng [8] and Lai and Yau [9] is similar to the model considered in this paper. They considered normal random effects only while our model can accept non-normal random effects. Their estimation method is based on BLUP and REML methods and thus is faster than the methods proposed in this paper. Our methods, on the other hand, produce maximum likelihood estimates of the parameters in the model. A simulation study demonstrates that the maximum likelihood estimates have small biases and mean square errors. The resulting maximum likelihoods from the proposed methods can be used to form information criteria that are particularly useful in model selection.

An R program was written for the proposed method and can be requested from the first author if one is interested in using the model to analyze clustered survival data with a cure fraction.

Appendix A: Monte-Carlo methods in E-step

In this appendix, we briefly outline two methods we used to generate random numbers from (12): rejection sampling and importance sampling methods [20]. If we denote the expectations in (9') and (10) generically by $E[\psi(\mathbf{u}, \mathbf{v})]$, the two sampling methods approximate the expectations by $\sum_m w^{(m)}\psi(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})/\sum_m w^{(m)}$, where $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})$, $m = 1, \dots, M$, are a sample from the density $\tilde{q}(\mathbf{u}, \mathbf{v})$ and $w^{(m)}$, $m = 1, \dots, M$, are weights for the samples. In the rejection sampling method, $w^{(m)} \equiv 1$ and the weighted average reduces to the simple average.

Our implementation of the rejection sampling is as follows. We sample from a multivariate normal distribution that approximates $q(\mathbf{u}, \mathbf{v}|\theta^{(r)})$. Particularly, we maximize $\log q(\mathbf{u}, \mathbf{v}|\theta^{(r)})$ with respect to (\mathbf{u}, \mathbf{v}) and let $g(\mathbf{u}, \mathbf{v})$ be the multivariate normal density function with mean equal to $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) = \arg \max_{\mathbf{u}, \mathbf{v}} \log q(\mathbf{u}, \mathbf{v}|\theta^{(r)})$ and the variance equal to the inverse of minus the second derivative of $\log q(\mathbf{u}, \mathbf{v}|\theta^{(r)})$ evaluated at $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$. We sample $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})$ in the following steps:

1. Sample $(\mathbf{u}^*, \mathbf{v}^*)$ independently from $g(\mathbf{u}, \mathbf{v})$.
2. Sample w from a uniform distribution in $(0, 1)$.
3. If $w \leq q(\mathbf{u}^*, \mathbf{v}^*)/(Cg(\mathbf{u}^*, \mathbf{v}^*))$, accept \mathbf{u}^* and \mathbf{v}^* as a sample $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})$ from $q(\mathbf{u}, \mathbf{v})$. Otherwise, reject $(\mathbf{u}^*, \mathbf{v}^*)$ and return to step 1.

The constant C is an upper bound of the ratio $q(\mathbf{u}, \mathbf{v})/g(\mathbf{u}, \mathbf{v})$. We set C to $\max\{q(\mathbf{u}^*, \mathbf{v}^*)/g(\mathbf{u}^*, \mathbf{v}^*)\}$ among those sampled $(\mathbf{u}^*, \mathbf{v}^*)$. The steps are repeated until M samples of \mathbf{u} and \mathbf{v} are obtained.

The importance sampling method to obtain $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)}, w^{(m)})$ is implemented as follows:

1. Sample $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})$ independently from $g(\mathbf{u}, \mathbf{v})$.
2. Compute weight $w^{(m)} = q(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})/g(\mathbf{u}^{(m)}, \mathbf{v}^{(m)})$.

The steps are repeated until M triples of $(\mathbf{u}^{(m)}, \mathbf{v}^{(m)}, w^{(m)})$ are obtained. As in other applications, the importance sampling tends to be faster than the rejection sampling, but the latter produces a sample from the conditional distribution of (\mathbf{u}, \mathbf{v}) .

Acknowledgements

The authors thank the Editor, an Associate Editor, and three reviewers for their insightful comments, which have led to a great improvement in the presentation of the study. The study of Yingwei Peng is supported in part by research grants from the Natural Sciences and Engineering Research Council of Canada and Queen's University.

References

1. Withers HR, Peters LJ, Taylor JMG, Owen JB, Morrison WH, Schultheiss TE, Keane T, O'Sullivan B, van Dyk J, Gupta N, Wang CC, Jones CU, Doppke KP, Myint S, Thompson M, Parsons JT, Mendenhall WM, Dische S, Aird EGA, Henk JM, Bidmean MAM, Svoboda V, Chon Y, Hanlon AL, Peters TL, Hanks GE. Local control of carcinoma of the tonsil by radiation therapy: an analysis of patterns of fractionation in nine institutions. *International Journal of Radiation Oncology, Biology, Physics* 1995; **33**:549–562.
2. Peng Y, Taylor JMG, Yu B. A marginal regression model for multivariate failure time data with a surviving fraction. *Lifetime Data Analysis* 2007; **13**:351–369.
3. Chen MH, Ibrahim JG, Sinha D. Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis* 2002; **80**:101–126.
4. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**:141–152.
5. Yin G. Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics* 2005; **61**:552–558.
6. Chatterjee N, Shih J. A bivariate cure-mixture approach for modeling familial association in disease. *Biometrics* 2001; **57**:779–786.
7. Wienke A, Lichtenstein P, Yashin AI. A bivariate frailty model with a cure fraction for modeling familial correlation in disease. *Biometrics* 2003; **59**:1178–1183.
8. Yau KKW, Ng ASK. Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in Medicine* 2001; **20**:1591–1607.
9. Lai X, Yau KKW. Long-term survivor model with bivariate random effects: applications to bone marrow transplant and carcinoma study data. *Statistics in Medicine* 2008; **27**:5692–5708.
10. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**:237–243.
11. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.
12. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 1995; **4**:12–35.
13. Johansen S. An extension of Cox's regression model. *International Statistical Review* 1983; **51**:165–174.
14. Ripatti S, Larsen K, Palmgren J. Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Analysis* 2002; **8**:349–360.
15. Oakes D. Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1999; **61**:479–482.

16. Meng XL, Rubin DB. Using EM to obtain asymptotic variance–covariance matrices: the SEM algorithm. *Journal of the American Statistical Association* 1991; **86**:899–909.
17. Meilijson I. A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B* 1989; **51**:127–138.
18. Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1982; **44**:226–233.
19. Hougaard P. *Analysis of Multivariate Survival Data*. Springer: New York, 2000.
20. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall/CRC: Boca Raton, FL, U.S.A., 2004.