

A Frailty Model for Informative Censoring

Xuelin Huang* and Robert A. Wolfe

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

*email: xueling@umich.edu

SUMMARY. To account for the correlation between failure and censoring, we propose a new frailty model for clustered data. In this model, the risk to be censored is affected by the risk of failure. This model allows flexibility in the direction and degree of dependence between failure and censoring. It includes the traditional frailty model as a special case. It allows censoring by some causes to be analyzed as informative while treating censoring by other causes as noninformative. It can also analyze data for competing risks. To fit the model, the EM algorithm is used with Markov chain Monte Carlo simulations in the E-steps. Simulation studies and analysis of data for kidney disease patients are provided. Consequences of incorrectly assuming noninformative censoring are investigated.

KEY WORDS: Clustered data; Competing risks; Dependent censoring; EM algorithm; Survival analysis.

1. Introduction

This article proposes a model to analyze clustered survival data with dependent censoring. There is a rich literature on the analysis of clustered or, more generally, correlated survival data. There are also many studies considering the problem of dependent censoring. However, these two problems, namely dependent censoring and the correlation introduced by clustering, are rarely considered simultaneously. Nevertheless, this type of data are very common in real practice. To illustrate, we present the following example.

In the Dialysis Outcome and Practice Patterns Study (Young et al., 2000), there are 10,290 kidney disease patients from 152 centers. All of them received dialysis treatment. Dialysis practice patterns vary from center to center and can affect mortality. Some features of the data set are the following. First, the survival outcomes of patients in the same center are correlated because they share the same service or practice patterns. Second, some patients withdraw during the study. Their withdrawal is very likely to be related to their health status. As an initial step to assess dependent censoring, we calculate the percentage dead and percentage withdrawn for each center. The correlation coefficient between these two groups of percentages is 0.25; that is to say, we cannot naively assume that all censoring is noninformative.

One major objective of this study is to identify good and bad dialysis practice patterns. To do this, it is desirable to rank facilities according to patient mortality. The crude percentages dead can only serve as a rough index to rank facilities because they do not take the effects of covariates into account. For example, some facilities may have higher mortality rates just because their patients are older. This article provides a method to rank those centers, taking into account covariate effects and the information provided by censoring.

Various approaches have been proposed to identify and ac-

count for dependent censoring. Emoto and Matthews (1990) assumed a bivariate Weibull model for failure and censoring times. Zheng and Klein (1995) used an assumed copula to study dependent competing risks. These articles include excellent reviews of the literature. Some authors used data collected after censoring to identify dependent censoring (e.g., Lin, Robins, and Wei, 1996; Lee and Wolfe, 1998). If there are marker process data available, then the method developed by Robins and Rotnitzky (1992) can be used to identify dependent censoring. However, all of these studies considered uncorrelated subjects only.

To analyze survival data collected from correlated subjects, frailty models can be used. The term frailty, which means propensity to failure, was first introduced by Vaupel, Manton, and Stallard (1979) to model population heterogeneity. Frailty models assume that subjects in the same cluster share a common unobserved frailty. By doing this, the within-cluster correlation between failure times is taken into account. Clayton (1978) used it to study the familial tendency in chronic disease incidence. Hougaard, Harvald, and Holm (1992) analyzed the lifetimes of twins. Gray (1994) examined the institutional effects in a multicenter cancer clinical trial.

Either explicitly or implicitly, all of these studies made the assumption that censoring is noninformative for frailty. This assumption was first clearly stated by Nielsen et al. (1992). They emphasized that this assumption is “for mere validity of the inference, not just for possible efficiency.” However, this assumption may not be true in real life. As can be seen from the above example, a kidney center’s mortality rate is positively correlated with its withdrawal rate. Therefore, withdrawal appears to be informative for failure.

Link (1989) did a sensitivity analysis using a frailty model with informative censoring. His model assumes that censoring only occurs in a subpopulation defined by the frailty distri-

bution. However, he used data from independent subjects. Frailty models for independent subjects are overparameterized. Therefore, some parameters in his model are specified by prior knowledge rather than being estimated. Nevertheless, his study produced “a class of survival functions so that the potential effect of incorrectly assuming the independent censoring model can be assessed.”

Our frailty model uses clustered data and is based on two assumptions. The first is that subjects in the same cluster share a common frailty. The second is that, within each cluster, censoring is independent of survival. These two assumptions are common to all standard frailty models. The new feature of our model is that it allows the censoring rate to be associated with failure rate at cluster level so that censoring can be informative for failure. Previous frailty models assumed independence between a cluster’s censoring and failure rates so that censoring is noninformative for failure.

The proposed model allows censoring by some causes to be analyzed as informative while treating censoring by other causes as noninformative. It can also be applied to multiple competing risk problems. Moreover, when many medical facilities participate in the study, the new model can identify facilities doing relatively better and those doing worse, with patient withdrawal information taken into account. This is more reasonable than assuming that patients’ withdrawals are independent of their survival outcomes.

To fit frailty models, Nielsen et al. (1992) proposed using the EM algorithm (Dempster, Laird, and Rubin, 1977). Klein (1992) and Andersen et al. (1997) modified and improved the algorithm. When integrals are not available in closed form in the E-steps of the EM algorithm, many authors have used the Markov chain Monte Carlo (MCMC) method, e.g., Wei and Tanner (1990) and Chan and Ledolter (1995). In this article, we also use the EM algorithm with the MCMC method to fit the proposed model. We assume a log-normal distribution for frailty. Different distributions for frailty and their properties were discussed by Hougaard (1984) and Oakes (1989), among others.

The frailty model with informative censoring is described in Section 2. The model is fitted by the EM algorithm in Section 3. The simulation studies in Section 4 evaluate the finite-sample properties of the proposed method. A data analysis for kidney disease patients is given in Section 5. Consequences of incorrectly assuming noninformative censoring are investigated in these two sections. Some further discussions about the proposed model are given in Section 6. Detailed formulas are listed in the Appendix.

2. Frailty Models with Informative Censoring

In this article, we consider right censoring only. Suppose there are two types of censoring, namely dropout and administrative censoring. Dropout could be any type of potentially informative censoring, such as initiating a nonrandomized therapy in the middle of a randomized trial. Administrative censoring is usually caused by the end of the study.

Denote the failure, dropout, and administrative censoring time for a subject by T , C , and S , respectively. Let $X = \min(T, C, S)$. Define $\Delta^{(T)} = 1$ if $X = T$ and $\Delta^{(T)} = 0$ otherwise. Similarly, define $\Delta^{(C)}$. S is assumed to be independent of both T and C and will not be discussed in detail. From now on, censoring will be used to mean dropout only. The

end of the study is denoted by τ . $Z^{(T)}$ and $Z^{(C)}$ are vectors of covariates associated with failure and censoring, with size $p \times 1$ and $q \times 1$, respectively. They may be completely distinct or overlapping or even identical. They are assumed to be time independent in this article, but the method can also be applied to external time-dependent covariates (Kalbfleish and Prentice, 1980, p. 122).

Suppose there are m clusters, n_i subjects in the i th cluster. The total sample size is then $N = \sum_{i=1}^m n_i$. Variables for the j th subject in the i th cluster are indicated by subscripts, e.g., T_{ij} , $\Delta_{ij}^{(T)}$, etc. The corresponding lowercase letters are used to indicate realized values.

Following Cox (1972), the baseline hazards for failure and censoring are nonparametric functions of time u , denoted by $\lambda_0^{(T)}(u)$ and $\lambda_0^{(C)}(u)$, respectively. Subjects in the i th cluster are assumed to share a common unobserved frailty e^{B_i} . The shared frailty introduces correlation among subjects in the same cluster. It is further assumed that, conditional on covariates and frailty, all N subjects are independent. Moreover, B_i , $i = 1, \dots, m$, are assumed to be independent with identical normal distribution with mean zero and variance θ , i.e.,

$$B_i \sim N(0, \theta), \quad \text{i.i.d., } i = 1, \dots, m. \quad (1)$$

The hazard functions for failure and censoring are then written as follows:

$$\begin{aligned} \lambda_{ij}^{(T)}(u | Z_{ij}^{(T)}, Z_{ij}^{(C)}, B_i, C_{ij}) \\ = \lambda_0^{(T)}(u) \exp(\beta^{(T)'} Z_{ij}^{(T)} + B_i) \end{aligned} \quad (2)$$

$$\begin{aligned} \lambda_{ij}^{(C)}(u | Z_{ij}^{(T)}, Z_{ij}^{(C)}, B_i, T_{ij}) \\ = \lambda_0^{(C)}(u) \exp(\beta^{(C)'} Z_{ij}^{(C)} + \alpha B_i). \end{aligned} \quad (3)$$

Note that T and C are independent conditional on covariates and B_i . The marginal dependence between T and C is solely due to the fact that the unobserved B_i affects both failure and censoring. In this model, the risk of dropout for a subject with frailty e^{b_i} is inflated by a factor of $(e^{b_i})^\alpha$, as opposed to a subject with frailty $e^0 = 1$. If $\alpha > 0$, higher frailty will result in earlier dropout. On the other hand, if $\alpha < 0$, a subject with higher frailty will be more likely to stay (in the study, in hospital, etc.). Although typically we shall have $\alpha > 0$ in real life, it is also possible that $\alpha < 0$. In either case, frailty affects censoring, so censoring is informative for frailty. Conditional on covariates only, failure time T and censoring time C are dependent. The correlation between $\log(T)$ and $\log(C)$ introduced by this model does not have a closed form. To illustrate the relationship between this correlation and α , we chose $\lambda_0^{(T)}(u) = \lambda_0^{(C)}(u) = 0.1$, no covariates, and α in the interval from -1.5 to 1.5 in (2) and (3). We estimated the correlation by simulation, and the results are plotted in Figure 1. It can be seen that the correlation has the same sign as α and that, as $|\alpha|$ increases, the magnitude of the correlation increases.

On the other hand, the assumption for the traditional frailty model is that censoring is noninformative for frailty, i.e., $\alpha = 0$ in (3). By this assumption, frailty does not affect censoring, so censoring cannot provide any information about frailty. In

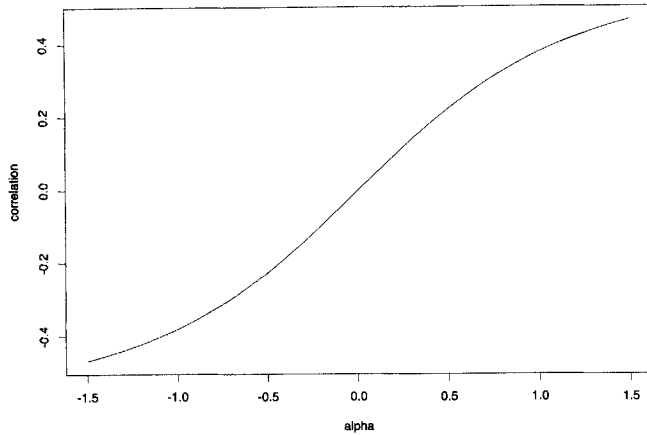


Figure 1. The relationship between α and the correlation coefficient between $\log(T)$ and $\log(C)$ when the hazard functions for T and C are the following: $\lambda^{(T)}(u | B) = 0.1 \exp(B)$, $\lambda^{(C)}(u | B) = 0.1 \exp(\alpha B)$, and $B \sim N(0, 1)$.

this case, a model for censoring does not need to be specified because it does not affect the inference for $\beta^{(T)}$ and θ . By the traditional frailty model, failure time T and censoring time C are independent, conditional on covariates. The purpose of the traditional model is to account for the within-cluster correlation between failure times, not to account for the dependence between T and C .

In the proposed model, censoring is independent (of failure) within each cluster but dependent (on failure) in the whole population. This means, e.g., within each kidney center, patients withdraw randomly. However, those centers with higher failure rates also have higher withdrawal rates, i.e., $\alpha > 0$ in (3). Hence, in the whole population, failure and withdrawal are positively correlated. For another example, let T and C denote potential failure times due to heart attack and diabetes, respectively. The frailty term can be viewed as the presence or absence of some disease genes shared by family members that cause both heart attacks and diabetes. Within each family, these two risks are independent. However, those families with more heart attacks also have more diabetes. This corresponds to $\alpha > 0$ in (3). Consequently, in the whole population, the two risks are positively correlated. These scenarios are the background for the proposed frailty model.

The proposed model can be easily extended to deal with more than two competing risks. For example, we can assume the hazard functions for T_1, T_2 , and T_3 are the following, where T_1, T_2 , and T_3 are the potential failure times due to three different causes:

$$\begin{cases} \lambda_1(u | Z_1, Z_2, Z_3, B, T_2, T_3) = \lambda_{1,0}(u) \exp(\beta'_1 Z_1 + B) \\ \lambda_2(u | Z_1, Z_2, Z_3, B, T_1, T_3) = \lambda_{2,0}(u) \exp(\beta'_2 Z_2 + \alpha_2 B) \\ \lambda_3(u | Z_1, Z_2, Z_3, B, T_1, T_2) = \lambda_{3,0}(u) \exp(\beta'_3 Z_3 + \alpha_3 B). \end{cases} \quad (4)$$

This model will be used to analyze the data set for kidney disease patients in Section 5. More generally, we can assume that the three types of failures have frailty B_1, B_2 , and B_3 and assume a multivariate normal distribution for them.

3. Fitting the Model by the MCEM Algorithm

We now study how to fit the proposed frailty model in (2) and (3). Its extended versions can be fitted similarly and are not discussed. Let $\Lambda_0^{(T)}(u) = \int_0^u \lambda_0^{(T)}(v)dv$ and $\Lambda_0^{(C)}(u) = \int_0^u \lambda_0^{(C)}(v)dv$, the nonparametric baseline cumulative hazards for failure and censoring, respectively. For simplicity, denote parameters and data by single letters as follows (the letter \mathbf{o} is used to denote observed data):

$$\begin{aligned} \Psi &= \left(\beta^{(T)'}, \beta^{(C)'}, \alpha, \theta, \lambda_0^{(T)}(u), \lambda_0^{(C)}(u), 0 \leq u \leq \tau \right)', \\ \mathbf{o}_{ij} &= \left(x_{ij}, \delta_{ij}^{(T)}, z_{ij}^{(T)'}, \delta_{ij}^{(C)}, z_{ij}^{(C)'} \right)', \\ \mathbf{o}_i &= (\mathbf{o}_{i1}, \dots, \mathbf{o}_{in_i})', \\ \mathbf{o} &= (\mathbf{o}'_1, \dots, \mathbf{o}'_m)', \\ \mathbf{B} &= (B_1, \dots, B_m)', \\ \mathbf{b} &= (b_1, \dots, b_m)'. \end{aligned} \quad (5)$$

Then, conditional on $B_i = b_i$, the likelihood term for subject (i, j) is

$$\begin{aligned} L_{ij}(\mathbf{o}_{ij}; \Psi, b_i) &= \left\{ \lambda_0^{(T)}(x_{ij}) \exp(\beta^{(T)'} z_{ij}^{(T)} + b_i) \right\}^{\delta_{ij}^{(T)}} \\ &\times \exp \left\{ -\Lambda_0^{(T)}(x_{ij}) \exp(\beta^{(T)'} z_{ij}^{(T)} + b_i) \right\} \\ &\times \left\{ \lambda_0^{(C)}(x_{ij}) \exp(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i) \right\}^{\delta_{ij}^{(C)}} \\ &\times \exp \left\{ -\Lambda_0^{(C)}(x_{ij}) \exp(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i) \right\}. \end{aligned} \quad (6)$$

Note the first two lines are for failure and the last two lines are for censoring. Usually the noninformative censoring assumption is made, which makes it appropriate to use only the first two lines and ignore the last two lines of (6) to make inference about failure risks. When censoring is informative, the last two lines cannot be ignored. Actually, there is a third part, which is for administrative censoring. However, it is assumed to be independent of both parameters of interest and frailty term b_i . Therefore, the third part is ignored. When subject (i, j) is administratively censored, both $\delta_{ij}^{(T)}$ and $\delta_{ij}^{(C)}$ are zero. The marginal likelihood for the i th cluster is

$$P_i(\mathbf{o}_i; \Psi) = \int_{-\infty}^{\infty} f(b_i; \theta) \prod_{j=1}^{n_i} L_{ij}(\mathbf{o}_{ij}; \Psi, b_i) db_i, \quad (7)$$

where $f(\cdot; \theta)$ is the probability density function of the $N(0, \theta)$ distribution. This integral does not have a closed form. The probability density function of B_i conditional on observed data is

$$f_{B_i | \mathbf{o}_i}(b_i; \Psi) = \frac{f(b_i; \theta) \prod_{j=1}^{n_i} L_{ij}(\mathbf{o}_{ij}; \Psi, b_i)}{P_i(\mathbf{o}_i; \Psi)}. \quad (8)$$

The marginal full likelihood is the product of (7) over $i = 1, \dots, m$. It is very difficult to maximize it directly.

However, if those $b_i, i = 1, \dots, m$, are known, then we can fit (2) and (3) separately. This feature makes the EM algorithm a good choice to fit the model. In the E-steps,

the density of B_i in (8) is used to generate Q random numbers, $b_i^{(s)}$, $s = 1, \dots, Q$, by the Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). Then B_i and functions of B_i are estimated by their sample means conditional on observed data. For example,

$$\hat{E}(B_i | \mathbf{o}_i) = \frac{1}{Q} \sum_{s=1}^Q b_i^{(s)}, \quad (9)$$

$$\hat{E}\{\exp(B_i) | \mathbf{o}_i\} = \frac{1}{Q} \sum_{s=1}^Q \exp(b_i^{(s)}). \quad (10)$$

Note that (10) gives an estimator for frailty that will be used later in Section 5 for data analysis.

By the profile likelihood viewpoint of Johansen (1983), $\lambda_0^{(T)}(\cdot)$ and $\lambda_0^{(C)}(\cdot)$ are allowed to be discrete. Then they have point masses at failure and censoring time points, respectively, and are equal to zero everywhere else. The following four profile score equations are used in the M-steps to update parameter estimates for $\beta^{(T)}$, $\beta^{(C)}$, α , and θ , respectively:

$$U_T \left\{ \beta^{(T)} \right\} = \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(T)} \left[z_{ij}^{(T)} - \bar{z}^{(T)}(x_{ij}) \right], \quad (11)$$

$$U_C \left\{ \beta^{(C)}, \alpha \right\} = \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(C)} \left[z_{ij}^{(C)} - \bar{z}^{(C)}(x_{ij}) \right], \quad (12)$$

$$U_A \left\{ \beta^{(C)}, \alpha \right\} = \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(C)} \left[\hat{E}(B_i | \mathbf{o}_i) - \bar{b}(x_{ij}) \right], \quad (13)$$

$$U_V(\theta) = \frac{1}{2\theta^2} \sum_{i=1}^m \left\{ \hat{E}(B_i^2 | \mathbf{o}_i) - \theta \right\}, \quad (14)$$

where

$$\begin{aligned} \bar{z}^{(T)}(x_{ij}) &= \frac{\sum_{x_{kl} \geq x_{ij}} z_{kl}^{(T)} w_{kl}^{(T)} \left\{ \beta^{(T)} \right\}}{\sum_{x_{kl} \geq x_{ij}} w_{kl}^{(T)} \left\{ \beta^{(T)} \right\}}, \\ \bar{z}^{(C)}(x_{ij}) &= \frac{\sum_{x_{kl} \geq x_{ij}} z_{kl}^{(C)} w_{kl}^{(C)} \left\{ \beta^{(C)}, \alpha \right\}}{\sum_{x_{kl} \geq x_{ij}} w_{kl}^{(C)} \left\{ \beta^{(C)}, \alpha \right\}}, \\ \bar{b}(x_{ij}) &= \frac{\sum_{x_{kl} \geq x_{ij}} \hat{E}\{B_k \exp(\alpha B_k) | \mathbf{o}_i\} \exp\left\{ \beta^{(C)'} z_{kl}^{(C)} \right\}}{\sum_{x_{kl} \geq x_{ij}} w_{kl}^{(C)} \left\{ \beta^{(C)}, \alpha \right\}}, \end{aligned}$$

and where

$$\begin{aligned} w_{kl}^{(T)} \left\{ \beta^{(T)} \right\} &= \exp \left\{ \beta^{(T)'} z_{kl}^{(T)} \right\} \hat{E}\{\exp(B_k) | \mathbf{o}_i\}, \\ w_{kl}^{(C)} \left\{ \beta^{(C)}, \alpha \right\} &= \exp \left\{ \beta^{(C)'} z_{kl}^{(C)} \right\} \hat{E}\{\exp(\alpha B_k) | \mathbf{o}_i\}. \end{aligned}$$

Note the first three equations have the same format as the Cox (1972) score equations. Once $\beta^{(T)}$, $\beta^{(C)}$, and α are obtained, baseline hazard estimates are updated by the Breslow (1972)

formula,

$$\lambda_0^{(T)}(x_{ij}) = \frac{\delta_{ij}^{(T)}}{\sum_{x_{kl} \geq x_{ij}} w_{kl}^{(T)} \left\{ \beta^{(T)} \right\}} \quad (15)$$

$$\lambda_0^{(C)}(x_{ij}) = \frac{\delta_{ij}^{(C)}}{\sum_{x_{kl} \geq x_{ij}} w_{kl}^{(C)} \left\{ \beta^{(C)}, \alpha \right\}}. \quad (16)$$

The computation of the covariance matrix for parameter estimates is given in the Appendix.

4. Simulation Studies

Simulation studies are conducted to assess the performance of the method developed in the previous section. In all simulations, there are 40 clusters and 5 subjects in each cluster. Each subject has probability 0.5 of receiving a treatment ($TR = 1$) and probability 0.5 of being a control ($TR = 0$). Subject age after being centered has a uniform($-10, 10$) distribution. All subjects are independent to each other with regard to covariate distributions.

The true model used to generate data is a frailty model with informative censoring. Covariate effects are specified by the following scenario. The treatment is effective to lower failure risk but may also be toxic to some patients and make them drop out. Older people are more likely to fail and also more likely to drop out. Specifically, the model is

$$B_i \sim N(0, 1), \quad \text{i.i.d., } i = 1, \dots, m,$$

$$\lambda_{ij}^{(T)}(u) = 0.2u \exp(0.1AGE_{ij} - 1.4TR_{ij} + B_i),$$

$$\lambda_{ij}^{(C)}(u) = 0.04u \exp(0.2AGE_{ij} + 1.2TR_{ij} + 1.0B_i),$$

i.e., $\text{var}(B_i) = \theta = 1$, $\beta_{\text{age}}^{(T)} = 0.1$, $\beta_{\text{tr}}^{(T)} = -1.4$, $\beta_{\text{age}}^{(C)} = 0.2$, $\beta_{\text{tr}}^{(C)} = 1.2$, and $\alpha = 1.0$. That α is positive indicates that clusters with higher frailty levels have higher dropout rates. The administrative censoring time S_{ij} is generated by a uniform(a_1, a_2) distribution, where a_1, a_2 are chosen to get desired censoring rates. Five hundred data sets were generated. Each data set was analyzed by two frailty models, one assuming informative censoring and the other assuming noninformative censoring. The results are summarized in Table 1.

First, we can see that parameter estimates by our proposed model are consistent. The magnitudes of empirical bias are very small. Also, the estimated parameter standard errors are fairly close to the empirical standard errors. The coverage probabilities of the 95% confidence intervals are close to the nominal level. Second, it can be seen that, if we misspecify censoring as noninformative, a frailty model can result in substantial bias in parameter estimates. The coverage probabilities of the 95% confidence intervals are far from the nominal level. Each data set is also analyzed by a Cox model assuming independent subjects. Parameters are biased by about 30% toward zero (results not shown).

To assess the bias directions by frailty models in different situations when informative censoring is ignored, we set α to -1.0 and keep everything else the same as previously. Again, 500 data sets are generated and analyzed by both models. The results are listed in the bottom half of Table 1. It can be seen

Table 1
 Parameter estimates by frailty models; 500 replicates: 40 clusters with 5 subjects in each cluster; 35% dropout, 20% administration censoring^{a,b}

	True	Informative censoring					Noninformative censoring				
		Mean	Bias (%)	SE	SEM	CP	Mean	Bias ^c (%)	SE	SEM	CP
$\alpha = 1$											
$10\beta_{\text{age}}^{(T)}$	1.0	1.023	2.3	0.251	0.256	.952	0.887	-11.3	0.248	0.233	.818
$\beta_{\text{tr}}^{(T)}$	-1.4	-1.423	1.6	0.288	0.291	.954	-1.510	+7.9	0.300	0.289	.940
θ	1.0	0.977	2.3	0.405	0.417	.930	0.813	-18.7	0.392	0.188	.482
$10\beta_{\text{age}}^{(C)}$	2.0	2.012	0.6	0.311	0.302	.942					
$\beta_{\text{tr}}^{(C)}$	1.2	1.225	2.1	0.322	0.331	.960					
α	1.0	1.027	2.7	0.291	0.322	.962					
$\alpha = -1$											
$10\beta_{\text{age}}^{(T)}$	1.0	0.996	0.4	0.251	0.243	.940	1.085	+8.5	0.263	0.257	.898
$\beta_{\text{tr}}^{(T)}$	-1.4	-1.415	1.1	0.276	0.289	.954	-1.335	-4.6	0.285	0.262	.908
θ	1.0	1.001	0.1	0.458	0.497	.928	0.846	-15.4	0.407	0.193	.510
$10\beta_{\text{age}}^{(C)}$	2.0	2.052	2.6	0.305	0.308	.956					
$\beta_{\text{tr}}^{(C)}$	1.2	1.199	0.1	0.306	0.312	.952					
α	-1.0	-1.034	3.4	0.337	0.328	.930					

^a SE is the empirical standard error calculated from 500 parameter estimates; SEM is the mean of the 500 estimated standard errors of the parameters; CP is the empirical coverage probability of the estimated 95% confidence intervals.

^b The ‘informative censoring’ columns are the results from the proposed model, which analyzes dropout as informative censoring, specified as follows:

$$\begin{aligned}
 B_i &\sim N(0, \theta), \quad \text{i.i.d., } i = 1, \dots, m, \\
 \lambda_{ij}^{(T)}(u) &= \lambda_0^{(T)}(u) \exp\left(\beta_{\text{age}}^{(T)} \text{AGE}_{ij} + \beta_{\text{tr}}^{(T)} \text{TR}_{ij} + B_i\right), \\
 \lambda_{ij}^{(C)}(u) &= \lambda_0^{(C)}(u) \exp\left(\beta_{\text{age}}^{(C)} \text{AGE}_{ij} + \beta_{\text{tr}}^{(C)} \text{TR}_{ij} + \alpha B_i\right).
 \end{aligned}$$

The ‘noninformative censoring’ columns are the results from the following frailty model, which treats dropout as noninformative censoring:

$$\begin{aligned}
 B_i &\sim N(0, \theta), \quad \text{i.i.d., } i = 1, \dots, m, \\
 \lambda_{ij}^{(T)}(u) &= \lambda_0^{(T)}(u) \exp\left(\beta_{\text{age}}^{(T)} \text{AGE}_{ij} + \beta_{\text{tr}}^{(T)} \text{TR}_{ij} + B_i\right).
 \end{aligned}$$

No model for censoring is specified except the assumption that frailty B_i does not affect censoring. Both models treat administrative censoring as noninformative.

^c + indicates overestimation, or bias away from zero, and - indicates underestimation, or bias toward zero.

that the estimates are consistent when informative censoring is taken into account whereas ignoring informative censoring gives biased results.

The consequences of falsely assuming noninformative censoring can be explained as following. These results are verified by extensive simulations. When $\alpha > 0$, those subjects with higher failure risks are more likely to drop out before we observe their failures. Then the result is that failure risks are underestimated. In contrast, when $\alpha < 0$, failure risks are overestimated. The bias direction in the estimation of $\beta^{(T)}$ is more complicated. It depends not only on α but also on $\beta^{(C)}$. Consider the estimation of treatment effect. When $\alpha > 0$, the failure risks in both the treated and the control groups are underestimated. However, when $\beta^{(C)} > 0$, the treated group has

a higher censoring rate. Hence, the degree of underestimation for failure risks is more severe in the treated group than in the control group. Therefore, the reduction of failure risk by the treatment is inflated, i.e., $\beta^{(T)}$ is overestimated when $\alpha > 0$ and $\beta^{(C)} > 0$. On the other hand, when $\alpha > 0$ but $\beta^{(C)} < 0$, $\beta^{(T)}$ is underestimated. The bias directions in other settings of α and $\beta^{(C)}$ can be similarly explained. In summary, ignoring informative censoring causes bias in the estimation of survival probability. With unbalanced censoring rates across covariate groups, it causes bias in the estimation of covariate effects. When censoring rates are balanced across covariate groups, the bias in covariate effect due to ignoring informative censoring is small. The estimation for θ is severely biased when censoring is incorrectly assumed to be noninformative.

5. Example: Mortality at Kidney Dialysis Centers

End-stage renal disease is a chronic condition of total and irreversible kidney failure. Dialysis is the treatment for patients before they receive kidney transplants. During the treatment, a tube is inserted into the patient's arm. A special fluid in the tube is driven by a machine outside the patient's body. It flows opposite to the patient's vein blood. Toxin in the blood is filtered through the tube into the fluid and taken away. This treatment usually takes 3–4 hours. Patients are usually treated three times a week. The total volume of the fluid used in a treatment depends on the patient's weight. However, different kidney dialysis facilities have different formulas to calculate the total volume. Many treatment characteristics, including treatment-time length, speed of the fluid flow, and type of dialysis machine, vary greatly from facility to facility. We refer to these factors as practice patterns.

The Dialysis Outcomes and Practice Patterns Study (DOPPS) is a prospective, observational study for the associations between dialysis treatment practices and patient outcomes (Young et al., 2000). Data were collected during 1996–1999 in seven countries. Here we use only the subset of the data for the United States. It contains 10,290 patients from 152 dialysis facilities. The number of patients in a facility ranged from 21 to 124. During the study, 3188 (31%) patients died, 424 (4%) withdrew from the study, 673 (6.5%) received transplants, and the rest remained in the study and on dialysis. It is suspected that the most common reason for withdrawal is worsened health status. Thus, withdrawal is likely to be an informative type of censoring. Transplantation may also be informative. Censoring by the end of the study is assumed to be noninformative.

In this data set, each facility has a moderate or large number of patients. Thus, we can calculate percentage dead and percentage withdrawn for each facility. Facility-level percentages of patients who died range from 0 to 48% (median = 31%, mean = 29%, standard deviation [SD] = 12%). Percentages withdrawn range from 0 to 20% (median = 2.9%, mean = 3.6%, SD = 3.7%). The percentages dead and withdrawn are positively correlated (Pearson correlation coefficient $r = 0.25$). Percentages of transplant range from 0 to 22% (median = 5.8%, mean = 6.1%, SD = 4.4%). The correlation coefficient between facility-level percentage dead and percentage of transplants is 0.17. This suggests that these two types of censoring, i.e., due to withdrawal and transplant, are informative.

Because patients in the same facility share the same practice patterns, their survival outcomes are likely to be correlated. Therefore, frailty models are used to do the analysis. Each facility is viewed as a cluster and its patients as subjects in the cluster. Four important patient-level covariates in kidney studies (Wolfe, 1994) are included in the models. They are age (in years), race (black = 1, other = 0), gender (male = 1, female = 0), and diabetes status (diabetes = 1, no diabetes = 0). The data set is analyzed by the frailty model with two types of informative censoring, specified in (4). It is also analyzed by a frailty model assuming noninformative censoring. Parameter estimates are reported in Table 2. If we use the informative censoring model as a gold standard, then the relative bias of the noninformative censoring model may be up to 30%. The effect of frailty on withdrawal is substantial ($\alpha_2 = 1.432$, with SE = 0.321). That α_2 is positive

indicates that facilities with higher mortality rates also had higher withdrawal rates after being adjusted by covariates. Similarly, $\alpha_3 = 0.745$ (SE = 0.226) tells us that, in this data set, those facilities with higher mortality rates also had higher rates of transplantation.

A major objective of this study is to identify good and bad dialysis practice patterns. To this end, we rank facilities according to patient mortality. The crude percentages of dead can only serve as a rough index to rank facilities because they do not take the effects of covariates into account. For example, some facilities may have higher mortality rates just because their patients are older. However, frailty models can be used to rank the performance of facilities. After adjustment by patient covariates, the facility-level frailty estimated by (10) can be viewed as a score for facility performance. Higher frailty levels imply worse performance. Again, two frailty models are used, one assuming informative censoring and the other noninformative censoring. The correlation coefficient between the ranks by the two models is 0.88. The ranks are plotted in Figure 2. Facilities are divided into two groups, high dropout and low dropout, by the median of their withdrawal percentages. The two groups are denoted by dots and circles. For simplicity, we do not show transplantation information in the plot. It can be seen that the disagreement between ranks by the two models has a pattern. Facilities with high withdrawal rates are penalized when informative censoring is taken into account. We believe this is more likely to be fair.

Note that some facilities have withdrawal rates up to 20% while many other facilities have withdrawal rates near zero. Therefore, even though the overall withdrawal rate is as low as 4%, when ranking those facilities, withdrawal still plays an important role. Moreover, if we consider that only 31% died, the 4% of patients withdrawn could also have a substantial effect on parameter estimates. For example, if those withdrawn patients had died immediately after withdrawal, then the percentage of dead would have increased to 35%. Ignoring the 4% withdrawn would be ignoring one ninth of the total number of deaths. This could produce roughly an 11% relative bias on survival probability. This is another reason for considering informative censoring even though the overall withdrawal rate is low. Similar arguments can be applied to transplantation. Withdrawal and transplantation combined account for 10% of the patients. A 10% potentially informative censoring rate cannot be simply taken as noninformative. However, it is not appropriate to combine them as a single type of informative censoring. This article provides a method to analyze this type of data.

6. Discussion

We view our contribution in this article as bias reduction, the ability to estimate the correlation between failure and censoring, and more reasonable ranks for medical facilities (or physicians, geographic areas, etc.).

In this article, for convenience, the distribution of frailty is chosen to be log normal. This assumption can be checked. This distributional assumption could influence the relative ranking of the facilities. Many other authors have used gamma frailty distributions, also for convenience. Shih and Louis (1995) and Glidden (1999) proposed methods to check the appropriateness of gamma frailty distributions. It should not

Table 2
Parameter and standard error estimates for DOPPS^a

	Parameter	Informative censoring		Noninformative censoring		Bias ^b (%)
		Estimate	SE	Estimate	SE	
Failure	$\beta_{\text{age}}^{(T)}$	0.0323	0.00138	0.0326	0.00138	+0.6
	$\beta_{\text{race}}^{(T)}$	-0.0988	0.0445	-0.132	0.0469	+33.6
	$\beta_{\text{gender}}^{(T)}$	0.0184	0.0361	0.0231	0.0362	+25.5
	$\beta_{\text{diabetic}}^{(T)}$	0.192	0.0361	0.189	0.0363	-1.6
	θ	0.0987	0.0204	0.102	0.0113	+3.3
Withdrawal	$\beta_{\text{age}}^{(C)}$	-0.0272	0.00303			
	$\beta_{\text{race}}^{(C)}$	-0.473	0.115			
	$\beta_{\text{gender}}^{(C)}$	-0.143	0.0979			
	$\beta_{\text{diabetic}}^{(C)}$	0.0220	0.100			
	α_2	1.432	0.321			
Transplant	$\beta_{\text{age}}^{(R)}$	-0.0411	0.00236			
	$\beta_{\text{race}}^{(R)}$	-0.754	0.0923			
	$\beta_{\text{gender}}^{(R)}$	0.202	0.0799			
	$\beta_{\text{diabetic}}^{(R)}$	-0.229	0.0834			
	α_3	0.745	0.226			

^a The ‘informative censoring’ columns are the results from the following frailty model, which analyzes withdrawal and transplantation as two types of informative censoring:

$$\begin{aligned}
 & B_i \sim N(0, \theta), \quad \text{i.i.d., } i = 1, \dots, m, \\
 & \lambda_{ij}^{(T)}(u) = \lambda_0^{(T)}(u) \exp \left(\beta_{\text{age}}^{(T)} \text{Age}_{ij} + \beta_{\text{race}}^{(T)} \text{Race}_{ij} + \beta_{\text{gender}}^{(T)} \text{Gender}_{ij} + \beta_{\text{diabetic}}^{(T)} \text{Diabetic}_{ij} + B_i \right), \\
 & \lambda_{ij}^{(C)}(u) = \lambda_0^{(C)}(u) \exp \left(\beta_{\text{age}}^{(C)} \text{Age}_{ij} + \beta_{\text{race}}^{(C)} \text{Race}_{ij} + \beta_{\text{gender}}^{(C)} \text{Gender}_{ij} + \beta_{\text{diabetic}}^{(C)} \text{Diabetic}_{ij} + \alpha_2 B_i \right), \\
 & \lambda_{ij}^{(R)}(u) = \lambda_0^{(R)}(u) \exp \left(\beta_{\text{age}}^{(R)} \text{Age}_{ij} + \beta_{\text{race}}^{(R)} \text{Race}_{ij} + \beta_{\text{gender}}^{(R)} \text{Gender}_{ij} + \beta_{\text{diabetic}}^{(R)} \text{Diabetic}_{ij} + \alpha_3 B_i \right).
 \end{aligned}$$

The ‘noninformative censoring’ columns are the results from the following frailty model, which treats withdrawal and transplantation as noninformative censoring:

$$\begin{aligned}
 & B_i \sim N(0, \theta), \quad \text{i.i.d., } i = 1, \dots, m, \\
 & \lambda_{ij}^{(T)}(u) = \lambda_0^{(T)}(u) \exp \left(\beta_{\text{age}}^{(T)} \text{Age}_{ij} + \beta_{\text{race}}^{(T)} \text{Race}_{ij} + \beta_{\text{gender}}^{(T)} \text{Gender}_{ij} + \beta_{\text{diabetic}}^{(T)} \text{Diabetic}_{ij} + B_i \right).
 \end{aligned}$$

No model for censoring is specified except the assumption that frailty B_i does not affect censoring. Both models treat censoring by the end of the study as noninformative.

^b Bias is calculated using the results in the ‘informative censoring’ column as the gold standard.

be hard to develop similar methods to check the appropriateness of the log-normal frailty distribution, but we have not yet seen any publications on this topic. Presumably it is more desirable to develop a method to find the right frailty distribution.

The ability of clustered data to identify dependent censoring comes from the assumption that subjects in the same cluster are similar, i.e., after being adjusted by covariates, they have the same risks to fail and the same risks to be censored. Subjects in the same cluster serve as replicates of each other.

The frailty distribution assumption is not essential for identifying dependent censoring. We expect other distributions lead to similar conclusions. We may even do it nonparamet-

rically. If cluster sizes are large, the B_i ’s can be reliably estimated as fixed intercept terms rather than random variables with a parametric distribution. Thus, both the conditional failure time distribution and the frailty values can be estimated nonparametrically. The model is parametric for the covariate effects and the effect of frailty on censoring, which is modeled through α in (3). These together yield a semiparametric estimator for the marginal failure time distribution. When cluster sizes are small, this approach will give unstable estimates for B_i ’s and other parameters. In this case, a parametric distribution for B_i is useful, similar to the use of a prior in Bayesian analysis. Different distributions can lead to different inferences for small sample sizes but will give es-

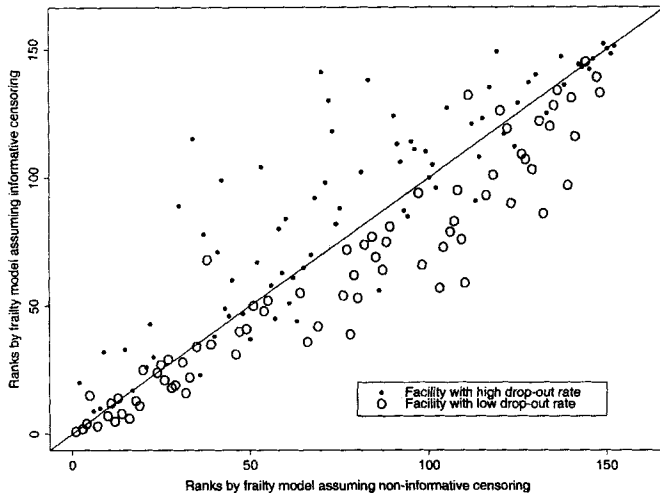


Figure 2. Facilities in the DOPPS study are ranked by the estimates of their frailty e^{B_i} . The facility with the lowest frailty is ranked as the first. The ranks on the y -axis are obtained from the new frailty model assuming informative censoring. The ranks on the x -axis are from a frailty model with the same covariates, but censoring is assumed to be non-informative. The median of the dropout percentages of the 152 facilities is used as the cutting point for high and low dropout rates.

essentially the same inference when the sample size is large in each cluster.

We assumed a common baseline hazard function for all clusters with a random effect for each cluster. Alternatively, when cluster sizes are large, a model with fixed cluster-specific effects or cluster-stratified hazards could be used. These models have fewer parametric assumptions than our model but could be inefficient, especially when the number of clusters is large and the sample size is small in each cluster. Our proposed model is a proportional hazards model with random intercepts (B_i in (2) and αB_i in (3)). It has an added advantage because it can be easily extended to include random covariate effects, i.e., random slopes. Vaida and Xu (2000) provided examples to show the usefulness of proportional hazards models with random slopes.

In this article, we have assumed that there is a significant clustering effect, i.e., the parameter θ , variance of log frailty, is assumed to be greater than zero. If this is in question, the tests provided by Commenges and Andersen (1995), Gray (1995), and Andersen, Klein, and Zhang (1999) can be used. Our model can be applied only if significant clustering effects are found.

Murphy (1994, 1995) and Parner (1998) derived the large-sample theory for the gamma-frailty model assuming non-informative censoring. Their proof may be modified to give asymptotics for our frailty model with informative censoring. This is still an open question, but our simulation results seem to point to the asymptotic validity of the proposed method.

RÉSUMÉ

Nous proposons un nouveau modèle de fragilité qui prend en compte la corrélation entre censure et décès pour des données

en grappe. Dans ce modèle, le risque d'être censuré est affecté par le risque de décès. Ce modèle s'adapte au sens et au degré de dépendance entre décès et censure. Le modèle de fragilité traditionnel en est un cas particulier. Notre modèle permet de traiter la censure pour certaines causes comme informative, et la censure pour d'autres causes comme non-informative. Il permet aussi d'analyser des données avec risques compétitifs. Pour ajuster le modèle un algorithme EM avec simulations MCMC dans l'étape E est utilisé. La méthode est illustrée par des simulations et une analyse de données de malades atteints de maladies du rein. Les conséquences d'une hypothèse incorrecte de censure non-informative sont étudiées.

REFERENCES

- Andersen, P. K., Klein, J. P., Knudsen, K. M., and Palacios, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics* **53**, 1475–1484.
- Andersen, P. K., Klein, J. P., and Zhang, M. (1999). Testing for centre effects in multi-centre survival studies: A Monte Carlo comparison of fixed and random effects tests. *Statistics in Medicine* **18**, 1489–1500.
- Breslow, N. E. (1972). Discussion of the paper by D. R. Cox cited below. *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation of time series models involving counts. *Journal of the American Statistical Association* **90**, 242–252.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.
- Commenges, D. and Andersen, P. K. (1995). Score test of homogeneity for survival data. *Lifetime Data Analysis* **1**, 145–160.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood estimation for incomplete data via the EM algorithms (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Emoto, S. E. and Matthews, P. C. (1990). A Weibull model for dependent censoring. *Annals of Statistics* **18**, 1556–1577.
- Glidden, D. V. (1999). Checking the adequacy of gamma frailty model for multivariate failure data. *Biometrika* **86**, 381–393.
- Gray, R. J. (1994). A Bayesian analysis of institutional effects in a multi-center clinical trial. *Biometrics* **50**, 244–253.
- Gray, R. J. (1995). Tests for variation over groups in survival data. *Journal of the American Statistical Association* **90**, 198–203.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- Hoggard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, 75–83.

- Hougaard, P., Harvald, B., and Holm, N. V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *Journal of the American Statistical Association* **87**, 17–25.
- Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51**, 165–174.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–806.
- Lee, S. and Wolfe, R. A. (1998). A simple test for independent censoring under the proportional hazard model. *Biometrics* **54**, 1176–1182.
- Lin, D. Y., Robins, J. M., and Wei, L. J. (1996). Comparing two failure time distributions on the presence of dependent censoring. *Biometrika* **83**, 381–393.
- Link, W. A. (1989). A model for informative censoring. *Journal of American Statistical Association* **84**, 749–752.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226–233.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.
- Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics* **22**, 712–731.
- Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Annals of Statistics* **23**, 182–198.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.
- Oakes, D. (1989). Bivariate survival models induced by frailty. *Journal of the American Statistical Association* **84**, 487–493.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty. *Annals of Statistics* **26**, 183–214.
- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell (eds), 297–331. Boston: Birkhäuser.
- Shih, J. H. and Louis, T. A. (1995). Assessing gamma frailty model for clustered failure time data. *Lifetime Data Analysis* **1**, 205–220.
- Vaida, F. and Xu, R. (2000). Proportional hazard model with random effects. *Statistics in Medicine* **19**, 3309–3324.
- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the Poor Man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.
- Wolfe, R. A. (1994). The standardized mortality ratio revisited: improvements, innovations, and limitations. *American Journal of Kidney Disease* **24**, 290–297.
- Young, E. W., Goodkin, D. A., Mapes, D. L., Port, F. K., Keen, M. L., Chen, K., Maroni, R. L., Wolfe, R. A., and Held, P. J. (2000). The dialysis outcomes and practice patterns study (DOPPS): An international hemodialysis study. *Kidney International* **57**(Suppl. 74), S74–S81.
- Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–138.

Received April 2001. Revised February 2002.

Accepted March 2002.

APPENDIX

Derivatives and Covariance Estimates

Denote by $\lambda^{(T)}$ the vector of the discrete baseline failure hazards on those death time points. Similarly, use $\lambda^{(C)}$ to denote that for censoring hazards. Then let $\eta = (\beta^{(T)'}, \lambda^{(T)'}, \beta^{(C)'}, \alpha, \lambda^{(C)'}, \theta)'$. If those frailty terms $B_i, i = 1, \dots, m$, were known, then the logarithm of the likelihood for complete data (\mathbf{o}, \mathbf{b}) is the following, where \mathbf{o} denotes the observed data and \mathbf{b} the unobserved frailty:

$$\begin{aligned}
 l(\mathbf{o}, \mathbf{b}; \eta) = & \sum_{i=1}^m \left(-\frac{1}{2} \log \theta - \frac{b_i^2}{2\theta} \right) \\
 & + \sum_{i=1}^m \sum_{j=1}^{n_i} \left(\delta_{ij}^{(T)} \left[\log \left\{ \lambda_0^{(T)}(x_{ij}) \right\} + \beta^{(T)'} z_{ij}^{(T)} + b_i \right] - \Lambda_0^{(T)}(x_{ij}) \exp \left\{ \beta^{(T)'} z_{ij}^{(T)} + b_i \right\} \right) \\
 & + \sum_{i=1}^m \sum_{j=1}^{n_i} \left(\delta_{ij}^{(C)} \left[\log \left\{ \lambda_0^{(C)}(x_{ij}) \right\} + \beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right] - \Lambda_0^{(C)}(x_{ij}) \exp \left\{ \beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right\} \right). \quad (\text{A.1})
 \end{aligned}$$

Denote $l(\mathbf{o}, \mathbf{b}; \eta)$ simply by $l(\eta)$. Note that the second derivative matrix of $l(\eta)$ is not the observed information matrix for η because the uncertainty about \mathbf{b} has not been accounted for. The observed information matrix $I(\hat{\eta})$ is given by Louis (1982) as the following (Parner (1998) proved the consistency of this formula in nonparametric maximum likelihood estimation for

gamma-frailty models):

$$I(\hat{\eta}) = -\hat{E} \left\{ \frac{\partial^2 l(\eta)}{\partial \eta \partial \eta'} \mid \mathbf{o}, \hat{\eta} \right\} - \hat{E} \left\{ \frac{\partial l(\eta)}{\partial \eta} \frac{\partial l(\eta)}{\partial \eta'} \mid \mathbf{o}, \hat{\eta} \right\}, \quad (\text{A.2})$$

where expectations are taken with respect to frailty terms \mathbf{b} . The components of $\partial l(\eta)/\partial \eta$ are

$$\frac{\partial l}{\partial \beta^{(T)}} = \sum_{i=1}^m \sum_{j=1}^{n_i} z_{ij}^{(T)} \left\{ \delta_{ij}^{(T)} - \Lambda_0^{(T)}(x_{ij}) \exp \left(\beta^{(T)'} z_{ij}^{(T)} + b_i \right) \right\}, \quad (\text{A.3})$$

$$\frac{\partial l}{\partial \lambda_0^{(T)}(x_{ij})} = \frac{\delta_{ij}^{(T)}}{\lambda_0^{(T)}(x_{ij})} - \sum_{x_{kl} \geq x_{ij}} \exp \left(\beta^{(T)'} z_{kl}^{(T)} + b_k \right), \quad (\text{A.4})$$

$$\frac{\partial l}{\partial \beta^{(C)}} = \sum_{i=1}^m \sum_{j=1}^{n_i} z_{ij}^{(C)} \left\{ \delta_{ij}^{(C)} - \Lambda_0^{(C)}(x_{ij}) \exp \left(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right) \right\}, \quad (\text{A.5})$$

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^m \sum_{j=1}^{n_i} b_i \left\{ \delta_{ij}^{(C)} - \Lambda_0^{(C)}(x_{ij}) \exp \left(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right) \right\}, \quad (\text{A.6})$$

$$\frac{\partial l}{\partial \lambda_0^{(C)}(x_{ij})} = \frac{\delta_{ij}^{(C)}}{\lambda_0^{(C)}(x_{ij})} - \sum_{x_{kl} \geq x_{ij}} \exp \left(\beta^{(C)'} z_{kl}^{(C)} + \alpha b_k \right), \quad (\text{A.7})$$

$$\frac{\partial l}{\partial \theta} = -\frac{1}{2} \left(\frac{m}{\theta} - \frac{\sum_{i=1}^m b_i^2}{\theta^2} \right). \quad (\text{A.8})$$

Solving (A.4) and (A.7) gives Breslow estimators for $\lambda^{(T)}$ and $\lambda^{(C)}$, respectively. Plugging them into (A.3), (A.5), and (A.6) leads to score equations (11)–(13). The second derivatives are

$$\frac{\partial^2 l}{\partial \beta^{(T)} \partial \beta^{(T)'}} = - \sum_{i=1}^m \sum_{j=1}^{n_i} z_{ij}^{(T)} z_{ij}^{(T)'} \Lambda_0^{(T)}(x_{ij}) \exp \left(\beta^{(T)'} z_{ij}^{(T)} + b_i \right), \quad (\text{A.9})$$

$$\frac{\partial^2 l}{\partial \left\{ \lambda_0^{(T)}(x_{ij}) \right\}^2} = - \frac{\delta_{ij}^{(T)}}{\left\{ \lambda_0^{(T)}(x_{ij}) \right\}^2}, \quad (\text{A.10})$$

$$\frac{\partial^2 l}{\partial \beta^{(T)} \partial \lambda^{(T)}(x_{ij})} = - \sum_{x_{kl} \geq x_{ij}} z_{kl}^{(T)} \exp \left(\beta^{(T)'} z_{kl}^{(T)} + b_k \right), \quad (\text{A.11})$$

$$\frac{\partial^2 l}{\partial \beta^{(C)} \partial \beta^{(C)'}} = - \sum_{i=1}^m \sum_{j=1}^{n_i} z_{ij}^{(C)} z_{ij}^{(C)'} \Lambda_0^{(C)}(x_{ij}) \exp \left(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right), \quad (\text{A.12})$$

$$\frac{\partial^2 l}{\partial \alpha^2} = - \sum_{i=1}^m \sum_{j=1}^{n_i} b_i^2 \Lambda_0^{(C)}(x_{ij}) \exp \left(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right), \quad (\text{A.13})$$

$$\frac{\partial^2 l}{\partial \left\{ \lambda_0^{(C)}(x_{ij}) \right\}^2} = - \frac{\delta_{ij}^{(C)}}{\left\{ \lambda_0^{(C)}(x_{ij}) \right\}^2}, \quad (\text{A.14})$$

$$\frac{\partial^2 l}{\partial \beta^{(C)} \partial \alpha} = - \sum_{i=1}^m \sum_{j=1}^{n_i} b_i z_{ij}^{(C)} \Lambda_0^{(C)}(x_{ij}) \exp \left(\beta^{(C)'} z_{ij}^{(C)} + \alpha b_i \right), \quad (\text{A.15})$$

$$\frac{\partial^2 l}{\partial \beta^{(C)} \partial \lambda^{(C)}(x_{ij})} = - \sum_{x_{kl} \geq x_{ij}} z_{kl}^{(C)} \exp \left(\beta^{(C)'} z_{kl}^{(C)} + \alpha b_k \right), \quad (\text{A.16})$$

$$\frac{\partial^2 l}{\partial \alpha \partial \lambda^{(C)}(x_{ij})} = - \sum_{x_{ki} \geq x_{ij}} b_k \exp \left(\beta^{(T)'} z_{kl}^{(T)} + \alpha b_k \right), \quad (\text{A.17})$$

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{1}{2} \left(-\frac{m}{\theta^2} + \frac{2 \sum_{i=1}^m b_i^2}{\theta^3} \right). \quad (\text{A.18})$$

All other off-diagonal elements of the second derivatives are zero. Using matrix techniques, we can avoid solving high-dimensional equations in the M-steps of the EM algorithm.

The computation of the expectation in the first term of (A.2) is similar to that of the E-steps. To compute that in the second term, we calculate $(\partial l(\eta)/\partial \eta)(\partial l(\eta)/\partial \eta')$ by letting $b_i = b_i^{(s)}$ in (A.3)–(A.8) for each $s = 1, \dots, Q$, and then take the average. The inverse of $I(\hat{\eta})$ gives the covariance matrix for $\hat{\eta}$.