Additive Mixed Effect Model for Clustered Failure Time Data

Jianwen Cai[*] and **Donglin Zeng**[**]
Department of Biostatistics, University of North Carolina at Chapel Hill,North Carolina 27599-7420, USA

Summary

We propose an additive mixed effect model to analyze clustered failure time data. The proposed model assumes an additive structure and include a random effect as an additional component. Our model imitates the commonly used mixed effect models in repeated measurement analysis but under the context of hazards regression; our model can also be considered as a parallel development of the gamma-frailty model in additive model structures. We develop estimating equations for parameter estimation and propose a way of assessing the distribution of the latent random effect in the presence of large clusters. We establish the asymptotic properties of the proposed estimator. The small sample performance of our method is demonstrated via a large number of simulation studies. Finally, we apply the proposed model to analyze data from a diabetic study and a treatment trial for congestive heart failure.

Keywords

Additive models; Clustered survival; Goodness of fit; Hazards rate; Moment methods; Random effects

1. Introduction

Clustered failure time data are commonly seen in biomedical studies. A popular model for analyzing clustered failure time data is the gamma-frailty model, which models the intra-cluster dependence by incorporating an unobserved random effect, the so-called frailty, into the Cox (1972) proportional hazards model. The asymptotic distribution of the maximum likelihood estimator for the gamma frailty model has been rigorously established by Murphy (1994; 1995) and by Parner (1998).

The multiplicative hazards model focuses on estimating hazard ratios and its multiplicative structure may not model real data well in many situations. In some cases, an additive effect could be a more reasonable association. Such an effect can be modeled using the so-called additive hazards model (Aalen, 1989; Huffer and McKeague, 1991; Lin and Ying, 1994; and McKeague and Sasieni, 1994; among others). In this model, an additive structure of a baseline hazards function and a covariate effect is assumed via the expression $d\Lambda(t|\mathbf{X}) = d\Lambda(t) + \mathbf{X}(t)^T \boldsymbol{\beta} dt$, where $\Lambda(t|\mathbf{X})$ denotes the cumulative hazard function for the given, possibly time-dependent, covariates $\mathbf{X}$ and (t) is the baseline cumulative hazard function. The regression coefficient $\boldsymbol{\beta}$ in the additive model is interpreted as risk difference and it has

[*]cai@bios.unc.edu
[**]dzeng@bios.unc.edu

been eloquently advocated and successfully utilized for right-censored independent survival data in many papers, e.g., Andersen, Borgan, Gill, and Keiding (1993, pp. 563–566), Lin and Ying (1994), McKeague and Sasieni (1994), Shen and Cheng (1999), and Gandy and Jensen (2005a; 2005b). The multivariate version of the above additive model has been used to model clustered failure time data in Yin and Cai (2004): $d\Lambda_{ij}(t|\mathbf{X}) = d\Lambda(t) + \mathbf{X}_{ij}(t)^T \boldsymbol{\beta} dt$, where $\Lambda_{ij}(t|\mathbf{X})$ denotes the cumulative hazard function for subject j in cluster i and $\mathbf{X}_{ij}$ is the associated covariate. Yin (2007) further developed a test for checking the additive structure using clustered data.

All the previous additive models assume the marginal relationship between covariates and survival times. In other words, they do not model the dependence among these events explicitly. Therefore, the marginal models cannot be used for individual prediction given the status of other individuals in the same cluster; neither can they be used to assess the dependence between individuals. In this paper, we propose an additive mixed effect model for clustered failure times. Our model includes a cluster-specific random effect as an additional component. Thus, our model imitates the commonly used mixed effect models in repeated measurement analysis but under the context of hazards regression; our model can also be considered as a parallel development of the gamma-frailty model in additive model structures. However, different from the gamma-frailty model, the proposed model still induces an additive marginal model. In Section 2, we introduce the proposed model and provide estimation procedure of the model parameters. We then establish the asymptotic distribution of the estimators. In Section 3, we demonstrate small-sample performance via a large number of simulation studies. When cluster size is large, we propose a consistent testing procedure in Section 4 to evaluate the distribution of random effects. The application to a diabetic study and the SOLVD treatment trial (SOLVD Investigators, 1990) is given in Section 5.

2. Model and Inference Procedure

2.1 Models and data

Suppose that data are collected from n i.i.d clusters. Within cluster i ($i = 1, \ldots, n$), there are n_i subjects and we use T_{ij} to denote the failure time of subject j in this cluster. We denote $\mathbf{X}_{ij}$ as the covariates associated with this subject.

Since the failure times from the same cluster may be correlated, in order to account for such dependence, we introduce a cluster-specific random effect ξ_i which is independent of covariates. We assume that $T_{i1}, \ldots, T_{in_i}$ are conditionally independent given all the covariates and the random effect ξ_i. Additionally, if letting $\Lambda_{ij}(t)$ be the cumulative hazards function given all the covariates and the cluster-specific random effect, our semiparametric additive mixed effect model assumes

$$d\Lambda_{ij}(t) = d\Lambda(t) + \mathbf{X}_{ij}(t)^T \beta dt + \xi_i dt. \tag{1}$$

Here, $\Lambda(t)$ is an unknown baseline function and $\boldsymbol{\beta}$ is an unknown coefficient vector. We also assume ξ_i follows a one-parameter distribution with density function $f(x; \theta)$ which has mean zero and a finite moment generating function. Note that our model (1) imitates the usual mixed effect model with a random intercept in the analysis of longitudinal data; more interestingly, it can be considered as an additive counterpart of the usual multiplicative frailty model for clustered failure time.

In practice, failure time T_{ij} may be right censored. We denote C_{ij} as the censoring time for subject j in cluster i and assume $(C_{i1}, \ldots, C_{in_i})$ are independent of $(T_{i1}, \ldots, T_{in_i})$ and ξ_i given all the covariates. Subject to censoring, the actual observations from these clusters are $(Z_{ij} = T_{ij} \wedge C_{ij}, \mathbf{X}_{ij}, \Delta_{ij} = I(T_{ij} \leq C_{ij})), j = 1, \ldots, n_i, i = 1, \ldots, n$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. We let τ be the study duration.

2.2 Parameter estimation and inference

To estimate parameters $(\boldsymbol{\beta}, \theta)$ and $\Lambda(t)$, we use the moment methods to construct estimating equations. First, from model (1), it is easy to calculate the marginal survival function T_{ij} given covariates $\mathbf{X}_{ij}$ which is given by

$$
\begin{aligned}
P(T_{ij} > t | \mathbf{X}_{ij}) &= E_{\xi_i} \left[\exp \left\{ -\Lambda(t) - \int_0^t \mathbf{X}_{ij}(u)^T \beta du - \xi_i t \right\} | \mathbf{X}_{ij} \right] \\
&= \exp \left\{ -\Lambda(t) - \int_0^t \mathbf{X}_{ij}(u)^T \beta du - G(t; \theta) \right\},
\end{aligned}
\tag{2}
$$

where $\exp\{-G(t; \theta)\} = \int e^{-xt} f(x; \theta) dx$. Equivalently, if $N_{ij}(t)$ denotes the counting process $\Delta_{ij} I(Z_{ij} \leq t)$ and Y_{ij} denotes the at-risk process $I(Z_{ij} \geq t)$, then the relationship in (2) implies

$$
E[dN_{ij}(t) | \mathbf{X}_{ij}, Z_{ij} \geq t] = Y_{ij}(t) \left\{ dH(t) + \mathbf{X}_{ij}(t)^T \beta dt \right\},
\tag{3}
$$

where $H(t) = \Lambda(t) + G(t; \theta)$.

Since equation (2) implies another marginal additive model studied in Lin and Ying (1994), the estimating equations developed in their paper can be used here. Following Lin and Ying (1994), we can construct the following estimating equation to estimate $\boldsymbol{\beta}$ and H:

$$
\sum_{i=1}^{n} \sum_{j=1}^{n_i} \int_0^\tau Y_{ij}(t) I(t \leq s) \left\{ dN_{ij}(t) - dH(t) - \mathbf{X}_{ij}(t)^T \beta dt \right\} = 0, \quad s > 0,
\tag{4}
$$

and

$$
\sum_{i=1}^{n} \sum_{j=1}^{n_i} \int_0^\tau Y_{ij}(t) \mathbf{X}_{ij}(t) \left\{ dN_{ij}(t) - dH(t) - \mathbf{X}_{ij}(t)^T \beta dt \right\} = 0.
\tag{5}
$$

The solutions to these equations have an explicit expression and we denote them as $\hat{\boldsymbol{\beta}}$ and $\hat{H}$ respectively. Specifically,

$$
\widehat{\beta} = \left[\sum_{i=1}^{n} \sum_{j=1}^{n_i} \int_0^\tau Y_{ij}(t) \{ \mathbf{X}_{ij}(t) - \overline{\mathbf{X}}(t) \}^{\otimes 2} dt \right]^{-1} \left[\sum_{i=1}^{n} \sum_{j=1}^{n_i} \int_0^\tau Y_{ij}(t) \{ \mathbf{X}_{ij}(t) - \overline{\mathbf{X}}(t) \} dN_{ij}(t) \right],
$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ and $\overline{\mathbf{X}}(t) = \sum_{k=1}^{n} \sum_{l=1}^{n_k} Y_{kl}(t) \mathbf{X}_{kl}(t) / \sum_{k=1}^{n} \sum_{l=1}^{n_k} Y_{kl}(t)$. Additionally,

$$\widehat{H}(s) = \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{ij}(t)\{dN_{ij}(t) - \mathbf{X}_{ij}(t)^T \widehat{\beta} dt\}}{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{ij}(t)}.$$

From (3), $\hat{H}(t)$ estimates the function which involves both $\Lambda(t)$ and θ. Thus, we need additional information to be able to discriminate between these two parameters. Note that θ describes the dependence among the failure times within the same cluster. This motivates us to consider the cross-moment among the marginal residual process ($dN_{ij}(t) - dH(t) - \mathbf{X}_{ij}(t)^T \beta dt$), also denoted as $d\varepsilon_{ij}(t)$. First, suppose $\mathbf{X}_i$. denotes the vector covariate ($\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}$). Then for $j \neq l$,

$$
\begin{aligned}
&E\{d\varepsilon_{ij}(t)d\varepsilon_{il}(s)|Z_{ij} \geq t, Z_{ij} \geq s, \mathbf{X}_i.\} \\
&= \frac{E\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}I(Z_{ij}\geq t)I(Z_{il}\geq s)|\mathbf{X}_i.\right]}{E\{I(Z_{ij}\geq t)I(Z_{il}\geq s)|\mathbf{X}_i.\}} \\
&= \frac{P(C_{ij}\geq t, C_{il}\geq s|\mathbf{X}_i.)E\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}I(T_{ij}\geq t)I(T_{il}\geq s)|\mathbf{X}_i.\right]}{P(C_{ij}\geq t, C_{il}\geq s|\mathbf{X}_i.)P(T_{ij}\geq t, T_{il}\geq s|\mathbf{X}_i.)} \\
&= \frac{E\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}I(T_{ij}\geq t)I(T_{il}\geq s)|\mathbf{X}_i.\right]}{P(T_{ij}\geq t, T_{il}\geq s|\mathbf{X}_i.)}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&E\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}I(T_{ij} \geq t)I(T_{il} \geq s)|\mathbf{X}_i.\right] \\
&= E\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}E\{I(T_{ij} \geq t)|\mathbf{X}_i., \xi_i\}E\{I(T_{il} \geq s)|\mathbf{X}_i., \xi_i\}\right] \\
&= E_{\xi_i}\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}e^{-\xi_i(t+s)}\right] \times \exp\left[-\{\Lambda(t)+\int_0^t \mathbf{X}_{ij}(u)^T \beta du\} - \{\Lambda(s)+\int_0^s \mathbf{X}_{il}(u)^T \beta du\}\right],
\end{aligned}
$$

and

$$P(T_{ij} \geq t, T_{il} \geq s|\mathbf{X}_i.) = E_{\xi_i}\left(\exp\left[-\{\Lambda(t)+\int_0^t \mathbf{X}_{ij}(u)^T \beta du + \xi_i t\} - \{\Lambda(s)+\int_0^s \mathbf{X}_{il}(u)^T \beta du + \xi_i s\}\right]\right).$$

Thus, we have

$$E\{d\varepsilon_{ij}(t)d\varepsilon_{il}(s)|Z_{ij} \geq t, Z_{il} \geq s, \mathbf{X}_i.\} = \frac{E_{\xi_i}\left[\{-dG(t;\theta)+\xi_i dt\}\{-dG(s;\theta)+\xi_i ds\}e^{-\xi_i(t+s)}\right]}{E_{\xi_i}\{e^{-\xi_i(t+s)}\}}.$$

. Since $E(\xi_i e^{-t\xi_i}) = G'(t; \theta) \exp\{-G(t; \theta)\}$ and $E(\xi_i^2 e^{-t\xi_i}) = \{G'(t;\theta)^2 - G''(t;\theta)\}\exp\{-G(t;\theta)\}$, where $G'(t; \theta)$ and $G''(t; \theta)$ denote the first and the second derivative of $G(t; \theta)$, respectively, with respect to t, some algebra yields

$$E\{d\varepsilon_{ij}(t)d\varepsilon_{ij}(s)|Z_{ij} \geq t, Z_{il} \geq s, \mathbf{X}_{i\cdot}\}=dG(t;\theta)dG(s;\theta)-G'(t+s;\theta)dG(t;\theta)ds-G'(t+s;\theta)dG(s;\theta)dt+\{G'(t+s;\theta)^2-G''(t+s;\theta)$$

(6)

After replacing $H(t)$ and $\boldsymbol{\beta}$ with their estimators, we can estimate θ by solving the following estimating equation

$$\sum_{i=1}^{n} \sum_{j\neq l, j,l=1}^{n_i} \int_0^\tau \int_0^\tau Y_{ij}(t)Y_{il}(s)\{d\widehat{\varepsilon}_{ij}(t)d\widehat{\varepsilon}_{il}(s) - Q(t, s;\theta)dtds\}=0,$$

(7)

where $d\hat{\varepsilon}_{ij}(t) = dN_{ij}(t) - d\hat{H}(t) - \mathbf{X}_{ij}(t)^T \hat{\boldsymbol{\beta}}dt$ and

$$Q(t, s;\theta)=G'(t;\theta)G'(s;\theta) - G'(t+s;\theta)\{G'(t;\theta)+G'(s;\theta)\}+G'(t+s;\theta)^2 - G''(t+s;\theta).$$

We denote the solution to (7) as $\hat{\theta}$. Then an estimator for $\Lambda(t)$ is given by $\hat{\Lambda}(t) = \hat{H}(t) - G(t; \hat{\theta})$.

In equation (7), $G(t; \theta)$ and its derivatives can be evaluated for each specific distribution for ξ_i. For example, if ξ_i is assumed to be from a normal distribution with mean zero and variance θ, then $G(t; \theta) = -t^2\theta/2$, $G'(t, \theta) = -t\theta$ and $G''(t; \theta) = -\theta$. Then, with time independent covariates, equation (7) becomes

$$\sum_{i=1}^{n} \sum_{j\neq l, j,l=1}^{n_i} \left[\left\{\Delta_{ij} - \widehat{H}(Z_{ij}) - \mathbf{X}_i^T\widehat{\beta}Z_{ij}\right\}\left\{\Delta_{ij} - \widehat{H}(Z_{ij}) - \mathbf{X}_i^T\widehat{\beta}Z_{ij}\right\} - \left(\theta^2 Z_{ij}^2 Z_{il}^2/4+\theta Z_{ij}Z_{il}\right)\right]=0,$$

which has an explicit solution. If $(\xi_i + \theta)$ is assumed to be from an exponential distribution with mean θ, then $G(t; \theta) = \log(1 + \theta t)$, $G'(t; \theta) = \theta(1 + \theta t)^{-1}$ and $G''(t; \theta) = -\theta^2(1 + \theta t)^{-2}$. Under this case, the solution to equation (7) can be obtained via Newton-Raphson iteration.

In Appendix, we will show that the estimators $\hat{\boldsymbol{\beta}}$, $\hat{H}$ and $\hat{\theta}$ are all asymptotic linear estimators for the respective true parameters. Thus, the asymptotic covariance can be consistently estimated using the empirical covariance of the corresponding estimated influence functions. The details are given in Web Appendix.

2.3 Asymptotic properties

The technical conditions for obtaining the asymptotic properties of the proposed estimators are given in Web Appendix. Let β_0, θ_0, and $\Lambda_0(t)$ denote the true values for β, θ, and $\Lambda_0(t)$, respectively. Under those conditions, we obtain the following results.

Theorem 1—Under conditions (C.1)–(C.4) in Web Appendix and assuming that the true density for ξ is $f(\xi; \theta)$, there exits a local consistent estimator $\hat{\theta}$ solving equation (7). Moreover, $\hat{\beta} \to_p \beta_0$ and $\hat{\Lambda}(t) \to_p \Lambda_0(t)$ uniformly in $[0, \tau]$.

Theorem 2—Under conditions (C.1)–(C.4) in Web Appendix and assuming that the true density for ξ is $f(\xi; \theta)$, $\sqrt{n}(\widehat{\beta} - \beta_0, \widehat{\theta} - \theta_0, \widehat{\Lambda}(t) - \Lambda_0(t))$ converges in distribution to a mean-zero Gaussian process in the metric space $R^{d+1} \times l^{\infty}[0, \tau]$, where d is the dimension of β_0 and $l^{\infty}[0, \tau]$ is the space of bounded functions in $[0, \tau]$ equipped with the uniform bounded norm.

The proofs of Theorem 1 and Theorem 2 are given in Appendix. In addition, we provide a consistent estimator of the asymptotic covariance in Web Appendix.

3. Simulation Studies

In the first simulation study, covariate $\mathbf{X}_{ij}$ contains one Bernoulli random variable and the other random variable is from the uniform distribution in $[0, 1]$. The survival time T_{ij} is generated using model (1), where $\Lambda(t) = 2t^2 + 3t$ and ξ_i follows a standard normal distribution. The true value for β is set to be $(0.5, 1)$ and $\theta_0 = 1$. The censoring time is generated from the uniform distribution in $[0, 3]$, which yields the average censoring rate of 23%. Finally, the cluster size is chosen randomly from 2, 3 and 4. Since in many multi-center studies, the center size can be large, we also consider the cluster size to be as large as 50 while the number of clusters is relatively small.

For each simulated data set, we solve equations (4) and (5) to estimate β and H then solve equation (7) for θ. Since ξ_i follows the normal distribution, the function $Q(t, s; \theta)$ is equal to $\theta^2 t^2 s^2/4 + \theta ts$. Thus, solving equation (7) is equivalent to solving a quadratic equation, which may have two solutions. To ensure θ to be positive, we choose the maximum between 0 and the larger of these two solutions as the estimate for θ. To estimate the asymptotic variance, we first estimate the influence functions as given in Appendix then calculate the empirical variance of these influence functions.

Table 1 summarizes the results from 1,000 replicates for each choice of sample size and cluster size. In the table, column "Est" is the average of the estimates from 1,000 replicates; column "SE" is the sample standard deviations of these estimates; column "ESE" is the average of the estimated standard errors; column "CP" is the coverage probabilities of 95% confidence intervals, which are given using the asymptotic normality. Table 1 reports the results for estimating β and θ. In addition, we also summarize the results for $\Lambda(t)$ for some t values which are chosen to be some quantiles of Z_{ij}. The table shows that the estimators for the regression coefficient β and the cumulative baseline functions $\Lambda(t)$ have negligible bias and the inference appears to be correct with the estimated variance being close to the empirical variance and with proper confidence interval coverage. Moreover, we observe that even if the number of clusters is as small as 20, the estimators for the regression coefficients and the baseline cumulative hazard functions still perform well when the cluster size is large. There seems to be some bias in estimating the frailty parameter, due to the small number of the clusters or small cluster size.

We also conduct a second simulation study, in which the data are generated using the same model except that the cluster-specific frailty ($\xi_i + 3$) is from an exponential distribution with mean $\theta = 3$. In this case, $G(t; \theta) = \log(1 + \theta t)$ and

$$Q(t, s; \theta) = \frac{\theta^2}{(1+\theta t)(1+\theta s)} - \frac{\theta^2}{\{1+\theta(t+s)\}} \left(\frac{1}{1+\theta t} + \frac{1}{1+\theta s} \right) + \frac{2\theta^2}{\{1+\theta(t+s)\}^2}.$$

Then the double integration of $Q(t, s; \theta)$ in equation (7) is equivalent to

$$\int_0^a \int_0^b Q(t, s; \theta) dt ds = \log(1+\theta a)\log(1+\theta b) + 2\log\frac{(1+\theta a)(1+\theta b)}{1+\theta(a+b)}$$
$$- \left\{ z\log(1-e^{-z}) + \frac{z^2}{2} - \sum_{k=1}^{\infty} \frac{e^{-kz}}{k^2} \right\} \Big|_{\log\{1+\theta(b+a)\}-\log(1+\theta a)}^{\log(1+\theta b)}$$
$$- \left\{ z\log(1-e^{-z}) + \frac{z^2}{2} - \sum_{k=1}^{\infty} \frac{e^{-kz}}{k^2} \right\} \Big|_{\log\{1+\theta(b+a)\}-\log(1+\theta b)}^{\log(1+\theta a)}.$$

Thus, we can solve equation (7) for $\hat{\theta}$ via one-dimensional numerical search algorithm. In our simulations, we find that for sample size 200, the proportion of equation (7) having no solution is about 5%; however, this number reduces to 0.5% for sample size 400. Table 2 summarizes the simulation results from 1,000 replicates after excluding these non-convergence cases. The same conclusion as in the previous simulation study can be made.

Finally, in another simulation study, we simulate data using the same setting as in the second simulation study. However, we misspecify the frailty distribution as a normal distribution. The simulation results which are reported in Table 3 indicate that due to the misspecification of the frailty distribution, the estimators for the baseline cumulative hazard function have more than 100% bias and the coverage probabilities are as low as 10%. However, the estimators for β's and also $H(t)$'s are the same as before. This is because the estimating equations for β's and H's are independent of the frailty distribution.

4. Assessing Frailty Distribution

In practice, it is impossible to know what distribution the frailty follows and our numerical experience showed that misspecifying the frailty distribution leads to large bias in the aforementioned approaches. Therefore, an important issue is how to assess the frailty distribution using empirical data.

By examining our approach, we have the following relationship $H(t) = \Lambda(t) + G(t; \theta)$, where $G(t; \theta)$ is the Laplace transformation of ξ's density. Moreover, we find that the estimators for $H(t)$ and β are obtained from equations (4) and (5), which never use the information about the distribution of the frailty. In fact, the proof of Theorem 1 shows that $\hat{H}(t)$ and $\hat{\beta}$ are always consistent estimators of $H_0(t)$ and β_0 respectively, no matter how the frailty distribution is assumed in the models. Hence, if we can obtain a consistent estimator of $\Lambda(t)$ disregarding the frailty distribution and denote it as $\tilde{\Lambda}(t)$, then the difference between $\hat{H}(t)$ and $\tilde{\Lambda}(t)$ should be close to $G(t; \theta)$ when the frailty distribution is correctly specified.

To obtain such an estimator $\tilde{\Lambda}$, we use the following fact:

$$E\{dN_{ij}(t) - \mathbf{X}_{ij}(t)^T \beta dt | Y_{ij}(t), \mathbf{X}_{ij}, \xi_i\} = Y_{ij}(t)\{d\Lambda(t) + \xi_i(dt)\}, \quad j=1,\ldots,n_i.$$

Thus, using the observations from cluster i and substituting $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, we can estimate $d\Lambda(t)+$ $\xi_i dt$ by $\sum_{j=1}^{n_i} Y_{ij}(t)\{dN_{ij}(t) - \mathbf{X}_{ij}(t)^T\widehat{\beta}dt\}/\sum_{j=1}^{n_i} Y_{ij}(t)$. After taking the average over all the clusters, we have

$$d\Lambda(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{n_i} Y_{ij}(t)\{dN_{ij}(t) - \mathbf{X}_{ij}(t)^T\widehat{\beta}dt\}}{\sum_{j=1}^{n_i} Y_{ij}(t)} - \frac{1}{n}\sum_{i=1}^{n}\xi_i dt.$$

Since $n^{-1}\sum_{i=1}^{n}\xi_i$ converges to zero, we thus drop the second term on the right-hand side and obtain an estimator for $\Lambda(t)$ as

$$\tilde{\Lambda}(t) = \int_0^t \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{n_i} Y_{ij}(u)\{dN_{ij}(u) - \mathbf{X}_{ij}(u)^T\widehat{\beta}du\}}{\sum_{j=1}^{n_i} Y_{ij}(u)}.$$

Hence, we compare $\hat{D}(t) \equiv \{\hat{H}(t) - \tilde{\Lambda}(t)\}$ with $G(t; \hat{\theta})$ to examine the validity of assuming ξ's density as $f(\xi; \theta)$. Additionally, we can obtain a confidence interval of $\{\hat{D}(t) - G(t; \hat{\theta})\}$ using the following result.

Theorem 3

Under conditions (C.1)–(C.4) in Web Appendix, $\sup_{t\in[0,\tau]} |\tilde{\Lambda}(t) - \Lambda_0(t)| \to_p 0$. Furthermore, if assume that ξ_i's density is $f(\xi; \theta_0)$, $\sqrt{n}\{\hat{D}(t) - G(t;\widehat{\theta})\}$ converges in distribution to a zero-mean Gaussian process in $l^\in [0, \tau]$.

In the proof of Theorem 3, we obtain $\sqrt{n}\{\hat{D}(t) - G(t;\widehat{\theta})\} = n^{-1/2}\sum_{i=1}^{n} W_i(t) + o_P(1)$, where

$$W_i(t) = S_{H,i}(t) - \int_0^t \frac{\sum_{j=1}^{n_i} Y_{ij}(u)\{dN_{ij}(u) - d\Lambda_0(u) - \mathbf{X}_{ij}(u)^T\beta_0 du\}}{\sum_{j=1}^{n_i} Y_{ij}(u)} + E\left\{\int_0^t \frac{\sum_{j=1}^{n_i} Y_{ij}(u)\mathbf{X}_{ij}(u)^T du}{\sum_{j=1}^{n_i} Y_{ij}(u)}\right\} S_{\beta,i} - \frac{\partial}{\partial\theta}G(t;\theta_0)S_{\theta,i}$$

with $S_{H,i}(t)$, $S_{\beta,i}$ and $S_{\theta,i}$ being the respective influence functions for $\hat{H}(t)$, $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ from the ith cluster observations. To construct a uniform confidence band for the true function of $\{D(t) - G(t; \hat{\theta})\}$, we use a resampling method: we generate $\omega_1, \ldots, \omega_n$ independently from the standard normal distribution and calculate $\sup_{t\in[0,\tau]}|\widehat{D}(t) - G(t;\widehat{\theta}) + n^{-1}\sum_{i=1}^{n}\omega_i\widehat{W}_i(t)|$, where $\hat{W}_i(t)$ has the same expression as $W_i(t)$ except that we replace Λ_0 and $\boldsymbol{\beta}_0$ by $\tilde{\Lambda}$ and $\hat{\boldsymbol{\beta}}$ and use consistent estimators for $S_{\beta,i}$ and $S_{\theta,i}$. We repeat this resampling procedure many times and determine the 95%-percentile of the calculated values, denoted by $c_{0.95}$. Then a uniform 95% confidence band is given by $\{\hat{D}(t) - G(t; \hat{\theta}) - c_{0.95}, \hat{D}(t) - G(t; \hat{\theta}) + c_{0.95}\}$, $t \in [0, \tau]$. If the band contains the horizontal line at zero, then there is no significant evidence to reject the assumed frailty distribution.

Finally, we conduct simulation studies to examine the performance of this assessment approach. The setting is similar to Section 4 but we only look at the case with large cluster size of 25. We simulate the frailty ξ from either normal or exponential distribution but we estimate all the parameters treating ξ as from a normal distribution. This way, we can

examine both the type I error and power of the proposed method. The results show that under the nominal level of 0.05, if using the correct frailty distribution, the type I error for the test statistic $\int_0^{t_0} |\widehat{D}(t) - G(t;\widehat{\theta})| dt$, where t_0 is the median time of the observed events, is 0.067 for both $n = 20$ and $n = 50$; while if using the incorrect frailty distribution, the power becomes 0.214 for $n = 20$ and increases to 0.520 for $n = 50$. Our additional numerical experience indicates that the proposed procedure may not work well if the cluster size is as small as 5. This is because in the estimation of each cluster-specific hazards function may $\tilde{\Lambda}$ be very inaccurate with small clusters. However, the above simulation study shows that the procedure is valid when the cluster size is at least 25.

5. Applications

5.1 Diabetic study

We now apply the proposed model to analyze the well-known Diabetic Retinopathy Study (Huster et al. 1989), which was conducted to assess the effectiveness of laser photocoagulation in delaying visual loss among patients with diabetic retinopathy. One eye of each patient was randomly selected to receive the laser treatment while the other eye was used as a control. The failure time of interest is the time to visual loss as measured by visual acuity less than 5/200. We confine our analysis to a subset of 197 high-risk patients, and consider three covariates: X_{1ij} indicates, by the values 1 versus 0, whether or not the jth eye ($j = 1$ for the left eye and $j = 2$ for the right eye) of the ith patient was treated with laser photocoagulation, $X_{2i1} = X_{2i2}$ indicates, by the values 1 versus 0, whether the ith patient had adult-onset or juvenile-onset diabetics, and $X_{3ij} = X_{1ij} * X_{2ij}$.

To justify why the covariates may have additive effects, we plot the differences among the cumulative hazards estimates for each combination of (X_{1ij}, X_{2ij}) (a total of four combinations), where the estimates are the Breslow-estimates of the cumulative hazards function within each group. The plot as shown in Figure 1 indicates that the differences are clearly linear in time, which imply that additive effects of the covariates may be plausible. Thus, we fit model (1) with these three covariates, along with ξ_i to account for the correlation between the two eyes of the same patient. We consider fitting the model with ξ_i from the normal distribution or the exponential distribution. The results are given in Table 4. They show that there is a high degree of dependence between the failure times of the two eyes from the same patient. Both the treatment indicator and the interaction term are significant, whereas the diabetic type is not. These findings agree with the results fitting the gamma-frailty proportional hazards model or proportional odds model (Zeng, Lin and Yin, 2005).

Another nice feature for our model is to make individual prediction. For example, we can predict the conditional survival probabilities of the treated eye given that it has not failed before 30 months while the untreated eye failed between 24 and 30 months, i.e., $\Pr(T_2 > t | T_2 \geq 30, 24 < T_1 < 30, X_{11} = 0, X_{12} = 1, X_2)$ for $t > 30$, where T_2 is the failure time for the treated eye and T_1 is the failure time for the untreated eye, X_{1k} is the treatment status for the kth eye ($k=1, 2$), and X_2 is the diabetic type for this patient. It is straightforward to show that

$$\Pr(T_2 > t | T_2 \geq 30, 24 < T_1 < 30, \mathbf{X}_1, \mathbf{X}_2) = \frac{\int_u e^{-\Lambda(t) - t\mathbf{X}_2^T \beta - ut} \{ e^{-\Lambda(24) - 24\mathbf{X}_1^T \beta - 24u} - e^{-\Lambda(30) - 30\mathbf{X}_1^T \beta - 30u} \} f(u;\theta) du}{\int_u e^{-\Lambda(30) - 30\mathbf{X}_2^T \beta - 30u} \{ e^{-\Lambda(24) - 24\mathbf{X}_1^T \beta - 24u} - e^{-\Lambda(30) - 30\mathbf{X}_1^T \beta - 30u} \} f(u;\theta) du}.$$

We can estimate this probability function by replacing $\boldsymbol{\beta}$, θ and Λ by their respective estimates. The variance function is given by the Delta method. Figure 2 displays the

estimated survival curves along with the 95% confidence intervals for the two diabetic types using model (1) with the normally distributed frailty, where the solid curve is the survival function for the patient with the juvenile-onset type, the dot-dash curve is for the patient with the adult-onset type, both the dashed and dotted curves are the confidence bands.

5.2 SOLVD study

We also apply our method to analyze the SOLVD (SOLVD Investigators, 1990) Treatment Trial data. This study was a randomized, double-masked, placebo-controlled trial conducted between 1986 and 1991. The participants were of age 21 to 80 years old, inclusive, with overt symptoms of congestive heart failure and left ventricular ejection fraction less than 35%. The latter is a measure of the efficiency of the heart in ejecting blood and is a number between 0 and 100%. The study was done at 23 medical centers in US, Canada and Belgium and the average number of patients per center was slightly over 100. The event of interest was the number of years to the first hospitalization for congestive heart failure or death from all causes, whichever happened first. The goal was to examine the effect of treatment by enalapril versus placebo but the participants' age, gender and ejection fraction could be potential confounders so they should also be adjusted for in the analysis. We fit our proposed model for this data, assuming that the frailty follows a normal distribution. Because of the large size in each cluster, the test process in Section 5 can be used to assess the goodness-of-fit of the assumed frailty model. It shows that there exists little evidence to reject the normal frailty (p-value is 0.33). The results are summarized in Table 5.

The results indicate that the treatment had a significant effect in reducing the risk of first hospitalization for congestive heart failure or death from all causes; younger patients had lower risk; patients with higher ejection fraction had lower risk as well. There is no strong evidence that significant difference exists between the two gender groups. The non-significance of the frailty parameter implies that the survival behaviors among all these centers were similar to one another. In fact, we plot the estimated cumulative hazards functions from all 23 centers in Figure 3 and they have very small differences except that one particulary medical center had a relatively larger baseline hazards. In Figure 4, we plot the proposed test process (the bold black curves) and the randomly simulated 100 curves from the null distribution. The figure indicates that using the normal frailty distribution may be appropriate.

Finally, we exclude that particular center identified from Figure 3 and re-do the analysis. The p-value for assessing the normality frailty assumption is around 0.30. The fitted results are also given in Table 5 and the conclusion are similar to before. For comparison, we also report the results by fitting the data using multiplicative gamma frailty models. The results show that the direction and significance of the estimates for the covariate effects are similar, although the interpretation of the coefficients is very different.

6. Discussion

We have proposed an additive model with random effects for clustered survival times. The proposed model takes a similar expression as one would fit longitudinal data via mixed effect models. Thus, the parameters in our model enjoy similar nice interpretations as in the usual mixed effect models, but in terms of hazard risks. The inference approach we have proposed can be treated as a generalization of the generalized estimating equations to the hazards models. As the result, any working dependence matrices can be used as weights in equations (4) and (5) to construct consistent estimators of regression parameters. One natural question is how to obtain the most efficient weight matrices; however, although in the longitudinal data analysis, the true dependence covariance matrices should give most

efficient estimation, it is unclear how the covariance matrices are even defined under survival context.

One limitation of the additive structure is that the estimated hazards rate function may not always be positive so the derived survival functions, $\tilde{S}(t)$, are not necessarily non-increasing. With an additional random effect introduced in our proposed model, the non-positive hazards problem may be lessened or worsened depending on the monotonicity of $G(t; \theta)$. For example, in a normal frailty case, $H(t) = \Lambda(t) - \theta t^2/2$, where $H(t)$ is the baseline after integrating out the random effect. Then even though $H(t)$ may be estimated to have negative jumps at some t, the estimator for $\Lambda(t)$ may not. In the exponential case, $H(t) = \Lambda(t) + \log(1 + \theta t)/\theta$. Then $\Lambda(t)$ can have more negative jumps but this also depends on the scale of θ. To fix the non-monotonicity problem in the estimated survival function, we can make a minor modification and define a new estimator as $\inf_{s \leq t} \tilde{S}(s)$. Theoretically, if $\tilde{S}(t)$ is a consistent estimator, the latter should also be consistent due to the increasing property of the truth. This type of modification has been employed in the literature for additive rate models (Lin and Yin, 1994; Yin and Cai, 2004).

In comparison to the multiplicative frailty model, the proposed model can be viewed as a parallel generalization of the additive hazards model to clustered data. Thus, when data (as shown in the diabetic study and seen in Figure 1) really present additive structure and cluster heterogeneity, the proposed model is more appropriate as the multiplicative assumption for the multiplicative model could be violated and may not fit the data well. Our proposed model provides information on hazard differences instead of hazard ratios. Additionally, the frailty variance represents the variability of the baseline hazards rates across clusters and therefore provide us information on the heterogenity of the hazard rate across clusters.

With the specified random effect distribution, the observed likelihood function can be expressed in terms of all the parameters. Therefore, some possible maximum likelihood approaches can be applicable to obtain the efficient parameter estimators. Unfortunately, due to the additive model structure as well as the semiparametric model structure, such inference and computation will be more complicated than the multiplicative frailty model. Further research is being investigated by us.

In model (1), we can further allow for heterogeneity of mixed effects among subjects. This can be achieved by including the interaction between random effects and some covariates. The proposed estimating equations should be applicable but in more complex forms. Finally, as a counterpart to the usual gamma-frailty multiplicative model, developing a valid test to discriminate our model from the gamma-frailty model is an important problem in practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aalen OO. A linear regression model for the analysis of lifetimes. Statistics in Medicine. 1989; 8:907–925. [PubMed: 2678347]

Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, D. Statistical Models Based on Counting Processes. New York: Springer-Verlag; 1993.

Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society Ser B. 1972; 34:187–220.

Gandy A, Jensen U. Checking a semiparametric additive risk model. Lifetime Data Analysis. 2005a; 11:451–472. [PubMed: 16328571]

Gandy A, Jensen U. On goodness-of-fit tests for Aalen's additive risk model. Scandinavian Journal of Statistics. 2005b; 32:425–445.

Huffer FW, McKeague IW. Weighted least squares estimation for Aalen's additive risk model. Journal of the American Statistical Association. 1991; 86:114–129.

Huster WJ, Brookmeyer R, Self SG. Modelling Paired Survival Data With Covariates. Biometrics. 1989; 45:145–156. [PubMed: 2655727]

Lin DY, Ying ZL. Semiparametric analysis of the additive risk model. Biometrika. 1994; 81:61–71.

McKeague IW, Sasieni PD. A partly parametric additive risk model. Biometrika. 1994; 81:501–514.

Murphy SA. Consistency in a proportional hazards model incorporating a random effect. The Annals of Statistics. 1994; 22:712–731.

Murphy SA. Asymptotic theory for the frailty model. The Annals of Statistics. 1995; 23:182–198.

Parner E. Asymptotic theory for the correlated gamma-frailty model. The Annals of Statistics. 1998; 26:183–214.

Shen Y, Cheng SC. Confidence bands for cumulative incidence curves under the additive risk model. Biometrics. 1999; 55:1093–1100. [PubMed: 11315053]

SOLVD Investigators. Studies of left ventricular dysfunction (SOLVD)– Rationale, design and methods: Two trials that evaluate the effect of enalapril in patients with reduced ejection fraction. The American Journal of Cardiology. 1990; 66:315–322. [PubMed: 2195865]

Yin G. Model checking for additive hazards model with multivariate survival data. Journal of Multivariate Analysis. 2007; 98:1018–1032.

Yin G, Cai J. Additive hazards model with multivariate failure time data. Biometrika. 2004; 91:801–818.

Zeng D, Lin DY, Yin G. Maximum likelihood estimation in proportional odds model with random effects. Journal of the American Statistical Association. 2005; 100:470–483.

Appendix

Proofs of Theorems 1–3

Proof of Theorem 1

We first show the consistency of the estimators. Clearly, since $\{Y_{kl}(t): t \in [0, \tau]\}$ is a P-Glivenko-Cantelli class, we obtain that uniformly in $t \in [0, \tau]$,

$$n^{-1}\sum_{k=1}^{n}\sum_{l=1}^{n_k}Y_{kl}(t)\mathbf{X}_{kl}(t)\to_{a.s.}E\left\{\sum_{l=1}^{n_k}Y_{kl}(t)\mathbf{X}_{kl}(t)\right\}, \quad n^{-1}\sum_{k=1}^{n}\sum_{l=1}^{n_k}Y_{kl}(t)\to_{a.s.}E\left\{\sum_{l=1}^{n_k}Y_{kl}(t)\right\}$$

and the second limit is strictly positive from condition (C.2) in Web Appendix. Thus,

$$\overline{\mathbf{X}}(t)\to_{a.s.}\frac{E\{\sum_{l=1}^{n_k}Y_{kl}(t)\mathbf{X}_{kl}(t)\}}{E\{\sum_{l=1}^{n_k}Y_{kl}(t)\}}\equiv\mu(t)$$

uniformly in $t \in [0, \tau]$. As a result,

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\overline{\mathbf{X}}(t)\}^{\otimes 2}dt \to_{a.s} E\left[\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\mu(t)\}^{\otimes 2}dt\right] \equiv \sum.$$

In Web Appendix, we show Σ is invertible. Therefore, from the expression of $\hat{\beta}$ and equation (3), we have

$$\widehat{\beta} \to_{a.s.} \sum{}^{-1} E\left[\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\mu(t)\}dN_{ij}(t)\right],$$

which is also equal to

$$\sum{}^{-1} E\left[\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\mu(t)\}\{dH_0(t)+\mathbf{X}_{ij}(t)^T\beta_0 dt\}\right]=\beta_0.$$

Similarly, for fixed $s \in [0, \tau]$,

$$\widehat{H}(s)=\int_0^s \frac{\sum_{i=1}^n\sum_{j=1}^{n_i}Y_{ij}(t)\{dN_{ij}(t)-\mathbf{X}_{ij}(t)^T\beta_0 dt\}}{\sum_{i=1}^n\sum_{j=1}^{n_i}Y_{ij}(t)}+o(1)\to_{a.s}\int_0^s \frac{E\left[\sum_{j=1}^{n_i}Y_{ij}\{dN_{ij}(t)-\mathbf{X}_{ij}(t)^T\beta_0\}dt\right]}{E\{\sum_{j=1}^{n_i}Y_{ij}(t)\}},$$

which is equal to $H_0(s)$. Since $H_0(s)$ is continuous in $[0, \tau]$, the pointwise consistency can be strengthened to the uniform convergence in $[0, \tau]$.

The consistency of $\hat{\theta}$ is based on the estimating equation for $\hat{\theta}$ as given in (7), i.e.,

$$n^{-1}\sum_{i=1}^{n}\sum_{j\neq l,j,l=1}^{n_i}\int_0^\tau\int_0^\tau Y_{ij}(t)Y_{il}(s)\{d\widehat{\varepsilon}_{ij}(t)d\widehat{\varepsilon}_{il}(s)-Q(t,s;\theta)dtds\}=0.$$

From the consistency of $(\hat{\beta}, \hat{H})$, it is clear that the left-hand side of the above equation converges to

$$E\left[\sum_{j\neq l,j,l=1}^{n_i}\int_0^\tau\int_0^\tau Y_{ij}(t)Y_{il}(s)\{d\varepsilon_{ij0}(t)d\varepsilon_{il0}(s)-Q(t,s;\theta)dtds\}\right],$$

where $d\varepsilon_{ij0}(t) = dN_{ij}(t) - dH_0(t) - \mathbf{X}_{ij}(t)^T\beta_0 dt$. The above limit, by the derivation in Section 2, is also equal to

$$E\left[\sum_{j\neq l,j,l=1}^{n_i}\int_0^\tau\int_0^\tau Y_{ij}(t)Y_{il}(s)\{Q(t,s;\theta_0)-Q(t,s;\theta)\}dtds\right].$$

Moreover, this convergence is uniformly in for in a compact set. By condition (C.4) in Web Appendix,

$$E\left[\sum_{j\neq l,j,l=1}^{n_i}\int_0^\tau\int_0^\tau Y_{ij}(t)Y_{il}(s)\nabla_\theta Q(t,s;\theta)dtds\right]$$

is non-singular in a neighborhood of θ_0. Thus, from the inverse mapping theorem, for any small ε, there exists a unique solution $\hat\theta$ to equation (7) such that $|\hat\theta-\theta_0|<\varepsilon$. This proves the consistency of $\hat\theta$.

Proof of Theorem 2

We use $\mathbf{P}_n$ to denote the empirical measure from n i.i.d clustered data and use $\mathbf{P}$ to denote the true probability measure. From the expression of $\hat{\boldsymbol\beta}$, we obtain

$$\hat\beta-\beta_0=\left[n^{-1}\sum_{i=1}^n\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\overline{\mathbf{X}}(t)\}^{\otimes 2}dt\right]^{-1}\times\mathbf{P}_n\left[\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\overline{\mathbf{X}}(t)\}d\varepsilon_{ij0}(t)\right].$$

Since $\mathbf{P}\left[\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\overline{\mathbf{X}}(t)\}d\varepsilon_{ij0}(t)\right]=0$ and $\overline{\mathbf{X}}(t)$ uniformly converges to $\boldsymbol\mu(t)$ and it belongs to some Donsker class, we have

$$\hat\beta-\beta_0=\{\textstyle\sum+o_p(1)\}^{-1}(\mathbf{P}_n-\mathbf{P})\left[\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\mu(t)\}d\varepsilon_{ij0}(t)\right]+o_p(n^{-1/2}).$$

$$\text{(A.1)}$$

That is, $\hat{\boldsymbol\beta}$ is an asymptotic linear estimator for $\boldsymbol\beta_0$ with the influence function

$$\mathbf{S}_\beta(\mathbf{O}_i)=\sum{}^{-1}\sum_{j=1}^{n_i}\int_0^\tau Y_{ij}(t)\{\mathbf{X}_{ij}(t)-\mu(t)\}d\varepsilon_{ij0}(t),$$

where $\mathbf{O}_i$ denotes the observed data from cluster i. Similarly, using the expression of $\hat{H}$, we obtain

$$\hat{H}(s)-H_0(s)=\int_0^s\frac{\mathbf{P}_n\left\{\sum_{j=1}^{n_i}d\varepsilon_{ij0}(t)\right\}}{E\left\{\sum_{j=1}^{n_i}Y_{ij}(t)\right\}+o_p(1)}-\int_0^s\overline{\mathbf{X}}(t)^T(\hat\beta-\beta_0)dt.$$

Thus, from (A.1),

$$\widehat{H}(s) - H_0(s) = (\mathbf{P_n} - \mathbf{P})\left[\int_0^s \frac{\sum_{j=1}^{n_i} Y_{ij}(t)d\varepsilon_{ij0}(t)}{E\left\{\sum_{j=1}^{n_i} Y_{ij}(t)\right\}} - \int_0^s \mu(t)^T dt \mathbf{S}_\beta(\mathbf{O}_i)\right] + o_p(n^{-1/2}),$$

(A.2)

Additionally, (A.2) holds in the metric space $l^\infty[0, \tau]$. Equivalently, $\hat{H}(s)$ is an asymptotic linear estimator for $H_0(s)$ with influence function

$$S_H(\mathbf{O}_i; s) = \int_0^s \frac{\sum_{j=1}^{n_i} d\varepsilon_{ij0}(t)}{E\left\{\sum_{j=1}^{n_i} Y_{ij}(t)\right\}} - \int_0^s \mu(t)^T dt \mathbf{S}_\beta(\mathbf{O}_i).$$

Finally, a similar expansion applies to the left-hand side of (7) and yields

$$\widehat{\theta} - \theta_0 = (\mathbf{P}_n - \mathbf{P})S_\theta(\mathbf{O}_i) + o_p(n^{-1/2})$$

(A.3)

for some influence function S_θ. The detail of S_θ is given in Web Appendix.

Proof of Theorem 3

According to the expression of $\tilde{\Lambda}(t)$, we obtain

$$d\tilde{\Lambda}(t) = d\Lambda_0(t) + \frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^{n_i} Y_{ij}(t)dM_{ij}(t)}{\sum_{j=1}^{n_i} Y_{ij}(t)} - \frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^{n_i} Y_{ij}(t)\mathbf{X}_{ij}(t)^T}{\sum_{j=1}^{n_i} Y_{ij}(t)}(\widehat{\beta} - \beta_0) + \frac{1}{n}\sum_{i=1}^n \xi_i dt,$$

(A.4)

where $dM_{ij}(t) = dN_{ij}(t) - d\Lambda_0(t) - \mathbf{X}_{ij}(t)^T\beta_0 dt - \xi_i dt$. Clearly, $E\{\sum_{j=1}^{n_i} Y_{ij}(t)dM_{ij}(t)/\sum_{j=1}^{n_i} Y_{ij}(t)\} = 0$ and $\{dM_{ij}(t) = t \in [0, \tau]\}$ and $\{Y_{ij}(t): t \in [0, \tau]\}$ are the Glivenko-Cantelli class. Thus, the second term on the right-hand side converges uniformly to zero. Finally, since $\widehat{\beta}$ converges to β_0 and $n^{-1}\sum_{i=1}^n \xi_i$ converges to zero, we obtain the uniformly convergence of $\tilde{\Lambda}(t)$ to $\Lambda_0(t)$.

The above expansion also implies

$$\sqrt{n}\{\tilde{\Lambda}(t) - \Lambda_0(t)\} = n^{-1/2}(\mathbf{P}_n - \mathbf{P})\left[\int_0^t \frac{\sum_{j=1}^{n_i} Y_{ij}(u)\{dM_{ij}(u) + \xi_i du\}}{\sum_{j=1}^{n_i} Y_{ij}(u)}\right] - \left[\int_0^t E\left\{\frac{\sum_{j=1}^{n_i} Y_{ij}(u)\mathbf{X}_{ij}(u)^T}{\sum_{j=1}^{n_i} Y_{ij}(u)}\right\} du + o_p(1)\right]\sqrt{n}(\widehat{\beta} - \beta_0) + o_p(1).$$

Thus, combining with the expansions in (A.1)–(A.3), we immediately have

$$\sqrt{n}\{\widehat{D}(t) - G(t, \widehat{\theta})\} = n^{-1/2}\sum_{i=1}^n W_i(t) + o_p(1),$$

where

$$W_i(t) = S_{H,i}(t) - \int_0^t \frac{\sum_{j=1}^{n_i} Y_{ij}(u)\{dN_{ij}(u) - d\Lambda_0(u) - \mathbf{X}_{ij}(u)^T \beta_0 du\}}{\sum_{j=1}^{n_i} Y_{ij}(u)} + E\left\{\int_0^t \frac{\sum_{j=1}^{n_i} Y_{ij}(u)\mathbf{X}_{ij}(u)^T du}{\sum_{j=1}^{n_i} Y_{ij}(u)}\right\} \mathbf{S}_{\beta,i} - \frac{\partial}{\partial \theta}G(t;\theta_0)S_{\theta,i}$$

with $S_{H,i}(t)$, $\mathbf{S}_{\beta,i}$ and $S_{\theta,\,i}$ being the respective influence functions for $\hat{H}(t)$, $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ from the observations in the ith cluster, as given in the expansions (A.1)–(A.3).
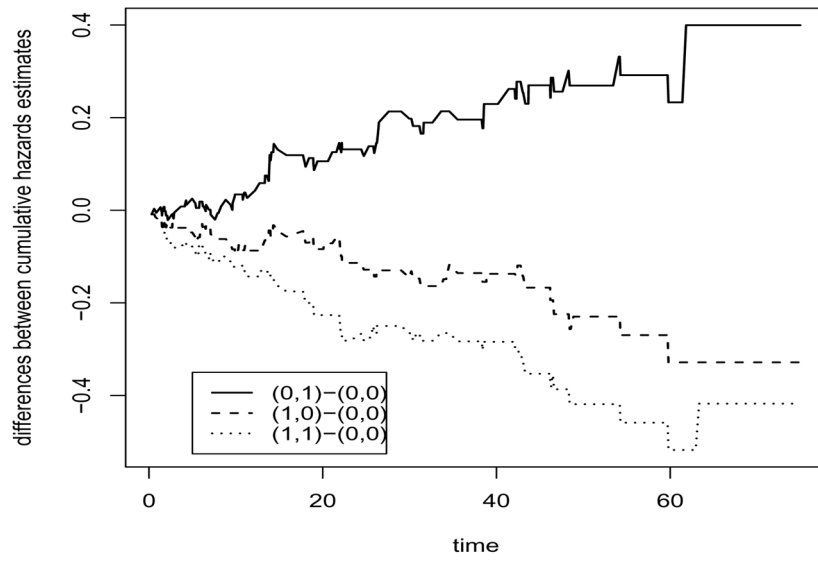
Figure 1.
Differences among the estimated cumulative hazard functions versus time in the DRS study
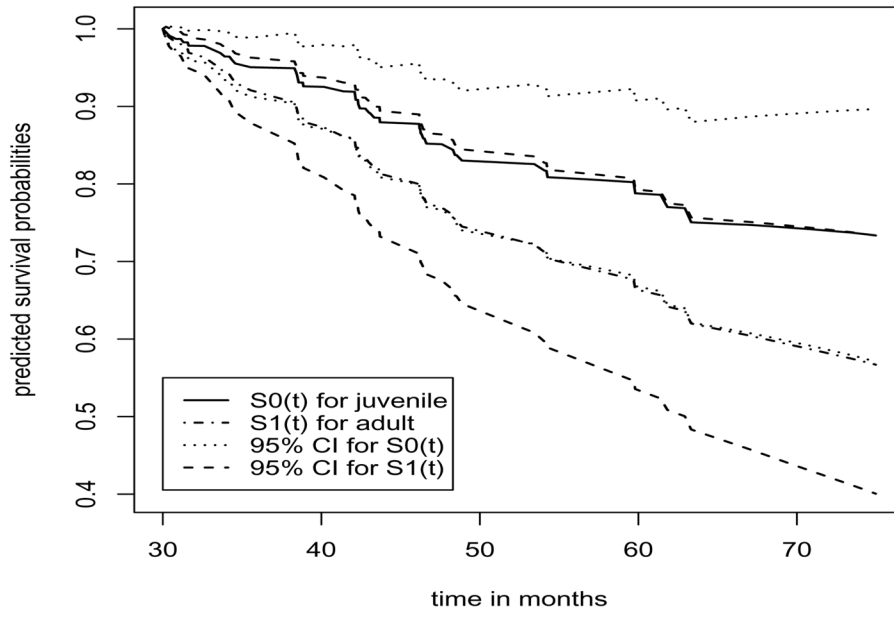
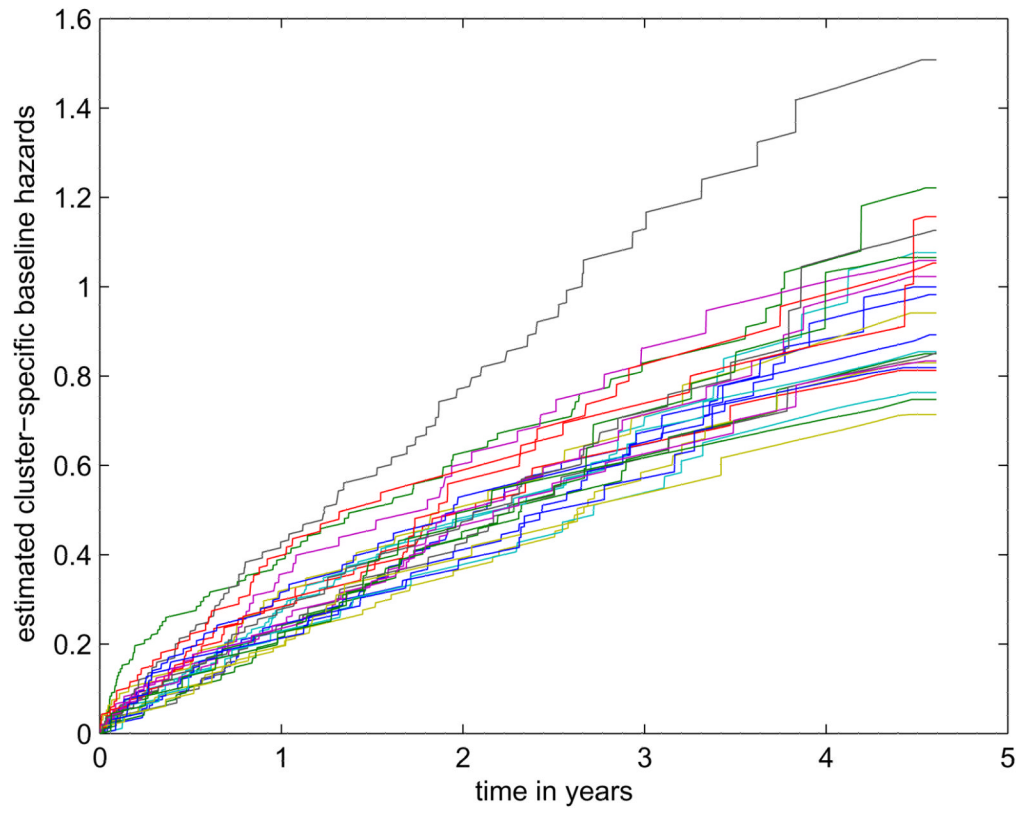**Figure 2.**
Predicted survival functions in the DRS study

Figure 3.
Estimated baseline hazards functions from all medical centers in the SOLVD study
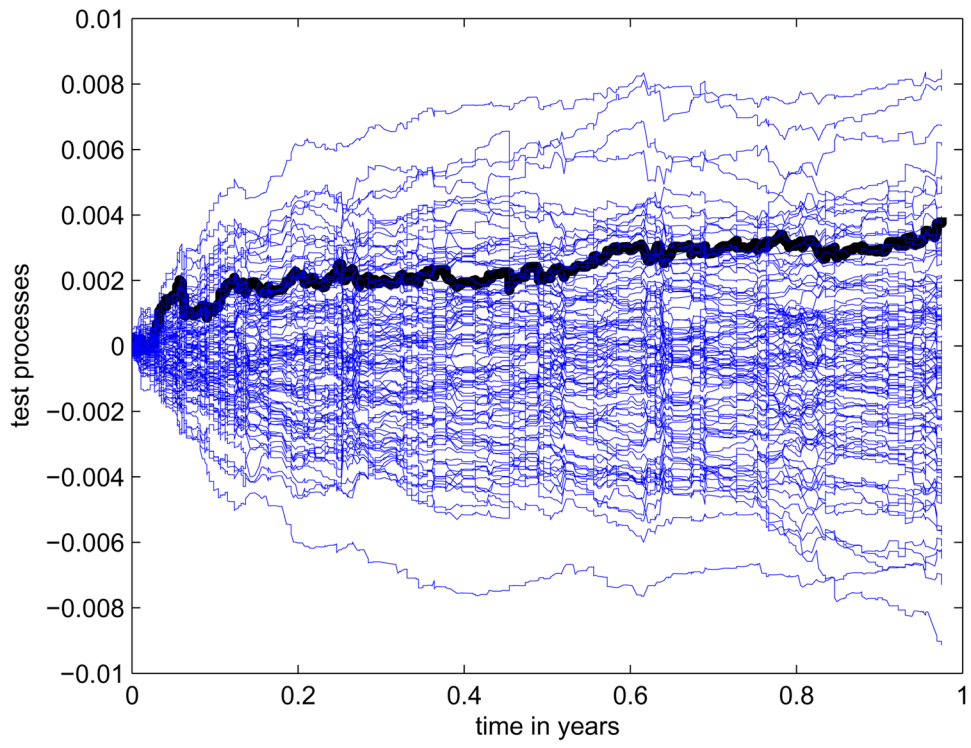
Figure 4.
Plot of the test process and simulated curves under the null in the SOLVD study

Table 1

Results from the simulation study with normal frailty

n	Cluster Size	Parameter	True value	Est	SE	ESE	CP
100	2 ~ 5	β_1	0.5	0.506	0.631	0.613	0.94
		β_2	1.0	1.015	1.091	1.059	0.95
		θ	1.0	1.043	1.096	1.424	0.97
		$\Lambda(t_1)$	0.473	0.477	0.103	0.103	0.95
		$\Lambda(t_2)$	1.076	1.094	0.213	0.215	0.94
		$\Lambda(t_3)$	1.958	2.000	0.377	0.397	0.96
200	2 ~ 5	β_1	0.5	0.504	0.445	0.433	0.94
		β_2	1.0	1.018	0.769	0.747	0.94
		θ	1.0	0.976	0.857	1.044	0.98
		$\Lambda(t_1)$	0.473	0.473	0.075	0.073	0.93
		$\Lambda(t_2)$	1.076	1.082	0.158	0.153	0.94
		$\Lambda(t_3)$	1.958	1.964	0.281	0.281	0.95
400	2 ~ 5	β_1	0.5	0.490	0.310	0.306	0.95
		β_2	1.0	1.001	0.549	0.528	0.94
		θ	1.0	0.921	0.655	0.761	0.98
		$\Lambda(t_1)$	0.473	0.474	0.054	0.052	0.94
		$\Lambda(t_2)$	1.076	1.079	0.111	0.108	0.94
		$\Lambda(t_3)$	1.958	1.959	0.200	0.200	0.94
20	25	β_1	0.5	0.509	0.430	0.416	0.94
		β_2	1.0	1.015	0.757	0.714	0.92
		θ	1.0	0.856	0.491	0.449	0.86
		$\Lambda(t_1)$	0.473	0.470	0.078	0.076	0.93
		$\Lambda(t_2)$	1.076	1.078	0.162	0.155	0.93
		$\Lambda(t_3)$	1.958	1.952	0.284	0.269	0.93
20	50	β_1	0.5	0.502	0.314	0.292	0.93
		β_2	1.0	0.992	0.554	0.503	0.91
		θ	1.0	0.817	0.404	0.328	0.79

n	Cluster Size	Parameter	True value	Est	SE	ESE	CP
		$\Lambda(t_1)$	0.473	0.476	0.064	0.058	0.91
		$\Lambda(t_2)$	1.076	1.080	0.128	0.117	0.92
		$\Lambda(t_3)$	1.958	1.953	0.220	0.204	0.92

Table 2

Results from the simulation study with exponential frailty

n	Cluster Size	Parameter	True value	Est	SE	ESE	CP
100	2 ~ 5	β_1	0.5	0.526	0.658	0.643	0.95
		β_2	1.0	1.002	1.107	1.104	0.95
		θ	3.0	2.889	1.192	1.264	0.99
		$\Lambda(t_1)$	0.111	0.116	0.069	0.077	0.98
		$\Lambda(t_2)$	0.298	0.307	0.157	0.173	0.98
		$\Lambda(t_3)$	0.549	0.579	0.259	0.281	0.98
200	2 ~ 5	β_1	0.5	0.501	0.462	0.455	0.95
		β_2	1.0	1.015	0.796	0.779	0.95
		θ	3.0	2.913	0.939	0.914	0.97
		$\Lambda(t_1)$	0.111	0.118	0.052	0.056	0.97
		$\Lambda(t_2)$	0.298	0.315	0.117	0.125	0.98
		$\Lambda(t_3)$	0.549	0.577	0.188	0.202	0.97
400	2 ~ 5	β_1	0.5	0.484	0.329	0.320	0.95
		β_2	1.0	1.016	0.568	0.552	0.95
		θ	3.0	2.929	0.660	0.653	0.95
		$\Lambda(t_1)$	0.111	0.114	0.040	0.040	0.96
		$\Lambda(t_2)$	0.298	0.308	0.088	0.089	0.96
		$\Lambda(t_3)$	0.549	0.564	0.140	0.143	0.96
20	25	β_1	0.5	0.534	0.514	0481	0.94
		β_2	1.0	0.980	0.867	0.810	0.93
		θ	3.0	2.805	0.865	0.859	0.91
		$\Lambda(t_1)$	0.111	0.118	0.045	0.051	0.95
		$\Lambda(t_2)$	0.298	0.316	0.107	0.117	0.96
		$\Lambda(t_3)$	0.549	0.569	0.191	0.197	0.95
20	50	β_1	0.5	0.489	0.344	0.334	0.93
		β_2	1.0	1.011	0.625	0.576	0.91
		θ	3.0	2.826	0.772	0.790	0.90

n	Cluster Size	Parameter	True value	Est	SE	ESE	CP
		$\Lambda(t_1)$	0.111	0.118	0.033	0.042	0.97
		$\Lambda(t_2)$	0.298	0.317	0.080	0.096	0.97
		$\Lambda(t_3)$	0.549	0.586	0.140	0.161	0.97

Table 3

Results from the simulation study with misspecified frailty

n	Cluster Size	Parameter	True value	Est	SE	ESE	CP
100	2 ~ 5	β_1	0.5	0.521	0.659	0.645	0.95
		β_2	1.0	1.014	1.125	1.112	0.95
		θ	3.0	2.559	1.390	1.345	0.93
		$\Lambda(t_1)$	0.111	0.272	0.050	0.051	0.09
		$\Lambda(t_2)$	0.298	0.695	0.117	0.117	0.05
		$\Lambda(t_3)$	0.549	1.251	0.214	0.213	0.05
200	2 ~ 5	β_1	0.5	0.510	0.462	0.454	0.95
		β_2	1.0	1.004	0.796	0.782	0.95
		θ	3.0	2.685	1.034	0.966	0.92
		$\Lambda(t_1)$	0.111	0.271	0.035	0.036	0.00
		$\Lambda(t_2)$	0.298	0.699	0.082	0.083	0.00
		$\Lambda(t_3)$	0.549	1.255	0.149	0.150	0.00
20	25	β_1	0.5	0.524	0.514	0.485	0.94
		β_2	1.0	1.004	0.867	0.824	0.93
		θ	3.0	2.680	1.041	0.985	0.85
		$\Lambda(t_1)$	0.111	0.270	0.051	0.049	0.06
		$\Lambda(t_2)$	0.298	0.699	0.119	0.115	0.03
		$\Lambda(t_3)$	0.549	1.255	0.208	0.203	0.03
20	50	β_1	0.5	0.487	0.347	0.340	0.93
		β_2	1.0	1.002	0.621	0.582	0.91
		θ	3.0	2.724	0.928	0.900	0.86
		$\Lambda(t_1)$	0.111	0.270	0.043	0.041	0.01
		$\Lambda(t_2)$	0.298	0.701	0.103	0.098	0.00
		$\Lambda(t_3)$	0.549	1.263	0.180	0.173	0.00

Table 4

Results of analyzing DRS data

Covariates	Est	SE	Z	p-value
treated vs untreated	−.0046	.0020	−2.2847	.022
diabetic type	.0057	.0034	1.705	.088
treat*type	−.0091	.0034	−2.673	.007
Normal frailty	−.000066	.000018	3.741	<.001
Exponential frailty	.0545	.0092	5.944	<.001

Table 5

Results of analyzing SOLVD data

Covariates	Est	SE	p-value	Est	SE	p-value
	Results using data from all 23 centers					
	Additive model			Multiplicative model		
Treatment vs placebo	−.0628	.0135	<.001	−.4590	.0717	<.001
Age	.0076	.0034	.03	.0645	.0201	.001
Male vs female	−.0211	.0154	.17	−.1314	.0861	.13
Ejection fraction	−.0040	.0007	<.001	−.0284	.0054	<.001
Frailty parameter	.0008	.0006	.083	.0449	.0125	<.001
	Results using data after excluding one center					
	Additive model			Multiplicative model		
Treatment vs placebo	−.0662	.0134	<.001	−.5013	.0751	<.001
Age	.0099	.0027	<.001	.0751	.0208	<.001
Male vs female	−.0172	.0159	.278	−.1254	.0902	.16
Ejection fraction	−.0039	.0007	<.001	−.0287	.0056	<.001
Frailty parameter	.0002	.0003	.404	.0144	.0274	.17