Published in final edited form as:
Biometrics. 2011 March ; 67(1): 8–17. doi:10.1111/j.1541-0420.2010.01444.x.

A Positive Stable Frailty Model for Clustered Failure Time Data with Covariate-Dependent Frailty

Dandan Liu[*], **John D. Kalbfleisch**, and **Douglas E. Schaubel**
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

Summary

In this article, we propose a positive stable shared frailty Cox model for clustered failure time data where the frailty distribution varies with cluster-level covariates. The proposed model accounts for covariate-dependent intracluster correlation and permits both conditional and marginal inferences. We obtain marginal inference directly from a marginal model, then use a stratified Cox-type pseudo-partial likelihood approach to estimate the regression coefficient for the frailty parameter. The proposed estimators are consistent and asymptotically normal and a consistent estimator of the covariance matrix is provided. Simulation studies show that the proposed estimation procedure is appropriate for practical use with a realistic number of clusters. Finally, we present an application of the proposed method to kidney transplantation data from the Scientific Registry of Transplant Recipients.

Keywords

Bridge distribution; Clustered failure times; Covariate-dependent frailty; Cox model; Positive stable frailty; Shared frailty

1. Introduction

Clustered failure time data are frequently observed in biomedical studies. For example, in the kidney transplantation setting, transplant failure times are of interest and can be taken as clustered failure times with transplant facilities as clusters. In family disease studies, time to disease onset is of interest and families are natural clusters. Subjects within cluster are correlated, with the intracluster dependence possibly due to sharing similar environmental and/or genetic conditions.

Several methods have been proposed for clustered failure time data. In general, these can be categorized into two broad strategies. In marginal models, the cluster structure is usually ignored when estimating the population-averaged covariate effect, but is used to derive valid standard error estimates. Marginal models can be used when the comparison of lifetimes across clusters is of interest. Examples include Wei, Lin, and Weissfeld (1989); Lee, Wei, and Amato (1992); and Spiekerman and Lin (1998). These authors used generalized estimating equations with an independence working assumption and the intracluster correlation structure left unspecified. As a result, some efficiency loss may occur, potentially affecting the significance of estimated covariate effects.

When the comparison of lifetimes within the same cluster is of interest, frailty models may be more appropriate. In this case, the correlation structure is specified by incorporating a

[*]dandanl@umich.edu.

random effect (frailty) that is common to subjects within the same cluster. The covariate effect is then interpreted as being conditional on the frailties and is cluster specific. One can also obtain marginal covariate effects by making additional assumptions about the frailty distribution as was done by Glidden and Self (1999) and Pipper and Martinussen (2003) under the Clayton-Oakes model. In frailty models, it is usually assumed that the frailty variables follow the same distribution across clusters, which implies equal intracluster dependence as well as between-cluster heterogeneity. This assumption may be violated in practice.

In studies comparing U.S. kidney transplant centers to the national average, the ratio of observed to expected deaths, known as the standardized mortality ratio (SMR), is used, with the expected deaths obtained from a marginal Cox model. An SMR > 1 indicates a mortality rate above the national average. In the shared frailty model, this statistic is actually a nonparametric Poisson-type estimator (Glidden and Vittinghoff, 2004) for the corresponding frailty, given the observed data in the center. An investigation of the SMRs suggests that there may be greater heterogeneity for smaller facilities, since SMRs for smaller centers are more frequently seen at either the top or at the bottom of the ordered list. Although this is partly due to sampling variance of the SMR estimator, it is also possible that an unequal degree of heterogeneity across centers results from varying cluster characteristics. This suggests a shared frailty model, but with the frailty distribution allowed to depend on cluster size. Other cluster-level covariates may also have an effect on the frailty distribution. For example, urban transplant facilities may exhibit more uniform practices than rural transplant hospitals, corresponding to less heterogeneity (smaller variance) for frailties of urban centers. In these examples of clustered failure time data, the population averaged effect is of primary interest. At the same time, however, the incorporation of cluster-level covariate effects on the frailty distribution is of practical interest and should be considered.

Similar situations exist for other types of clustered data. Prentice (1986) proposed a regression model for clustered binary data, in which the correlation between pairs of binary observation within clusters was assumed to depend on cluster-level covariates. Lin, Raz, and Harlow (1997) proposed a linear mixed model with heterogeneous within-cluster variances, where the within-cluster errors were assumed to follow a normal distribution with cluster-specific covariance matrix. Specifically, the variance of the measurement error was assumed to follow an inverse gamma distribution, where the mean depends on some linear combination of cluster-level covariates through a log link. Heagerty (1999) proposed a marginally specified logistic-normal model for longitudinal binary data in which the marginal mean, rather than the conditional mean, was regressed on covariates. In addition, a conditional model on a Gaussian latent variable is specified, where the random effect additively influences the logit of the conditional mean. Wang and Louis (2004) further extended this method to clustered binary data, allowing the distribution parameters of the random effect to depend on some cluster-level covariates. Their approach used a “bridge” distribution previously identified by Wang and Louis (2003) for the random effect to unify the form of the marginal and the conditional models. As a result, the conditional regression parameters can be expressed as functions of the marginal regression parameters and a parameter in the bridge distribution. Under this model, the regression parameter estimates have a direct marginal interpretation, while the conditional regression parameter estimates can easily be obtained. Moreover, the influence of the cluster-level covariates on the random effect can be estimated.

The positive stable distribution (Hougaard, 1986) serves as a bridge distribution for clustered failure time data under a Cox proportional hazards shared frailty model in the same sense as Wang and Louis (2003) since the resulting marginal regression parameter is a product of the conditional regression parameters and the frailty parameter. This relationship

allows both marginal and conditional inference, while accounting for intracluster dependence. The shared positive stable frailty model has attracted renewed attention recently (e.g., Fine, Glidden, and Lee, 2003; Martinussen and Pipper, 2005).

In this article, we propose a covariate-dependent positive stable shared frailty model. The bridge-type frailties are allowed to depend on cluster-level covariates and so to follow different distributions across clusters. Under this unified framework, the marginal regression parameters and the covariate effects on the frailty distribution can be consistently estimated. The major contributions of this paper are the methods proposed for modeling the effects of the cluster-level covariates on the frailty distribution and the corresponding estimation of the marginal regression effects.

The remainder of this article is organized as follows. In Section 2, we introduce the proposed covariate-dependent frailty model and describe the estimation procedures. We obtain the large sample properties of the model parameter estimators in Section 3 and Section 4 presents simulation studies. The proposed method is then applied to kidney transplant data from the Scientific Registry of Transplant Recipients (SRTR) in Section 5. In Section 6, we provide some concluding remarks and discussion. Proofs of the results are provided in the Appendix.

2. Model Specification and Estimation

2.1 The Positive Stable Shared Frailty Cox Proportional Hazards Model with Covariate-Dependent Frailty

In this section, we specify a positive stable shared frailty Cox model, with the frailty distribution depending on cluster-level covariates and the corresponding marginal hazard having a proportional hazards form. Our ultimate purpose is to estimate cluster-level covariate effects on the frailty distribution, as well as the correlation within clusters and heterogeneity between clusters. We first define the Cox-type conditional and marginal hazard functions through the “bridge” property of the positive stable distribution. The relationship between the conditional hazard parameters, marginal hazard parameters, and frailty distribution parameter can be obtained accordingly. Cluster-level covariates are related to the frailty distribution parameter through a link function. Finally, we derive the individual intensity process given the observed history of all the individuals with the parameters of interest. We begin this section by establishing the requisite notation.

Suppose we have measurements from subjects in K clusters and that the cluster sizes n_k (k = 1, 2, …, K) are independent and identically distributed bounded random variables. Given n_k, let D_{ik} and C_{ik} be the failure and censoring times for the ith individual (i = 1, …, n_k) in the kth cluster; let $T_{ik} = D_{ik} \wedge C_{ik}$ be the follow-up time and $\Delta_{ik} = I(D_{ik} \leq C_{ik})$ the observed death indicator. Let W_k denote the positive stable distributed frailty with dependence parameter α_k for the kth cluster that we use to describe within-cluster dependence possibly due to unobserved covariate information. Let Z_{ik} be a p-vector of time-independent covariates measured on individual (i, k). In addition, let X_k be a q-vector of time-independent cluster-level covariates that may influence α_k. Let $D_k = (D_{1k}, \ldots, D_{n_k k})$, with C_k and Z_k defined similarly. We assume that (D_k, C_k, Z_k, X_k, n_k, W_k) are independent and identically distributed for k = 1, …, K. Define the at-risk process $Y_{ik}(t) = I(T_{ik} \geq t)$ and the individual counting process $N_{ik}(t) = \Delta_{ik} I(T_{ik} \leq t)$. We define the filtrations

$$\mathcal{F}_t = \sigma\{N_{ik}(s),\ Y_{ik}(s),\ Z_{ik},\ X_k,\ n_k : k = 1, \ldots, K,\ i = 1, \ldots, n_k,\ 0 \leq s \leq t\}$$

and

$$\mathscr{H}_t=\sigma\{N_{ik}(s),\, Y_{ik}(s),\, Z_{ik},\, X_k,\, n_k,\, W_k{:}k=1,\, \ldots,\, K,\, i=1,\, \ldots,\, n_k,\, 0 \le s \le t\}.$$

Similar to Martinussen and Pipper (2005), we term $\mathscr{F}_t$ the observed filtration and $\mathscr{H}_t$ the conditional filtration.

We assume that W_k follows a positive stable distribution with shape parameter α_k ($0 < \alpha_k \le 1$). The positive stable distribution has been used by Hougaard (1986) for multivariate failure time data; its density function and Laplace transform are given by

$$f_{\alpha_k}(w)=-\frac{1}{\pi w}\sum_{i=1}^{\infty}\frac{\Gamma(i\alpha_k+1)}{i!}(-w^{-\alpha_k})^i\sin(\alpha_k i\pi),$$

and

$$L(s)=\varepsilon\{\exp(-sW_k)\}=\exp(-s^{\alpha_k}) \quad (s \ge 0),$$

respectively.

Given (Z_k, X_k, W_k, n_k), the failure time D_{ik}, $i = 1, \ldots, n_k$ are assumed to be independent with hazard function

$$\lim_{h\to 0^+} P(t \le D_{ik} \le t+h \mid D_{ik} \ge t,\, Z_k,\, X_k,\, n_k,\, W_k)/h=W_k\lambda_{0k}(t)e^{\beta_k^T Z_{ik}}, \quad (1)$$

where $\lambda_{0k}(t)$($k = 1, \ldots, K$) are unknown cluster-specific baseline hazard functions and β_k ($k = 1, \ldots, K$) are p-vectors of unknown cluster-specific regression parameters, all of which rely on α_k through the derived marginal hazard function below.

Since W_k has the positive stable distribution, the marginal hazard function of D_{ik} is given by

$$\lim_{h\to 0^+} P(t \le D_{ik} \le t+h \mid D_{ik} \ge t,\, Z_k,\, X_k,\, n_k)/h=h_0(t)e^{\gamma^T Z_{ik}}, \quad (2)$$

where $h_0(t)$ is an unspecified baseline hazard and γ is a p-vector of unknown marginal regression parameters. In this, we have assumed a constant marginal log hazard ratio γ, which, given (1) and (2), imposes the restriction $\gamma = \alpha_k\beta_k$, $k = 1, \ldots, K$. Note also that $\Lambda_{0k}(t)=H_0(t)^{\alpha_k^{-1}}$, where $\Lambda_{0k}(t)=\int_0^t\lambda_{0k}(s)\,ds$ and $H_0(t)=\int_0^t h_0(s)\,ds$.

We further relate X_k and α_k through a link function $\alpha_k = g(\eta, X_k)$ and let $\alpha_k^{-1}=g(\eta;X_k)$, where η is a $(q + 1)$-vector of unknown parameters. Here, we assume that $g(\cdot)$ is monotone and twice differentiable with respect to η. Since $\alpha_k \in (0, 1]$, a natural choice for g is the logit link function and we set

$$g(\eta;X_k)=1+e^{-\eta^T\tilde{X}_k}, \quad (3)$$

with $\tilde{X}_k=(1,X_k^T)^T$ and $\eta=(\eta_1,\eta_2^T)^T$ where η_1 is a scalar intercept and η_2 is a q-vector of regression parameters.

In addition, we assume that the D_{ik} and C_{ik} are independent given Z_{ik} for $i = 1, \ldots, n_k$. Under this conditional independent censoring assumption, model (1) implies that the individual intensity process with respect to the conditional filtration $\mathscr{H}_t$ is

$$\lambda_{ik}(t \mid \mathscr{H}_{t^-})=Y_{ik}(t)W_k\lambda_{0k}(t)e^{\beta_k^T Z_{ik}}. \quad (4)$$

By applying the innovation theorem (Andersen et al., 1993) to (4) and inserting the link function (3), the individual intensity process with respect to the observed filtration $\mathscr{F}_t$ is

$$\lambda_{ik}(t \mid \mathscr{F}_{t^-})=Y_{ik}(t)f_k(t)\lambda_{0k}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}, \quad (5)$$

Where $f_k(t) = E(W_k \mid \mathscr{F}_{t^-})$ has the explicit form

$$f_k(t)=\frac{E_{W_k}\left[W_k^{N_{\cdot k}(t^-)+1}e^{-W_k\sum_{i=1}^{n_k}\int_0^{t^-}Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}dH_0^{g(\eta;X_k)}(s)}\right]}{E_{W_k}\left[W_k^{N_{\cdot k}(t^-)}e^{-W_k\sum_{i=1}^{n_k}\int_0^{t^-}Y_{ik}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}}dH_0^{g(\eta;X_k)}(s)}\right]}, \quad (6)$$

with "·" denoting summation over a subscript.

2.2 Estimation

Model (4) differs from the existing positive stable shared frailty Cox proportional model in that it allows the frailty distribution parameter α_k to depend on cluster-level covariates, which induces the cluster-specific conditional regression parameter $\beta_k=\alpha_k^{-1}\gamma$ and the cluster-specific conditional baseline hazard $\lambda_{0k}(t)$. It can be easily seen that when $\eta_2 = 0$, α_k is a constant and the proposed model reduces to the common positive stable shared frailty model for which several estimation procedures have been developed. For example, Wang, Klein, and Moeschberger (1995) applied the E-M algorithm for parameter estimation. Fine et al. (2003) presented a simple estimation procedure that fitted a marginal model and stratified model separately and utilized the relationship $\alpha = \gamma/\beta$. Martinussen and Pipper (2005) proposed a likelihood-based estimation procedure based on the individual intensity process with respect to an observed filtration similar to (5), but with $\alpha_k = \alpha$ and $\beta_k = \beta$. However, we are not able to extend these estimation procedures in the proposed model, since the regression parameter β_k in the conditional hazard is cluster specific.

As can be seen in the existing literature, simulations and applications of the positive stable shared frailty model are usually based on small clusters, such as twin or family studies, especially when the estimation of frailties is needed. In order to apply the positive stable frailty model to studies with large clusters, it is useful to avoid the estimation of $f_k(t)$ in (6). We notice that model (5) can be written as

$$\lambda_{ik}(t \mid \mathscr{F}_{t^-})=Y_{ik}(t)\tilde{\lambda}_{0k}(t)e^{g(\eta;X_k)\gamma^T Z_{ik}},$$

where $\tilde{\lambda_{0k}}(t) = \lambda_{0k}(t) f_k(t)$, which is actually a stratified Cox model, except that the covariate effect is cluster specific and depends on a function of cluster-level covariates. The stratified partial likelihood approach (Cox, 1975; Kalbfleisch and Prentice, 2002) can be directly applied here. Due to the loss of information in $f_k(t)$ and the multiplicative relationship between g and γ, we cannot estimate the intercept term η_1 and the remaining parameters simultaneously. Therefore, our estimation procedure is actually based on two results from models (2) and (5), respectively.

Before proceeding, it is convenient to introduce the following two sets of notation for $k = 1, \ldots, K$ and $r = 0, 1, 2$,

$$\begin{aligned} S^{(r)}(\gamma, t) &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k} Y_{ik}(t) e^{\gamma^T Z_{ik}} Z_{ik}^{\otimes r}, \\ E(\gamma, t) &= S^{(1)}(\gamma, t)/S^{(0)}(\gamma, t), \\ V(\gamma, t) &= S^{(2)}(\gamma, t)/S^{(0)}(\gamma, t) - \{E(\gamma, t)\}^{\otimes 2}, \end{aligned}$$

Where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$, and

$$\begin{aligned} S_k^{(r)}(\eta;\gamma, t) &= \sum_{i=1}^{n_k} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} \left\{ g_1(\eta;X_k)\gamma^T Z_{ik} \right\}^{\otimes r}, \\ S_k^{(3)}(\eta;\gamma, t) &= \sum_{i=1}^{n_k} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} g_2(\eta;X_k)\gamma^T Z_{ik}, \\ S_k^{(4)}(\eta;\gamma, t) &= \sum_{i=1}^{n_k} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} g_1(\eta;X_k) Z_{ik}^T, \\ S_k^{(5)}(\eta;\gamma, t) &= \sum_{i=1}^{n_k} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} g(\eta;X_k) Z_{ik}^T, \\ S_k^{(6)}(\eta;\gamma, t) &= \sum_{i=1}^{n_k} Y_{ik}(t) e^{g(\eta;X_k)\gamma^T Z_{ik}} g_1(\eta;X_k)\gamma^T Z_{ik}^{\otimes 2} g(\eta;X_k), \\ E_k^1(\eta;\gamma, t) &= S_k^{(1)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t), \\ E_k^3(\eta;\gamma, t) &= S_k^{(3)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t), \\ E_k^4(\eta;\gamma, t) &= S_k^{(4)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t), \\ E_k^5(\eta;\gamma, t) &= S_k^{(5)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t), \\ V_k^1(\eta;\gamma, t) &= S_k^{(2)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t) - \{E_k^1(\eta;\gamma, t)\}^{\otimes 2}, \\ V_k^2(\eta;\gamma, t) &= S_k^{(6)}(\eta;\gamma, t)/S_k^{(0)}(\eta;\gamma, t) - E_k^1(\eta;\gamma, t) E_k^5(\eta;\gamma, t), \\ g_1(\eta;X) &= \partial g(\eta;X)/\partial\eta,\ g_2(\eta;X) = \partial g_1(\eta;X)/\partial\eta^T. \end{aligned}$$

We first estimate γ from model (2) by maximizing the pseudo partial log-likelihood

$$\ell_1(\gamma) = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \int_0^\tau \left\{ \gamma^T Z_{ik} - \log S^{(0)}(\gamma, t) \right\} dN_{ik}(t)$$

under the working independence assumption (Wei et al., 1989). The corresponding estimating equation can be written as

$$U_1(\gamma)=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\{Z_{ik}-E(\gamma,\,t)\}\,dN_{ik}(t).$$

Given an estimator $\hat{\gamma}$ of γ from model (2), we estimate η from model (5) by maximizing the pseudo-stratified partial log-likelihood

$$\ell_2(\eta;\hat{\gamma})=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\left\{g(\eta;X_k)\hat{\gamma}^T Z_{ik}-\log S_k^{(0)}(\eta;\hat{\gamma},\,t)\right\}dN_{ik}(t),$$

with corresponding score function,

$$U_2(\eta;\hat{\gamma})=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\left\{g_1(\eta;X_k)\hat{\gamma}^T Z_{ik}-E_k^1(\eta;\hat{\gamma},\,t)\right\}dN_{ik}(t).$$

Solving $U_2(\eta,\hat{\gamma})=0$, we can obtain the estimator $\hat{\eta}$ for η.

3. Asymptotic Properties

Denote γ_0 and η_0 as the true values of the parameters γ and η, respectively. In this section, we emphasize the large sample results for $\hat{\eta}$. We begin by restating a previously derived result. We list the assumed conditions, state a previously derived result, and then state the theorems for our estimators. Proofs are provided in the Appendix.

The following conditions are assumed throughout this article, where for all $k = 1, \ldots, K$ and some constant $\tau > 0$:

- **a.** $(D_k, C_k, Z_k, X_k, n_k, W_k)$ are independent and identically distributed;
- **b.** $P\{Y_{ik}(\tau) = 1\} > 0$ for $i = 1, \ldots, n_k$;
- **c.** $|Z_{ikl}| < B_Z < \infty$ and $|X_{kj}| < B_X < \infty$ for all $l = 1, \ldots, p$ and $j = 1, \ldots, q$ and some constants B_Z and B_X;
- **d.** $g(\cdot)$ is twice continuously differentiable with respect to η;
- **e.** γ_0 and η_0 are interior to the parameter space.
- **f.** The following matrices are positive definite,

$$\begin{aligned} A_1 &= \varepsilon\left\{\int_0^\tau V(\gamma_0,\,t)S^{(0)}(\gamma_0,\,t)dH_0(t)\right\},\\ A_2 &= \varepsilon\left\{\int_0^\tau V_k^1(\eta_0;\gamma_0,\,t)S_k^{(0)}(\eta_0;\gamma_0,\,t)f_k(t)\,d\Lambda_{0k}(t)\right\}. \end{aligned}$$

Large sample results for $\hat{\gamma}$ have been provided by Lee et al. (1992), who showed that $K^{1/2}(\hat{\gamma} - \gamma_0)$ is asymptotically mean zero normal with variance $\sum_1 = A_1^{-1}B_1A_1^{-1}$, where A_1 and B_1 can be consistently estimated by $\hat{A}_1 = K^{-1}\hat{I}$ and $\hat{B}_1 = K^{-1}\sum_{k=1}^{K}\hat{\psi}_k^{\otimes 2}$, with

$$\begin{aligned}\widehat{I} &= \sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau V(\widehat{\gamma},\, t)\, dN_{ik}(t),\\ \widehat{\psi}_k &= \sum_{i=1}^{n_k}\int_0^\tau \{Z_{ik} - E(\widehat{\gamma},\, t)\} \times \left\{dN_{ik} - Y_{ik}(t)e^{\widehat{\gamma}^T Z_{ik}} d\widehat{H}_0(t)\right\},\end{aligned}$$

where

$$\widehat{H}_0(t)=\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^t dN_{ik}(u)/S^{(0)}(\widehat{\gamma},\, u).$$

Theorem 1

Under conditions (*a*)−(*f*), $\hat{\eta}$ *is unique and converges almost surely to* η_0 *as* $K \to \infty$.

The proof of the consistency of $\hat{\eta}$ is similar to that of Prentice and Self (1983) and Lemma 3.1 in Andersen and Gill (1982).

Theorem 2

Under conditions (*a*)−(*f*), *the random vector* $K^{1/2}$ ($\hat{\eta} - \eta_0$) *converges weakly to a* (q + 1)*-variate normal vector with mean 0 and covariance matrix*

$$\sum\nolimits_2 = A_2^{-1}\left(A_2 + B_2 \sum\nolimits_1 B_2^T - 2CB_2^T\right) A_2^{-1},$$

where A_2 *is defined in condition (f) and*

$$\begin{aligned}B_2 &= \varepsilon\left\{\int_0^\tau V_k^2(\eta_0;\gamma_0,\, t) S_k^{(0)}(\eta_0;\gamma_0,t) f_k(t) d\Lambda_{0k}(t)\right\},\\ C &= \varepsilon\left\{u_k \psi_k^T\right\} A_1^{-1},\end{aligned}$$

with

$$\begin{aligned}u_k &= \sum_{i=1}^{n_k}\int_0^\tau \left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,t)\right\} dN_{ik}(t),\\ \psi_k &= \sum_{i=1}^{n_k}\int_0^\tau \{Z_{ik} - e(\gamma_0,t)\} \left\{dN_{ik} - Y_{ik}(t)e^{\gamma_0^T Z_{ik}}\, dH_0(t)\right\},\end{aligned}$$

where

$$e(\gamma,t)=\frac{\varepsilon\{S^{(1)}(\gamma,t)\}}{\varepsilon\{S^{(0)}(\gamma,t)\}}.$$

Using the proof of Theorems 1 and 2, together with the results from Lee et al. (1992), we can show that Σ_2 can be consistently estimated by

$$\widehat{\sum}_2 = \widehat{A}_2^{-1} \left(\widehat{A}_2 + \widehat{B}_2 \widehat{\sum}_1 \widehat{B}_2^T - 2\widehat{C}\widehat{B}_2^T \right) \widehat{A}_2^{-1}$$

with

$$\begin{aligned}
\widehat{A}_2 &= -K^{-1}\partial U_2(\eta;\gamma)/\partial \eta^T|_{\eta=\widehat{\eta},\gamma=\widehat{\gamma}} \\
&= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{ V_k^1(\widehat{\eta};\widehat{\gamma},t) - g_2(\widehat{\eta};X_k)\widehat{\gamma}^T Z_{ik} + E_k^3(\widehat{\eta};\widehat{\gamma},t) \right\} dN_{ik}(t), \\
\widehat{B}_2 &= -K^{-1}\partial U_2(\eta;\gamma)/\partial \gamma^T|_{\eta=\widehat{\eta},\gamma=\widehat{\gamma}} \\
&= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left\{ V_k^2(\widehat{\eta};\widehat{\gamma},t) - g_1(\widehat{\eta};X_k)Z_{ik}^T - E_k^4(\widehat{\eta};\widehat{\gamma},t) \right\} dN_{ik}(t), \\
\widehat{C} &= K^{-1}\sum_{k=1}^{K}\widehat{u}_k\widehat{\psi}_k^T\widehat{A}_1^{-1},
\end{aligned}$$

where

$$\widehat{u}_k = \sum_{i=1}^{n_k}\int_0^\tau \left\{ g_1(\widehat{\eta};X_k)\widehat{\gamma}^T Z_{ik} - E_k^1(\widehat{\eta};\widehat{\gamma},t) \right\} dN_{ik}(t).$$

4. Numerical Studies

Simulation studies were conducted to assess the finite sample behavior of $\hat{\eta}$. We also compare our method to that of Fine et al. (2003) under the special case where α_k is common among clusters.

In the first simulation study, clustered failure time data were simulated from models (3) and (4) with K = 50, 100; $H_0(t) = t$; $\gamma = (0.5, 1)^T$; η_1 = −0.5, −0.25, 0, 0.25, 0.5; and η_2 = 0.5. Cluster sizes were simulated from a discrete uniform distribution in the following four intervals [5, 20], [21, 50], [51, 100], and [101, 200] with approximately equal number of clusters in each interval. The cluster-level covariate X_k was the cluster size measured in units of 100 subjects. The positive stable frailties, W_k, were simulated following the method in Chambers, Mallows, and Stuck (1976),

$$W_k = \frac{\sin(\alpha_k W_{1k})}{\sin(W_{1k})^{1/\alpha_k}} \left[\frac{\sin\{(1-\alpha_k)W_{1k}\}}{W_{2k}} \right]^{(1-\alpha_k)/\alpha_k},$$

where W_{1k} and W_{2k} are independent, with W_{1k} following a uniform distribution $U(0, \pi)$ and W_{2k} following an exponential distribution with mean 1. The individual-level covariate $Z_{ik} = (Z_{ik1}, Z_{ik2})^T$ was independently generated, with Z_{ik1} from a Bernoulli distribution with p = 0.5 and Z_{ik2} from $N(0, 1)$ distribution. The censoring times were simulated from the uniform

distribution, U (0.25, 1), yielding censoring probabilities of approximately 46%. For each scenario, 1000 replicates were carried out.

The results are summarized in Table 1. We report bias of the sampling mean of the estimators (BIAS), the mean of the standard error estimators (ASE), empirical standard deviation of the estimators (ESD), and the 95% empirical coverage probability (CP). In the last column, we present the approximate range of a_k for the simulated data. We also present the results for γ. We can see that the estimator $\hat{\eta}$ is nearly unbiased. The (ASE) is generally fairly close to the ESD and, correspondingly, 95% empirical coverage probabilities are generally close to the nominal values. As the number of clusters increases from $K = 50$ to $K = 100$, the coverage probability is generally closer to the nominal value. In addition, as the value of a_k decreases, the coverage probability becomes lower. This may partly be due to the fact that, for a fixed sample size, the amount of independent information decreases as a_k decreases; that is, smaller value of a_k corresponds to stronger association within clusters.

To assess the asymptotic normality of the regression parameter estimates, we study the quantile-quantile (Q-Q) plots of $\hat{\eta}$ after being standardized against standard normal variable. In Figure 1, we show the Q-Q plots of $\hat{\eta}_1$ and $\hat{\eta}_2$ when $K = 100$ and $\eta_1 = -0.5$, 0, and 0.5. All six plots exhibit diagonal lines, which suggests that the asymptotic normal approximation is reasonable.

In the second simulation study, we compare the proposed method (LKS) with Fine et al. (2003) (FGL) when a_k is fixed for all clusters. We keep the same setting for H_0, γ, and K. The individual-level covariates and the censoring variable follow the same distribution as the first study. We fix $a_k = 0.5$ or $a_k = 0.75$ for all clusters. When using our method, we let $\eta_2 = 0$ and estimate η_1 only. For the FGL method, a is estimated by averaging the truncated ratio of the marginal and conditional regression parameter estimators. The results are displayed in Table 2, In order to facilitate the comparison, we show the results for $\hat{a}$ rather than $\hat{\eta}$.

Both methods give an almost unbiased estimator for a, and the estimated standard error and coverage probability are reasonable. Similar to the results in Table 1, when the number of clusters increases from 50 to 100, the asymptotic standard errors of the estimators decrease and the coverage probability tends to be closer to the nominal value. The asymptotic standard error estimators from the two methods are very close. The LKS method gives somewhat better coverage probability than FGL.

Simulations have been done under covariate-dependent frailty and common frailty settings. Since there is no existing method to compare with under the covariate-dependent frailty setting, we only make comparison under the common frailty setting. For this, three methods are available. Both the traditional EM method (Wang et al., 1995) and the Martinussen and Pipper (2005) method (MP) involve estimation of the frailties as missing data, which is computationally very slow when large number of deaths are observed for some clusters and does not yield standard error easily. On the other hand, FGL does not involve the estimation of the frailties as is the case with the LKS method. Since our primary application of interest has clusters with large number of observed deaths, we have compared our method to FGL only.

5. Application

We applied the proposed methods to data on deceased donor kidney transplants performed between 2000 and 2004 in the United States. Data were obtained from the SRTR. Failure time (recorded in days) was defined as the time from transplantation to graft failure, retransplantation, or death, whichever occurred first. There were 224 facilities and a total of

23,027 transplants included in the study. The facility size varied from 1 to 708. We fitted the proposed covariate-dependent frailty model to the data with the logit link function for the dependence parameter α_k. A total of 12 patient-level covariates and four cluster-level covariates are considered in the proportional hazards model. The same cluster-level covariates are included in the link function for α_k. Patient-level covariates included age at transplantation (by decade), race (African-American, Other), gender, time on dialysis (2 dummy variables), body mass index (BMI; 3 dummy variables) and primary cause of renal disease (4 dummy variables). Cluster-level covariates included percentage of female patients, percentage of African-American patients, percentage of patients caused by diabetes, and center size (per 100 patients) in a center.

We expect that any covariate that is associated with the between-cluster variability may also be related to within-cluster variation. Moreover, it is easier to interpret a covariate's effect on the frailty variance after adjusting for its effect on the hazard function itself. Therefore, as a modeling strategy, covariates included in the logit link function should also be represented in the marginal hazards model. Naturally, such cluster-level covariates will not be used in the second stage of the estimation procedures, due to the stratification.

Results of our analysis are shown in Table 3. Percentage of female patients has a significant effect ($p = 0.0063$) on the frailty parameter. It is found that facilities with fewer female patients tend to have a smaller value of α_k, which corresponds to greater heterogeneity in facility performance. The percentage of female patients also influences the hazards significantly. Upon examining the point estimates, one could interpret these results as being in the same direction, as higher percent female implies lower graft failure hazard and lower variation; both desirable outcomes.

6. Discussion

Covariate-dependent frailty models for clustered failure time data have rarely been studied previously. Wassell and Moeschberger (1993) proposed a bivariate survival model with the gamma frailty parameter depending on a pairwise covariate. Their approach only considered paired survival times in each cluster and cannot be applied to studies with larger cluster sizes. Wassell, Kulczycki, and Moyer (1995) also pointed out the increasing complexity of the application of a frailty model to clustered failure time data with larger group sizes. The model proposed in this article enables one to adjust for covariate effects on the frailty distribution and permits both marginal and conditional inference for clustered failure time data regardless of the group size. Further consideration of the proposed method reveals two additional advantages. First, model (5), on which we make inference, allows for covariate-by-cluster interaction. The covariate effect is multiplicatively influenced by clusters through the cluster-level covariate-dependent frailty parameter α_k. Second, with the rapid development of various methods for frailty models, researchers have begun to consider more carefully issues of ease of implementation and computation time (e.g., Fine et al., 2003; Liu and Huang, 2007). The proposed method performs well in both aspects. The method can be implemented using SAS IML. When we evaluated the computation time in the simulation study, it took approximately 4 hours for 1000 runs, with approximately one-third of the time spent on the PROC PHREG call.

Recalling that $\Lambda_{0k}(t) = H_0(t)^{\alpha_k^{-1}}$, we can estimate Λ_{0k} with $\hat{H}_0(t)^{g(\hat{\eta},\hat{X}_k)}$, $k = 1, \ldots, K$, where the estimator $\hat{H}_0(t)$ of $H_0(t)$ can be estimated from model (2) (see Spiekerman and Lin, 1998). Since the joint distribution of $\hat{H}_0(t)$ and $\hat{\eta}$ is complicated, we have not been able to obtain the asymptotic distribution of the Λ_{0k}'s.

We noted that when a cluster-level covariate is included in the conditional proportional hazard model, its effect is nearly nonidentifiable and does not interfere with the estimation of other covariate effects. This is due to the use of the stratified partial likelihood approach in the estimation. Since the motivation of the proposed method is to model cluster-level covariate effects on between-cluster heterogeneity and within-cluster association, the inclusion of a cluster-level covariate in the conditional hazard is not needed. On the other hand, one is able to obtain the marginal effect of a cluster-level covariate due to the proportional hazard in the marginal model.

For ease of computation and to avoid the estimation of the $f_k(t)$ (which is difficult for studies with large clusters), we first attempted using a stratified partial likelihood approach based on model (5) only. We found that this approach does not lead to useful estimators for the parameter η_1. As an alternative, we estimate γ from model (2), then use the estimator $\hat{\gamma}$ in model (5) to obtain a consistent estimator for η. The proposed estimation procedure is actually a two-step procedure. Such approach has been employed previously in the context of maximum likelihood by, for example, Gong and Samaniego (1981) and for the Clayton-Oakes model with a proportional hazards model for the margins by Glidden (2000). It should be noted that some efficiency is lost under the stratified partial likelihood approach in the second stage, as exemplified by the fact that the same estimation would be obtained if we let $f_k(t) = 1$.

Several areas of future research are possible. The proposed method relies on the specification of a link function, and model checking on this function is of potential interest. Future research on this method may also include the extension to other frailty distributions.

Acknowledgments

The authors thank the Scientific Registry of Transplant Recipients (SRTR) for access to the kidney transplant data. The SRTR is funded by a contract from the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services. This research was supported in part by National Institutes of Health grant R01 DK-70869 (DES). The authors are also grateful to the coordinating editor, associate editor, and a referee for suggestions that resulted in considerable improvement of the manuscript. The authors also thank Tempie Shearon of the University of Michigan Kidney Epidemiology and Cost Center for assistance with assembling the analysis files.

Appendix

Proof of Theorem 1

The individual counting process martingale for the observed filtration is

$$M_{ik}(t)=N_{ik}(t) - \int_0^t Y_{ik}(s) f_k(s) e^{g(\eta_0;X_k)\gamma_0^T Z_{ik}}\, d\Lambda_{0k}(s).$$

The proof of the consistency of $\hat{\eta}$ considers the following two processes,

$$\begin{aligned} G(\eta,\hat{\gamma}) &= K^{-1}\{l_2(\eta,\hat{\gamma},t) - l_2(\eta_0,\gamma_0,t)\} \\ &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau \left[\left\{g(\eta;X_k)\hat{\gamma}^T - g(\eta_0;X_k)\gamma_0^T\right\} Z_{ik} - \log\frac{S_k^{(0)}(\eta,\hat{\gamma},t)}{S_k^{(0)}(\eta_0,\gamma_0,t)}\right] dN_{ik}(t), \end{aligned}$$

and

$$\Xi(\eta)=K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\left[\left\{g(\eta;X_k)\gamma_0^T-g(\eta_0;X_k)\gamma_0^T\right\}Z_{ik}-\log\frac{S_k^{(0)}(\eta,\gamma_0,t)}{S_k^{(0)}(\eta_0,\gamma_0,t)}\right]Y_{ik}(t)f_k(t)e^{g(\eta_0;X_k)\gamma_0^TZ_{ik}}\,d\Lambda_{0k}(t).$$

The difference between them can be decomposed into two parts,

$$\begin{aligned}G(\eta,\hat{\gamma})-\Xi(\eta) &= \{G(\eta,\hat{\gamma})-G(\eta,\gamma_0)\}+\{G(\eta,\gamma_0)-\Xi(\eta)\}\\ &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\left\{g(\eta;X_k)Z_{ik}(\hat{\gamma}-\gamma_0)-\log\frac{S_k^{(0)}(\eta,\hat{\gamma},t)}{S_k^{(0)}(\eta,\gamma_0,t)}\right\}dN_{ik}(t)+K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^\tau\Big[\{g(\eta;X_k)-g(\eta_0;X_k)\}\times\gamma_0^T\end{aligned}$$

For each η, the first term on the right-hand side of the equation converges almost surely to zero due to the consistency of $\hat{\gamma}$ and under conditions (a) to (f), the second term is a summation of K independent and identical distributed zero mean random variables. By the Strong Law of Large Numbers (SLLN), as $K\to\infty$, $G(\eta,\hat{\gamma})$ converges almost surely to the same limiting function of η as $\Xi(\eta)$.

By the conditions (d) to (f), we can evaluate the first and the second derivatives of this limiting function by taking the partial derivatives inside the integral of $\Xi(\eta)$. The first derivative is thus

$$\varepsilon\left[\sum_{i=1}^{n_k}\int_0^\tau\left\{g_1(\eta;X_k)\gamma_0^TZ_{ik}-E_k^1(\eta;\gamma_0,t)\right\}Y_{ik}(t)f_k(t)\times e^{g(\eta_0;X_k)\gamma_0^TZ_{ik}}\,d\Lambda_{0k}(t)\right].$$

It is 0 at $\eta=\eta_0$. The second derivative

$$\varepsilon\left[-\sum_{i=1}^{n_k}\int_0^\tau V_k^1(\eta;\gamma_0,t)S_k^{(0)}(\eta;\gamma_0,t)f_k(t)d\Lambda_{0k}(t)\right]$$

is minus a positive definite matrix at $\eta=\eta_0$ by condition (f). Therefore, $G(\eta,\hat{\gamma})$ converges almost surely to a concave function of η with a unique maximum at $\eta=\eta_0$. Since $\hat{\eta}$ maximizes $G(\eta,\hat{\gamma})$, it follows that $\hat{\eta}\xrightarrow{a.s.}\eta_0$ as $K\to\infty$.

Proof of Theorem 2

The first-order Taylor series expansion of $K^{-1/2}U_2(\eta,\hat{\gamma})$ about $\eta=\eta_0$ and $\gamma=\gamma_0$ gives

$$K^{-1/2}U_2(\hat{\eta};\hat{\gamma})=K^{-1/2}U_2(\eta_0;\gamma_0)-\hat{B}_2(\eta_0;\gamma^*)K^{1/2}(\hat{\gamma}-\gamma_0)-\hat{A}_2(\eta^*;\hat{\gamma})K^{1/2}(\hat{\eta}-\eta_0),$$

where η^* is on the line segment between $\hat{\eta}$ and η_0 and γ^* is on the line segment between $\hat{\gamma}$ and γ_0. Thus, we have

$$K^{1/2}(\widehat{\eta}-\eta_0)=\widehat{A}_2^{-1}(\eta^*;\widehat{\gamma})\left\{K^{-1/2}U_2(\eta_0;\gamma_0)-\widehat{B}_2(\eta_0;\gamma^*)K^{1/2}(\widehat{\gamma}-\gamma_0)\right\}.$$

With the consistency of $\hat{\eta}$ and $\hat{\gamma}$ and the SLLN, we can show that $\widehat{A}_2(\eta^*;\widehat{\gamma}) \xrightarrow{p} A_2$, and $\widehat{B}_2(\eta_0;\gamma^*) \xrightarrow{p} B_2$ and that A_2 and B_2 can be consistently estimated by $\hat{A}_2$ and $\hat{B}_2$, respectively.

It has been noted in Section 3 that $K^{1/2}(\hat{\gamma} - \gamma_0)$ converges in distribution to $N(0, \Sigma_1)$. We will prove that $K^{-1/2}U_2(\eta_0; \gamma_0)$ converges in distribution to $N(0, A_2)$. It can be easily seen that the process $K^{-1/2}U_2(\eta_0; \gamma_0, t)$ can be written as a sum of orthogonal martingales,

$$K^{-1/2}U_2(\eta_0;\gamma_0,t)=K^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^t\left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik}-E_k^1(\eta_0;\gamma_0,s)\right\}\,dM_{ik}(s),$$

with predictable variation process

$$\begin{aligned}\left\langle K^{-1/2}U_2(\eta_0;\gamma_0)\right\rangle(t) &= K^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\int_0^t\left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik}-E_k^1(\eta_0;\gamma_0,s)\right\}^{\otimes 2}\times Y_{ik}(s)f_k(s)e^{g(\eta_0;X_k)\gamma_0^T Z_{ik}(s)}\,d\Lambda_{0k}(s)\\ &= K^{-1}\sum_{k=1}^{K}\int_0^t V_k^1(\eta_0;\gamma_0,s)S_k^{(0)}(\eta_0;\gamma_0,s)f_k(s)\,d\Lambda_{0k}(s).\end{aligned}$$

From Rebolledo's Theorem, the Weak Law of Large Numbers (WLLN) and condition (f), we can easily show that $K^{-1/2}U_2(\eta_0; \gamma_0, \tau)$ converges in distribution to a zero mean Gaussian vector with covariance matrix

$$\lim_{K=\infty}\left\langle K^{-1/2}U_2(\eta_0;\gamma_0)\right\rangle(\tau)=A_2.$$

Finally, we need to obtain the asymptotic covariance matrix of $K^{-1/2}U_2(\eta_0; \gamma_0)$ and $K^{1/2}(\hat{\gamma} - \gamma_0)$. We can see that both items can be written as a summation of K i.i.d. zero mean random vectors,

$$\begin{aligned}K^{-1/2}U_2(\eta_0;\gamma_0) &= K^{-1/2}\sum_{k=1}^{K}u_k,\\ K^{1/2}(\widehat{\gamma}-\gamma_0) &= K^{-1/2}\widehat{A}_1(\gamma^*)^{-1}\sum_{k=1}^{K}\psi_k+o_p(1),\end{aligned}$$

with

$$u_k(\eta_0;\gamma_0)=\sum_{i=1}^{n_k}\int_0^\tau \left\{g_1(\eta_0;X_k)\gamma_0^T Z_{ik} - E_k^1(\eta_0;\gamma_0,t)\right\}\, dN_{ik}(t),$$

and

$$\psi_k(\gamma_0,H_0)=\sum_{i=1}^{n_k}\int_0^\tau \{Z_{ik}-e(\gamma_0,t)\}\times\left\{dN_{ik}-Y_{ik}(t)e^{\gamma_0^T Z_{ik}}\, dH_0(t)\right\}.$$

With the consistency of $\hat{\gamma}$ and the WLLN, the asymptotic covariance matrix of $K^{-1/2}U_2$ (η_0; γ_0) and $K^{1/2}$ ($\hat{\gamma}-\gamma_0$) is $C=\varepsilon\left\{u_k\psi_k^T\right\}\, A_1^{-1}$.

In summary, $K^{1/2}$ ($\hat{\eta}-\eta_0$) converges in distribution to a N (0, Σ_2), where

$$\sum\nolimits_2=A_2^{-1}\left(A_2+B_2\sum\nolimits_1 B_2^T-2CB_2^T\right)\, A_2^{-1},$$

which can be consistently estimated by replacing each quantity with its corresponding estimator.

References

Andersen PK, Gill RD. Cox's regression model for counting process: A large sample study. Annals of Statistics. 1982; 10:1100–1120.

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

Chambers JM, Mallows CL, Stuck BW. A method for simulating stable random variables. Journal of the American Statistical Association. 1976; 71:340–344.

Cox DR. Partial likelihood. Biometrika. 1975; 62:269–276.

Fine JP, Glidden DV, Lee KE. A simple estimator for a shared frailty regression model. Journal of the Royal Statistical Society, Series B. 2003; 65:317–329.

Glidden DV. A two-stage estimator of the dependence parameter for the Clayton-Oakes model. Lifetime Data Analysis. 2000; 6:141–156. [PubMed: 10851839]

Glidden DV, Self S. Semiparametric likelihood estimation in the Clayton-Oakes failure time model. Scandinavian Journal of Statistics. 1999; 26:363–372.

Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. Statistics in Medicine. 2004; 23:369–388. [PubMed: 14748034]

Gong G, Samaniego FJ. Pseudo maximum likelihood estimation: Theory and applications. Annals of Statistics. 1981; 9:861–869.

Heagerty PJ. Marginally specified logistic-normal models for longitudinal binary data. Biometrics. 1999; 55:688–698. [PubMed: 11314994]

Hougaard P. A class of multivariate failure time distributions. Biometrika. 1986; 73:671–678.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2nd. New York: Wiley; 2002.

Lee, EW.; Wei, LJ.; Amato, DA. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein, JP.; Goel, PK., editors. Survival Analysis: State of the Art. Dordrecht: Kluwer Academic Publishers; 1992. p. 237-247.

Lin X, Raz J, Harlow S. Linear mixed models with heterogeneous within-cluster variances. Biometrics. 1997; 53:910–923. [PubMed: 9290222]

Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models. Statistics in Medicine. 2007; 27:2665–2683. [PubMed: 17910008]

Martinussen T, Pipper CB. Estimation in the positive stable shared frailty Cox proportional hazards model. Lifetime Data Analysis. 2005; 11:99–115. [PubMed: 15747592]

Pipper CB, Martinussen T. A likelihood based estimating equation for the Clayton-Oakes model with marginal proportional hazards. Scandinavian Journal of Statistics. 2003; 30:509–522.

Prentice RL. Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. Journal of the American Statistical Association. 1986; 81:321–327.

Prentice RL, Self SG. Asymptotic distribution theory for Cox-type regression models with general relative risk form. Annals of Statistics. 1983; 81:804–813.

Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. Journal of the American Statistical Association. 1998; 93:1164–1175.

Wang ST, Klein JP, Moeschberger ML. Semiparametric estimation of covariate effects using the positive stable frailty model. Applied Stochastic Models and Data Analysis. 1995; 11:121–133.

Wang Z, Louis TA. Matching conditional and marginal shapes in binary mixed-effect models using a bridge distribution function. Biometrika. 2003; 90:765–775.

Wang Z, Louis TA. Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. Biometrics. 2004; 60:884–891. [PubMed: 15606408]

Wassell JT, Moeschberger ML. A bivariate survival model with modified gamma frailty for assessing the impact of interventions. Statistics in Medicine. 1993; 12:241–248. [PubMed: 8456209]

Wassell JT, Kulczycki GW, Moyer ES. Frailty models of manufacturing effects. Lifetime Data Analysis. 1995; 1:161–170. [PubMed: 9385098]

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association. 1989; 84:1065–1073.

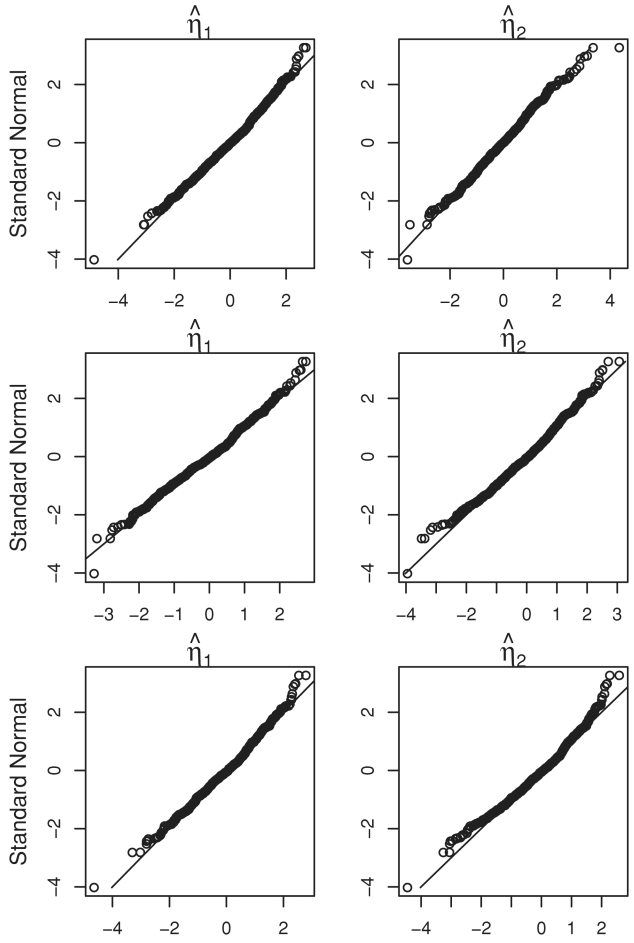

Figure 1.
Q-Q plots for $\hat{\eta_1}$ and $\hat{\eta_2}$ when $K = 100$ and $\eta_1 = -0.5$, 0, and 0.5.

Table 1

Summary of results for the first simulation study with $\eta_2 = 0.5$, $\gamma_1 = 0.5$, and $\gamma_2 = 1$ based on 1000 replicates

	$\hat{\eta}$						$\hat{\gamma}$					
K	**Parameter**	**True**	**BIAS**	**ASE**	**ESD**	**CP**	**Parameter**	**BIAS**	**ASE**	**ESD**	**CP**	**Range of a_k**
50	η_1	0.5	0.01	0.27	0.27	0.96	γ_1	0.01	0.06	0.06	0.92	0.63–0.82
	η_2		0.03	0.22	0.23	0.94	γ_2	0.00	0.08	0.08	0.90	
	η_1	0.25	0.02	0.24	0.24	0.96	γ_1	0.01	0.06	0.06	0.91	0.57–0.78
	η_2		0.01	0.18	0.19	0.94	γ_2	0.00	0.08	0.09	0.90	
	η_1	0	0.03	0.22	0.23	0.94	γ_1	0.01	0.06	0.07	0.92	0.51–0.73
	η_2		0.00	0.15	0.16	0.93	γ_2	0.00	0.09	0.10	0.91	
	η_1	−0.25	0.04	0.21	0.22	0.93	γ_1	0.01	0.07	0.07	0.91	0.44–0.68
	η_2		−0.01	0.12	0.13	0.92	γ_2	0.00	0.10	0.11	0.91	
	η_1	−0.5	0.07	0.20	0.22	0.91	γ_1	0.01	0.07	0.07	0.92	0.38–0.62
	η_2		−0.04	0.10	0.12	0.87	γ_2	0.01	0.11	0.11	0.91	
100	η_1	0.5	0.01	0.20	0.19	0.95	γ_1	0.01	0.04	0.04	0.94	0.63–0.82
	η_2		0.02	0.15	0.15	0.95	γ_2	0.00	0.06	0.06	0.92	
	η_1	0.25	0.02	0.17	0.18	0.95	γ_1	0.01	0.05	0.05	0.93	0.57–0.78
	η_2		0.00	0.13	0.12	0.95	γ_2	0.00	0.06	0.06	0.93	
	η_1	0	0.02	0.16	0.16	0.93	γ_1	0.01	0.05	0.05	0.93	0.51–0.73
	η_2		0.00	0.10	0.10	0.94	γ_2	0.00	0.07	0.07	0.93	
	η_1	−0.25	0.03	0.15	0.16	0.93	γ_1	0.01	0.05	0.05	0.93	0.44–0.68
	η_2		−0.02	0.09	0.09	0.92	γ_2	0.00	0.07	0.07	0.94	
	η_1	−0.5	0.04	0.14	0.15	0.91	γ_1	0.01	0.05	0.05	0.94	0.38–0.62
	η_2		−0.03	0.07	0.08	0.87	γ_2	0.00	0.08	0.08	0.93	

Table 2

Summary of results for the second simulation study comparing the proposed method (LKS) with FGL (see Fine et al., 2003) in the special case of constant $\alpha_k = \alpha$, $k = 1, \ldots, K$ with $\gamma_1 = 0.5$, $\gamma_2 = 1$, and 1000 replicates

			LKS				FGL			
K	**Parameter**	**True**	**BIAS**	**ASE**	**ESD**	**CP**	**BIAS**	**ASE**	**ESD**	**CP**
50	α	0.5	0.01	0.06	0.06	0.92	0.01	0.06	0.06	0.91
	γ_1		0.01	0.07	0.08	0.92				
	γ_2		0.01	0.11	0.12	0.91				
	α	0.75	0.01	0.06	0.06	0.91	0.01	0.06	0.06	0.87
	γ_1		0.02	0.06	0.06	0.91				
	γ_2		0.01	0.08	0.09	0.88				
100	α	0.5	0.00	0.04	0.04	0.94	0.00	0.04	0.05	0.94
	γ_1		0.01	0.05	0.05	0.94				
	γ_2		0.00	0.08	0.09	0.94				
	α	0.75	0.00	0.05	0.05	0.93	0.00	0.04	0.05	0.90
	γ_1		0.01	0.04	0.04	0.94				
	γ_2		0.00	0.06	0.06	0.92				

Table 3
Analysis of SRTR kidney transplant data

Covariates	Estimates	SE	p-value
γ (Patient Level)			
Age (in decades)	0.1541	0.0104	<.0001
African-American	0.2738	0.0293	<.0001
Female	−0.0957	0.0254	0.0002
Time on Dialysis (in years)			
≤1	−0.1379	0.0372	0.0002
>3	0.1153	0.0277	<.0001
Recipient BMI			
<20	0.0732	0.0502	0.1450
[25, 30)	0.0391	0.0298	0.1904
≥30	0.1369	0.0320	<.0001
Cause of ESRD			
Diabetes	0.2970	0.0359	<.0001
Hypertension	0.1646	0.0375	<.0001
Polycystic	−0.3106	0.0571	<.0001
Other	0.1156	0.0392	0.0032
γ (Cluster level)			
Percent of female (pct)	−0.0063	0.0022	0.0035
Percent of African-American (pct)	0.0039	0.0007	<.0001
Percent of diabetes (pct)	0.0084	0.0017	<.0001
Center size (in 100 patients)	0.0097	0.0068	0.1548
η			
Intercept	−2.3192	2.0945	0.2682
Percent of female (pct)	0.1046	0.0382	0.0063
Percent of African-American (pct)	0.0389	0.0325	0.2316
Percent of diabetes (pct)	0.0298	0.0531	0.5745
Center size (in 100 patients)	−0.2288	0.2705	0.3977