*The Canadian Journal of Statistics*
Vol. 44, No. 3, 2016, Pages 375–394
*La revue canadienne de statistique*



# Variable selection for frailty transformation models with application to diabetic complications

Xu LIU[1], Xinyuan SONG[2,3], Shangyu XIE[4]* and Yong ZHOU[1,5]

[1]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, P.R. China*
[2]*Shenzhen Research Institute, Chinese University of Hong Kong, Hong Kong, P.R. China*
[3]*Department of Statistics, Chinese University of Hong Kong, Hong Kong, P.R. China*
[4]*RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, P.R. China*
[5]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, P.R. China*

*Key words and phrases:* Diabetic complications; gamma frailty; MM algorithm; SCAD; transformation models; variable selection.

*MSC 2010:* Primary 62G08; secondary 62N01

*Abstract:* This article focuses on variable selection in the context of gamma frailty transformation models for multivariate survival times. We propose a method based on a nonconcave penalty function to select relevant regression predictors and simultaneously estimate model parameters and nonparametric functions. Our procedure performs as well as the oracle one in which the true model is assumed to be known. We conduct simulation studies to demonstrate the performance of the proposed methodology. We apply our method to a case study based on the Hong Kong Diabetes Registry, and obtain new insights on the risk factors of cardiovascular–renal complications for type 2 diabetic patients. *The Canadian Journal of Statistics* 44: 375–394; 2016 © 2016 Statistical Society of Canada

*Résumé:* Les auteurs se penchent sur la sélection de variables dans le contexte d'un modèle de fragilité gamma pour des durées de vie multivariées. Ils proposent une méthode basée sur une pénalité non concave pour choisir les prédicteurs utiles de la régression tout en estimant les paramètres du modèle et ses fonctions non paramétriques. La performance de leur procédure est équivalente à celle d'un oracle lorsque le vrai modèle est connu. Les auteurs présentent une étude de simulation pour évaluer la performance de la méthodologie proposée. Ils appliquent leur méthode à une étude de cas portant sur le registre du diabète de Hong Kong et obtiennent ainsi une nouvelle perspective sur les facteurs de risque pour les complications cardio-rénales des patients souffrant de diabète de type 2. *La revue canadienne de statistique* 44: 375–394; 2016 © 2016 Société statistique du Canada

## 1. INTRODUCTION

Diabetes is a major cause of morbidity, mortality, and health care costs (American Diabetes Association [ADA], 2008; CDCP, 2008). The health and economic burden of diabetes, primarily type 2 diabetes, is large today and expected to increase significantly in the next 15–25 years. By 2030, 4.4% of the global population (about 370 million people) is expected to have diabetes (ADA, 2008). In Asia type 2 diabetes is prevalent in young to middle-aged populations (Chan et al., 2009). The early onset of disease exposes patients to long disease duration with high risk

Additional supporting information may be found in the online version of this article at the publisher's website.
* *Author to whom correspondence may be addressed.*
*E-mail: xshyu@amss.ac.cn*

for cardiovascular–renal complications. Given that cardiovascular and renal diseases may share common risk factors, such as obesity, hypertension, and dyslipidaemia, it is of great interest to identify the risk factors that may have separate effects on one or more diabetic complications. Early detection of high-risk individuals and appropriate intervention treatments are beneficial for effectively preventing the progression of type 2 diabetic complications.

The Hong Kong Diabetes Registry was established at the Prince of Wales Hospital in 1995 as part of a quality improvement program with a weekly enrolment of 30–50 patients clinically diagnosed as having type 2 diabetes. In this article, a cohort of 2,871 patients with complete documentation of risk factors, complications, drug use, and clinical outcomes is considered. We focus on three types of diabetic complications, namely, stroke, heart failure, and kidney failure. The status and failure time to each of the three complications, together with possible risk factors related to age, sex, duration of diabetes, blood pressure, obesity, glycemic control, lipid control, and renal function, are measured. The main objective is to develop a sound statistical method to reveal the risk factors for the aforementioned diabetic complications.

Multivariate failure time data arise when each subject being studied can potentially experience several events (Kalbfleisch & Prentice, 2002). As multivariate failure times are often the records of times of occurrences of different events for the same subject, they are potentially dependent on one another; ignoring this dependence may lead to biased inference. To account for the dependence of correlated failure times and avoid modelling each type of event separately in a statistical model, it is natural and convenient to represent such dependence through a frailty term (i.e., random effect) (Clayton & Cuzick, 1985; Oakes, 1991; Zeng, Chen, & Ibrahim, 2009). Such a frailty term takes into account both the correlation of individuals and of failure times. A popular assumption is that the frailty variable comes from a gamma distribution (Wienke, Lichtenstein, & Yashin, 2003; Bhulai et al., 2009). For instance, Nielsen et al. (1992) incorporated covariates in the proportional hazards model (Cox, 1972) with gamma frailty. In this article we consider a gamma frailty transformation model with multivariate failure time data. This model is general because it incorporates a class of transformations that are indexed by some parameters and determined by the data. As pointed out by Zeng, Chen, & Ibrahim (2009), this model includes the gamma frailty model as a special case and allows the random effects to be zero. Further selecting a specific transformation function according to the data enables the misspecification problem to be effectively avoided. The model also includes the frailty model (Fan & Li, 2002) and common baseline hazard model (Cai et al., 2005) as its special cases.

Another important and challenging problem is how to select the most relevant covariates among a large number of candidates in substantive research. Traditional procedures for variable selection include the use of information criteria, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). For linear regression models, many regularization methods for variable selection have been developed in the literature including, but not limited to, the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), bridge regression (Frank & Friedman, 1993; Huang et al., 2009) smoothly clipped absolute deviation (SCAD) (Fan & Li, 2001), and their extensions, such as the adaptive LASSO (Zou, 2006), group LASSO (Yuan & Lin, 2006), and the elastic net (Zou & Hastie, 2005). Some of the variable selection criteria and procedures for linear regression analysis have been extended to the context of survival data analysis. For instance, Tsai (2009) extended his LASSO method to Cox models. Wang et al. (2009) proposed a variable selection procedure for penalized Cox regression with grouped variables. Fan & Li (2002) extended their SCAD-penalized likelihood approach to Cox and frailty models. Cai et al. (2005) extended the SCAD penalty method to multivariate failure time data. Recently Androulakis, Koukouvinos, & Vonta (2012) conducted variable selection for frailty models based on penalized likelihood. Liu & Zeng (2013) developed an adaptive LASSO procedure to perform variable selection for transformation models with right-censored data. However their models excluded either a transformation function or frailty term. Inspired by Fan & Li (2002) and Cai et al. (2005) we consider variable selection for a gamma frailty transformation model with multivariate failure time data. Compared with the existing literature, our model framework is more general and flexible.

The main contributions of this article are as follows. First a novel variable selection procedure based on a nonconcave penalty function is developed in the context of gamma frailty transformation models with multivariate failure time data. The proposed method has the advantage of conducting parameter estimation and variable selection through a single model estimation procedure. Second our proposed approach has the oracle property. That is our procedure performs as well as the oracle one in which the true model is assumed to be known. Third the developed methodology is applied to a study of diabetic complications. Potential risk factors, which may have separate or common effects on stroke, heart failure, and kidney failure, are examined and identified. The new insights will be beneficial to the effective intervention of prevalence and incidence of diabetic complications, thereby improving the quality of life of type 2 diabetic patients.

The rest of this article is organized as follows. Section 2.1 presents the log-likelihood function proposed by Zeng, Chen, & Ibrahim (2009) based on the cumulative hazard function for the gamma frailty transformation model with multivariate failure time data. To conduct variable selection and estimation, the penalized maximum likelihood estimation (PMLE) procedure using a SCAD penalty is proposed in Section 2.2. The consistency and oracle property of the PMLE are established in Section 2.3. Section 3 reports numerical comparisons and simulation studies. Section 4 applies the proposed methodology to a study of diabetic complications. Section 5 concludes the article. The minorization–maximization (MM) algorithm and the estimated covariance matrix of the parameter estimator are presented in the Appendix. Other technical details are available online as Supplementary Material.

## 2. METHODOLOGY

### 2.1. Gamma Frailty Transformation Model

We consider the gamma frailty transformation model proposed by Zeng, Chen, & Ibrahim (2009). Let $T_{ik}$ denote the failure time of event type $k$ ($k = 1, \ldots, K$). We assume that given $w_i$,

$$\Lambda_{ik}(t|w_i, \mathbf{X}_i) = G_k\{\Lambda_k(t)\exp(\boldsymbol{\beta}_k^\top \mathbf{X}_i)\}w_i, \tag{1}$$

where $G_k$ is a known transformation function, $\Lambda_k(t)$ is an unspecified baseline function, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})^\top$ is a vector of regression coefficients, $\mathbf{X}_i$ is a vector of covariates, and $w_i$ is a frailty term following the gamma distribution with mean unit and variance $\theta$, say $g_\theta(w)$. Let $\boldsymbol{\beta}_{k0} = (\beta_{k01}, \ldots, \beta_{k0p})^\top = (\boldsymbol{\beta}_{k10}^\top, \boldsymbol{\beta}_{k20}^\top)^\top$ be the true value of $\boldsymbol{\beta}_k$. Without loss of generality, it is assumed that $\boldsymbol{\beta}_{k20} = \mathbf{0}$, and all the components of $\boldsymbol{\beta}_{k10}$ are not equal to 0. This model is a natural generalization of the ordinary frailty model. It yields the ordinary one as a special case but allows flexibility in the choice of transformations.

We further assume that the study ends at some finite $\tau$. Under these assumptions we can rewrite the above model as

$$\Lambda_{ik}(t|w_i, \mathbf{X}_i) = G_k\{F_k(t)\exp(\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X}_i)\}w_i, \tag{2}$$

where $F_k(t) = \Lambda_k(t)/\Lambda_k(\tau)$ and $\alpha_k = \log \Lambda_k(\tau)$. Clearly $F_k(\tau) = 1$. That is $F_k(t)$ is a distribution function in $[0, \tau]$.

If we have a set of right-censored data, the observed data for a single cluster are $\{Y_k, \Delta_k, \mathbf{X}_k\}$, $k = 1, \ldots, K$, where $Y_k = \min(T_k, C_k)$, $\Delta_k = I(T_k \le C_k)$, $C_k$ is the censoring time for event type $k$, and $I(A)$ is an indictor function with 1 if event $A$ occurs and 0 otherwise. Therefore under the assumption that the censoring time is independent of the failure time and the frailty given the covariates, the likelihood function is

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F) = \Pi_{k=1}^{K} \left[ G_k'\{F_k(Y_k)\exp(\alpha_k + \beta_k^\top \mathbf{X})\} F_k(Y_k)\exp(\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X}) \right]^{\Delta_k}$$
$$\times \int w^{\left(\sum_{k=1}^{K} \Delta_k\right)} \exp\left( -w \left[ \sum_{k=1}^{K} G_k\{F_k(Y_k)\exp(\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X})\} \right] \right) g_\theta(w) dw,$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_K^\top)^\top$, $F = (F_1, \ldots, F_K)^\top$, $F_k(t)$ denotes the jump size of $F_k$ at $t$, and $G_k'(\cdot)$ denotes the first derivative of $G_k(\cdot)$. To estimate $\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta$, and $F$ we maximize the log-likelihood function $\ell_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F) = \log L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F)$ based on $n$ independently and identically distributed clusters.

Inspired by Zeng, Chen, & Ibrahim (2009) we regard $w_i$ as missing data and use the expectation–maximization (EM) algorithm to conduct the analysis. In the E-step, the conditional expectation of some function $h(w_i)$, given the observed data, is evaluated analytically or via a Laplace approximation, $\hat{E}\{h(w_i)\}$. In the M-step we maximize the following complete-data log-likelihood function:

$$\ell_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F) = \sum_{i=1}^{n}\sum_{k=1}^{K} \Delta_{ik} \big[ \log G_k'\{F_k(Y_{ik})\exp(\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X}_i)\} + \log F_k\{Y_{ik}\}$$
$$+ \alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X}_i + \hat{E}(\log w_i) \big] - \sum_{i=1}^{n} \hat{E}(w_i) \sum_{k=1}^{K} G_k\{F_k(Y_{ik})\exp(\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{X}_i)\}$$
$$- n\log(\theta^{1/\theta}\Gamma(1/\theta)) + (1/\theta - 1)\sum_{i=1}^{n} \hat{E}(\log w_i) - \theta^{-1}\sum_{i=1}^{n} \hat{E}(w_i).$$

The method proposed by Zeng, Chen, & Ibrahim (2009) works well in the case where the number of covariates is small. However substantive studies often involve a large number of covariates. Hence identification of the most relevant variables for achieving a parsimonious model and improving the efficiency of estimation is important. In the next section we propose a model selection method using the penalized log-likelihood function to select significant covariates and simultaneously estimate model parameters as well as the nonparametric baseline function.

## 2.2. Variable Selection

In this section we focus on variable selection in the context of model (2). We consider the penalized log-likelihood function

$$Q_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F) = \ell_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F) - n\sum_{k=1}^{K}\sum_{j=1}^{p} p_{\eta_k}(|\beta_{kj}|),$$

where $p_{\eta_k}(\cdot)$ is a penalty function with threshold parameter $\eta_k$. Maximizing $Q_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta, F)$ with respect to $\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta$, and $F$ leads to penalized maximum likelihood estimators, $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\theta}$, and $\hat{F}$. With the proper choice of $\eta_k$ and a penalty function, some of the estimated parameters will be zero and hence their corresponding variables do not appear in the model. This penalized procedure achieves the objective of variable selection. We usually can choose a distinct regular parameter, $\eta_k$, for each $k$. To avoid computational burden we choose the same $\eta$ for all $k$, $k = 1, \ldots, K$.

A good penalty function should result in an estimator with the following three properties (Fan & Li, 2001): (a) unbiasedness for large true coefficients to avoid excessive estimation bias; (b) sparsity with a threshold rule that automatically sets small estimated parameters to zero to reduce model complexity; and (c) continuity to avoid unnecessary variation in model prediction.

A simple penalty function that satisfies the above three properties is the SCAD penalty, which is defined by

$$p'_\eta(u) = \eta \left\{ I(u \leq \eta) + \frac{(a\eta - u)_+}{(a-1)\eta} I(u \geq \eta) \right\},$$

for some $a > 2$ and $u > 0$, where $\eta$ is a threshold parameter. Fan & Li (2001) suggested the use of $a = 3.7$ from a Bayesian point of view. We adopt this value throughout the article.

### 2.3. Oracle Properties to Variable Selection

Let $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_{k1}^\top, \boldsymbol{\beta}_{k2}^\top)^\top$, its true value $\boldsymbol{\beta}_{k0} = (\boldsymbol{\beta}_{k10}^\top, \boldsymbol{\beta}_{k20}^\top)^\top$, and its penalized maximum likelihood estimators $(\hat{\boldsymbol{\beta}}_{k1}^\top, \hat{\boldsymbol{\beta}}_{k2}^\top)^\top = (\hat{\beta}_{k1}, \ldots, \hat{\beta}_{kp})^\top$, $k = 1, \ldots, K$. Without loss of generality we assume that $\boldsymbol{\beta}_{k20} = \mathbf{0}$. Define $\boldsymbol{\beta}_{d10} = (\boldsymbol{\beta}_{110}^\top, \ldots, \boldsymbol{\beta}_{K10}^\top)^\top$ and $\boldsymbol{\beta}_{d20} = (\boldsymbol{\beta}_{120}^\top, \ldots, \boldsymbol{\beta}_{K20}^\top)^\top$. Let $\xi_{10} = (\boldsymbol{\alpha}_0^\top, \theta_0, \boldsymbol{\beta}_{d10}^\top)^\top$ and $\xi_0 = (\xi_{10}^\top, \boldsymbol{\beta}_{d20}^\top)^\top$, and let $s^*$ denote the number of components of $\xi_{10}$. We then can define the profile likelihood as follows:

$$p\ell_n(\xi) = \sup_{F \in \mathscr{F}} \ell_n(\xi, F),$$

where $\xi$ is the vector corresponding to $\xi_0$, $\mathscr{F} = \{F = (F_1(t_1^0), \ldots, F_K(t_K^0))^\top : F_k(t_k^0) = (\Lambda_k(t_{1k}^0)/\Lambda_k(\tau), \ldots, \Lambda_k(t_{Nk}^0)/\Lambda_k(\tau)), k = 1, \ldots, K\}$, and $t_{1k}^0 < \cdots < t_{Nk}^0$ denote the ordered observed failure times for event type $k$. Under some regularity conditions, Murphy & van der Vaart (2000) proved that for any random sequence $\xi_n \to \xi_0$ in probability,

$$\begin{aligned} p\ell_n(\xi_n) = p\ell_n(\xi_0) &+ (\xi_n - \xi_0)^\top \sum_{i=1}^{n} \ell_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\} \\ &- \frac{1}{2} n(\xi_n - \xi_0)^\top I_0(\xi_0)(\xi_n - \xi_0) + o_p(\sqrt{n}||\xi_n - \xi_0|| + 1)^2, \end{aligned} \tag{3}$$

where $\ell_0$ is the *efficient score function for* $\xi_0$, the ordinary score function minus its orthogonal projection onto the closed linear span of the score functions for the nuisance parameter, and $I_0(\xi_0)$ is the covariance matrix, the *efficient Fisher information matrix*. $I_0(\xi_0)$ is nonsingular according to Theorem 4 in Asgharian (2014) assuming the boundness of $\Lambda_k(\tau)$. The nonsingularity of $I_0(\xi_0)$ can also be shown in a similar way in Theorem 2 and Proposition A.2 of Zeng & Lin (2007).

When the asymptotic property (Eq. 3) based on MM algorithm holds, we can establish the oracle property (Fan & Li, 2001) for the penalized profile likelihood according to Equation (3) for the gamma frailty model. We denote the penalized profile likelihood by

$$Q_n(\xi) = p\ell_n(\xi) - n \sum_{k=1}^{K} \sum_{j=1}^{p} p_{\eta_n}(|\beta_{kj}|).$$

Let $\hat{\xi} = (\hat{\xi}_1^\top, \hat{\boldsymbol{\beta}}_{d2}^\top)^\top$ be the maximizer of $Q_n(\xi)$, where $\hat{\xi}_1 = (\hat{\boldsymbol{\alpha}}^\top, \hat{\theta}, \hat{\boldsymbol{\beta}}_{d1}^\top)^\top$, $\hat{\boldsymbol{\beta}}_{d1} = (\hat{\boldsymbol{\beta}}_{11}^\top, \ldots, \hat{\boldsymbol{\beta}}_{K1}^\top)^\top$, and $\hat{\boldsymbol{\beta}}_{d2} = (\hat{\boldsymbol{\beta}}_{12}^\top, \ldots, \hat{\boldsymbol{\beta}}_{K2}^\top)^\top$. To show the oracle property we introduce the following notations. Let $\Sigma = \text{diag}\{p''_{\eta_n}(|\tilde{\beta}_{10}|), \ldots, p''_{\eta_n}(|\tilde{\beta}_{s0}|)\}$, and $b = (p'_{\eta_n}(|\tilde{\beta}_{10}|), \ldots, p'_{\eta_n}(|\tilde{\beta}_{s0}|))^\top$, where $(\tilde{\beta}_{10}, \ldots, \tilde{\beta}_{s0})^\top = \boldsymbol{\beta}_{d10}$, and $s = s^* - K - 1$ is the number of its components. Define $\Sigma_1 = \text{diag}(\mathbf{0}_1, \Sigma)$ and $\mathbf{b}_1 = (\mathbf{0}_2, b^\top)^\top$, where $\mathbf{0}_1$ and $\mathbf{0}_2$ denote the $d \times d$ matrix with all elements 0 and the $d \times 1$ vector with all elements 0, respectively, in which $d = K + 1$ is the number of components of the vector $(\boldsymbol{\alpha}_0^\top, \theta_0)^\top$. The dimension of the matrix $\Sigma_1$ is $s^*$, and the number of components of the vector $\mathbf{b}_1$ is $s^*$.

The consistency and oracle property of penalized profile maximum likelihood estimators require some regularity conditions, which are given in Supplementary Material. Those conditions were provided by Zeng, Chen, & Ibrahim (2009) to show the consistency and asymptotic normality of nonparametric MLE, and can also assure that Equation (3) holds. As a result, the maximizer of the profile likelihood is consistent according to Murphy & van der Vaart (2000) and Zeng, Chen, & Ibrahim (2009). In Theorem 1 given below we provide the rates of convergence for the penalized logarithm of profile likelihood estimators, which depend on the regular parameter $\eta_n$.

**Theorem 1.** *Assume that the regularity conditions* (1)–(3) *hold in Supplementary Material,* $\{\mathbf{X}_{ik}, T_{ik}, C_{ik}\}_{k=1}^K$ *are independent random samples for* $i = 1, \ldots, n$, $w_i$ *and* $(\mathbf{X}_{ik}, Y_{ik}, C_{ik})$, $k = 1, \ldots, K$, *are independent and satisfy (2),* $T_{ik}$ *and* $C_{ik}$ *are conditionally independent given* $\mathbf{X}_i$, *and* $w_i$ *are i.i.d. from a gamma distribution with mean 1 and variance* $\theta$. *If* $b_n \to 0$, *then there exists a local maximizer* $\hat{\xi}$ *of* $Q_n(\xi)$ *such that* $||\hat{\xi} - \xi_0|| = O_p(n^{-1/2} + a_n)$, *where* $a_n = \max\{p'_{\eta_n}(\beta_{kj0}) : \beta_{kj0} \neq 0\}$, $j = 1, \ldots, p$ *and* $k = 1, \ldots, K$.

Theorem 1 shows that there exists a $\sqrt{n}$-consistent penalized logarithm of profile likelihood estimator when $a_n = O(n^{-1/2})$ if a proper regular parameter $\eta_n$ is chosen. The following theorem gives the oracle property under some regularity conditions.

**Theorem 2.** *Assume that the penalty function* $p_{\eta_n}(\cdot)$ *satisfies Condition (4) in Supplementary Material. Under the conditions of Theorem 1, if* $\eta_n \to 0$ *and* $\sqrt{n}\eta_n \to \infty$, *then with probability tending to 1, the penalized maximum likelihood estimator* $\hat{\xi} = (\hat{\xi}_1^\top, \hat{\boldsymbol{\beta}}_{d2}^\top)^\top$ *must satisfy:*

*(a) (Sparsity)* $\hat{\boldsymbol{\beta}}_{d2} = \mathbf{0}$;
*(b) (Asymptotic normality)*

$$\sqrt{n}\{\hat{\xi}_1 - \xi_{10} + [I_1(\xi_{10}) + \Sigma_1]^{-1}\mathbf{b}_1\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, V_1(\xi_{10})),$$

*where* $\xrightarrow{\mathcal{L}}$ *denotes convergence in distribution,* $V_1(\xi_{10}) = [I_1(\xi_{10}) + \Sigma_1]^{-1} I_1(\xi_{10})[I_1(\xi_{10}) + \Sigma_1]^{-1}$, *and* $I_1(\xi_{10})$ *consists of the first* $s^* \times s^*$ *submatrix of* $I_0(\xi_0)$.

The theoretical proofs of Theorems 1 and 2 are provided in Supplementary Material. The MM algorithm, the variance estimation, and the selection of tuning parameter $\eta_n$ are given in the Appendix.

## 3. NUMERICAL STUDY

In the following examples, the performance of the proposed procedures in terms of model error and model complexity is compared. With $\mu(\mathbf{X}) = E(T|\mathbf{X})$, the relative model error (RME) is defined as $E\{\hat{\mu}(\mathbf{X}) - \mu(\mathbf{X})\}^2$, where the expectation is taken with respect to a new observation $\mathbf{X}$ as in Hunter & Li (2005), $T$ is the failure time, and $\hat{\mu}(\mathbf{X})$ denotes $\mu(\mathbf{X})$ with replacement of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$. The RME of the proposed procedure is compared to that of the maximum likelihood estimators proposed by Zeng, Chen, & Ibrahim (2009). The median and standard deviation (in

TABLE 1: Relative model errors for Example 1.

| n | Method | RME Median | Zeros C | Zeros In | RME Median | Zeros C | Zeros In | RME Median | Zeros C | Zeros In |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $G_{\rho 1}(x;1)$[a] | | | $G_{\rho 2}(x;1)$ | | | Combination | | |
| 200 | PMLE | 0.499 (0.271) | 3.996 | 0.024 | 0.487 (0.263) | 3.996 | 0.024 | 1.003 (0.399) | 7.992 | 0.048 |
| | Oracle | 0.496 (0.258) | 4.000 | 0.000 | 0.479 (0.259) | 4.000 | 0.000 | 0.993 (0.384) | 8.000 | 0.000 |
| 400 | PMLE | 0.453 (0.257) | 4.000 | 0.002 | 0.497 (0.263) | 4.000 | 0.000 | 0.985 (0.400) | 8.000 | 0.002 |
| | Oracle | 0.453 (0.255) | 4.000 | 0.000 | 0.497 (0.263) | 4.000 | 0.000 | 0.983 (0.400) | 8.000 | 0.000 |
| | | $G_{\rho 1}(x;1)$ | | | $G_{r2}(x;1)$[a] | | | Combination | | |
| 200 | PMLE | 0.486 (0.265) | 3.996 | 0.012 | 0.491 (0.264) | 3.998 | 0.042 | 1.006 (0.396) | 7.994 | 0.054 |
| | Oracle | 0.483 (0.262) | 4.000 | 0.000 | 0.490 (0.259) | 4.000 | 0.000 | 1.002 (0.384) | 8.000 | 0.000 |
| 400 | PMLE | 0.458 (0.258) | 4.000 | 0.000 | 0.493 (0.262) | 4.000 | 0.006 | 0.961 (0.403) | 8.000 | 0.006 |
| | Oracle | 0.459 (0.258) | 4.000 | 0.000 | 0.493 (0.258) | 4.000 | 0.000 | 0.961 (0.399) | 8.000 | 0.000 |
| | | $G_{\rho 1}(x;0.5)$ | | | $G_{r2}(x;0.5)$ | | | Combination | | |
| 200 | PMLE | 0.475 (0.248) | 3.998 | 0.024 | 0.511 (0.262) | 3.998 | 0.032 | 0.985 (0.382) | 7.996 | 0.056 |
| | Oracle | 0.474 (0.246) | 4.000 | 0.000 | 0.504 (0.256) | 4.000 | 0.000 | 0.982 (0.373) | 8.000 | 0.000 |
| 400 | PMLE | 0.443 (0.259) | 4.000 | 0.004 | 0.503 (0.256) | 4.000 | 0.000 | 0.960 (0.394) | 8.000 | 0.004 |
| | Oracle | 0.443 (0.255) | 4.000 | 0.000 | 0.506 (0.256) | 4.000 | 0.000 | 0.960 (0.393) | 8.000 | 0.000 |
| | | $G_{r1}(x;1)$ | | | $G_{r2}(x;1)$ | | | Combination | | |
| 200 | PMLE | 0.467 (0.248) | 4.000 | 0.044 | 0.497 (0.262) | 3.998 | 0.038 | 0.955 (0.382) | 7.998 | 0.082 |
| | Oracle | 0.465 (0.244) | 4.000 | 0.000 | 0.495 (0.259) | 4.000 | 0.000 | 0.947 (0.376) | 8.000 | 0.000 |
| 400 | PMLE | 0.427 (0.252) | 4.000 | 0.006 | 0.500 (0.257) | 4.000 | 0.006 | 0.948 (0.379) | 8.000 | 0.012 |
| | Oracle | 0.426 (0.249) | 4.000 | 0.000 | 0.500 (0.254) | 4.000 | 0.000 | 0.948 (0.374) | 8.000 | 0.000 |

[a] $G_{\rho k}(x;c)$ and $G_{rk}(x;c)$ imply that $G_k(\cdot) = G_\rho(x;c)$ and $G_k(\cdot) = G_r(x;c)$, for the $k = 1$ and 2 type events, and $c = 1$ or 0.5, respectively.

parentheses) of RME over 500 simulated data sets are summarized in Tables 1 and 4. In this article, the conditional expectation $\mu(\mathbf{X})$ is

$$\mu(\mathbf{X}) = E(T|\mathbf{X}) = \int_0^\infty \int tG'\{\Lambda(t)\exp(\boldsymbol{\beta}^\top\mathbf{X})\}\lambda(t)\exp(\boldsymbol{\beta}^\top\mathbf{X})w\exp(-G\{\Lambda(t)\exp(\boldsymbol{\beta}^\top\mathbf{X})\}w)g_\theta(w)dwdt.$$

For the conditional expectation $\int w\exp(-G\{\Lambda(t)\exp(\boldsymbol{\beta}^\top\mathbf{X})\}w)g_\theta(w)dw$ we use Laplace approximation.

We conduct several simulation studies to examine the finite sample performance of the proposed estimation procedure. Similar to Zeng, Chen, & Ibrahim (2009) we assume two event types

($k = 1, 2$) with the following transformations:

$$G(x) = \begin{cases} G_\rho(x;\rho) = ((1+x)^\rho - 1)/\rho, & (\rho \geq 0) \\ G_r(x;r) = 1/r \log(1+rx), & (r \geq 0) \end{cases}, \tag{4}$$

where $\rho = 1$ or $r = 0$, namely, $G(x) = x$, yielding the proportional hazards model, whereas $\rho = 0$ or $r = 1$, namely, $G(x) = \log(1+x)$, yielding the proportional odds model. In our simulation examples we only consider four special cases: $G_\rho(x;1) = x$, $G_r(x;1) = \log(1+x)$, $G_\rho(x;0.5) = 2((1+x)^{1/2} - 1)$ with $\rho = 0.5$, and $G_r(x;0.5) = 2\log(1+x/2)$ with $r = 0.5$.

**Example 1.** *In this example we consider the case where the two event types share the same relevant predictors. The transformation model has a cumulative hazard function of the form*

$$\Lambda_{ik}(t|w_i, \mathbf{X}_i) = G_k\{\Lambda_k(t)\exp(\beta_{k0} + \beta_{k1}X_{i1} + \beta_{k2}X_{i2} + \beta_{k3}X_{i3} + \beta_{k4}X_{i4} + \beta_{k5}X_{i5} + \beta_{k6}X_{i6})\}w_i, \quad i = 1, \ldots, n;\ k = 1, 2, \tag{5}$$

*where the covariates* $\mathbf{X}_i^* = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6})^\top$, *in which* $X_{i1}$ *and* $X_{i3}$ *are generated from a uniform distribution on* $(0, 1)$, $X_{i2}$ *and* $X_{i5}$ *are generated from a Bernoulli distribution with success probability* $P = 0.4$, *and* $X_{i4}$ *and* $X_{i6}$ *are generated from a standard normal distribution. Furthermore* $w_i$ *is generated from a gamma* $(1/\theta, \theta)$ *distribution, and* $F_k(t) = (1 - e^{-t})/(1 - e^{-3})I(0 \leq t \leq 3) + I(t > 3)$. *We simulate 500 replicated data sets with sample size* $n = 200$ *and 400. The true population values of the parameters are set as* $\boldsymbol{\beta}_1^* = (\beta_{10}, \beta_{11}, \ldots, \beta_{16}) = (-1.5, 1.5, 2.0, 0, 0, 0, 0)$, $\boldsymbol{\beta}_2^* = (\beta_{20}, \beta_{21}, \ldots, \beta_{26}) = (2.0, -1.5, -2.0, 0, 0, 0, 0)$, *and* $\theta = 1.0$. *Hence the true relevant covariates for the two event types are all* $\{X_1, X_2\}$. *The censoring times of both event types are generated from a mixture distribution with probability 0.5 from a uniform distribution on* $(3/2, 3)$ *and probability 0.5 from a point mass at* $\tau = 3$.

We follow the idea of Fan & Li (2002) to propose the generalized cross-validation method to select the tuning parameter $\eta_n$, which is data driven. On the basis of the 500 replicated data sets, the average number of zero coefficients is summarized in Table 1, in which the column labelled "C" presents the average number restricted only to the true zero coefficients, the column labelled "In" reports the average number of the nonzero coefficients that are erroneously set to zero, and the column labelled "Combination" presents the average number corresponding to the whole model that combines the two event types. As expected, the median of RME is slightly larger for our proposed method than for the MLE in the oracle situation (Oracle) where the true model, $\Lambda_{ik}(t|w_i, X_{i1}, X_{i2}) = G_k\{\Lambda_k(t)\exp(\beta_{k0} + \beta_{k1}X_{i1} + \beta_{k2}X_{i2})\}w_i$, is used. However the difference becomes negligible as the sample size increases. Our method can correctly identify the significant variables for each event type and is comparable to the oracle MLE proposed by Zeng, Chen, & Ibrahim (2009) in terms of model error under the given sample sizes. We therefore conclude that our PMLE performs as well as the oracle procedure in reducing model error and model complexity when the sample size is large enough.

We next compare three methods, including our proposed PMLE, nonparametric MLE, and nonparametric MLE in the oracle situation (Oracle). On the basis of the 500 replicated data sets, the average of censored proportions for the two event types are about 50–60%, which may vary across different sample sizes and transformation functions. In the comparison, two sample sizes, $n = 200$ and $n = 400$, and four combinations of $\rho$ and $r$ (see Eq. (4)), denoted by $G_\rho(x;1)$, $G_r(x;1)$, $G_\rho(x;0.5)$, and $G_r(x;0.5)$, are considered. Tables 2 and 3 report simulation results, including the estimates of the nonzero parameters, the value of $\Lambda_k(t)$ at $\tau/4$, and the frailty variance $\theta$, $\{\beta_{11}, \beta_{12}, \Lambda_1(\tau/4), \beta_{21}, \beta_{22}, \Lambda_2(\tau/4), \theta\}$, as well as the magnitudes of the bias (Bias),

TABLE 2: Estimation of nonzero parameters for Example 1 ($n = 200$).

| | | | PMLE | | | | MLE | | | | Oracle | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tran | Par | True | Bias | SE | SD | CP | Bias | SE | SD | CP | Bias | SE | SD | CP |
| | | | | | | | $n = 200$ | | | | | | | |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | −0.006 | 0.489 | 0.533 | 0.938 | 0.024 | 0.503 | 0.530 | 0.934 | 0.000 | 0.494 | 0.510 | 0.942 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.015 | 0.316 | 0.331 | 0.940 | 0.016 | 0.320 | 0.342 | 0.934 | −0.015 | 0.314 | 0.331 | 0.938 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.003 | 0.070 | 0.071 | 0.950 | −0.002 | 0.070 | 0.071 | 0.946 | 0.003 | 0.070 | 0.071 | 0.944 |
| $G_{\rho 2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | 0.017 | 0.474 | 0.496 | 0.946 | −0.010 | 0.484 | 0.498 | 0.938 | 0.017 | 0.475 | 0.484 | 0.954 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.012 | 0.330 | 0.327 | 0.958 | −0.045 | 0.334 | 0.343 | 0.948 | −0.012 | 0.328 | 0.326 | 0.958 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.007 | 0.071 | 0.070 | 0.938 | 0.002 | 0.071 | 0.071 | 0.928 | 0.007 | 0.071 | 0.070 | 0.934 |
| | $\theta$ | 1.000 | −0.026 | 0.259 | 0.254 | 0.966 | −0.024 | 0.258 | 0.268 | 0.960 | −0.025 | 0.255 | 0.254 | 0.966 |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | −0.004 | 0.493 | 0.513 | 0.948 | 0.019 | 0.503 | 0.527 | 0.934 | −0.003 | 0.494 | 0.507 | 0.946 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.015 | 0.321 | 0.338 | 0.928 | 0.016 | 0.325 | 0.349 | 0.928 | −0.015 | 0.319 | 0.339 | 0.928 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.004 | 0.070 | 0.071 | 0.946 | −0.001 | 0.071 | 0.071 | 0.944 | 0.004 | 0.070 | 0.071 | 0.948 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | 0.027 | 0.615 | 0.659 | 0.934 | −0.021 | 0.650 | 0.674 | 0.944 | 0.021 | 0.632 | 0.646 | 0.948 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.004 | 0.419 | 0.408 | 0.942 | −0.052 | 0.428 | 0.435 | 0.948 | −0.004 | 0.416 | 0.408 | 0.944 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.014 | 0.084 | 0.082 | 0.936 | 0.007 | 0.084 | 0.084 | 0.934 | 0.014 | 0.084 | 0.082 | 0.930 |
| | $\theta$ | 1.000 | −0.033 | 0.274 | 0.277 | 0.958 | −0.030 | 0.274 | 0.288 | 0.952 | −0.033 | 0.270 | 0.277 | 0.958 |
| $G_{\rho 1}(x; 0.5)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.002 | 0.557 | 0.600 | 0.940 | 0.032 | 0.576 | 0.618 | 0.934 | 0.003 | 0.564 | 0.592 | 0.938 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.013 | 0.360 | 0.375 | 0.942 | 0.024 | 0.366 | 0.387 | 0.940 | −0.013 | 0.358 | 0.375 | 0.942 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.006 | 0.075 | 0.074 | 0.944 | 0.000 | 0.075 | 0.075 | 0.942 | 0.006 | 0.075 | 0.074 | 0.938 |
| $G_{r2}(x; 0.5)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | 0.022 | 0.557 | 0.599 | 0.944 | −0.018 | 0.581 | 0.603 | 0.944 | 0.016 | 0.567 | 0.582 | 0.944 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.011 | 0.384 | 0.386 | 0.944 | −0.054 | 0.391 | 0.407 | 0.942 | −0.011 | 0.381 | 0.385 | 0.944 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.006 | 0.080 | 0.074 | 0.956 | −0.001 | 0.080 | 0.076 | 0.954 | 0.006 | 0.080 | 0.074 | 0.954 |
| | $\theta$ | 1.000 | −0.030 | 0.271 | 0.282 | 0.946 | −0.025 | 0.271 | 0.290 | 0.940 | −0.030 | 0.267 | 0.282 | 0.942 |
| $G_{r1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.004 | 0.623 | 0.689 | 0.922 | 0.044 | 0.660 | 0.721 | 0.926 | 0.007 | 0.642 | 0.679 | 0.932 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.016 | 0.414 | 0.440 | 0.934 | 0.033 | 0.424 | 0.458 | 0.944 | −0.015 | 0.412 | 0.440 | 0.934 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.009 | 0.081 | 0.086 | 0.920 | 0.002 | 0.082 | 0.087 | 0.924 | 0.009 | 0.081 | 0.086 | 0.922 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | 0.024 | 0.616 | 0.654 | 0.942 | −0.023 | 0.651 | 0.673 | 0.946 | 0.021 | 0.632 | 0.646 | 0.952 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.004 | 0.422 | 0.413 | 0.944 | −0.054 | 0.432 | 0.440 | 0.940 | −0.004 | 0.420 | 0.412 | 0.942 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.014 | 0.084 | 0.083 | 0.928 | 0.007 | 0.085 | 0.085 | 0.936 | 0.014 | 0.084 | 0.083 | 0.934 |
| | $\theta$ | 1.000 | −0.036 | 0.289 | 0.290 | 0.946 | −0.028 | 0.290 | 0.303 | 0.946 | −0.035 | 0.286 | 0.290 | 0.946 |

the standard deviation (SD) of the estimated parameter, the average of the estimated standard errors (SE), and the coverage probability of 95% confidence intervals (CP). Several findings have been obtained. First the parameter estimators are unbiased, and the values of SD and SE are close to each other in all the cases. Second the CPs are very close to the nominal level 95%, ranging from 93 to 97% for all the cases. Third almost all the values of SD and SE of the proposed PMLE are smaller than those of the nonparametric MLE proposed by Zeng, Chen, & Ibrahim (2009), indicating that our PMLE can improve the efficiency of parameter estimators. Finally all the methods perform better when the sample size gets larger.

TABLE 3: Estimation of nonzero parameters for Example 1 ($n = 400$).

| | | | PMLE | | | | MLE | | | | Oracle | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tran | Par | True | Bias | SE | SD | CP | Bias | SE | SD | CP | Bias | SE | SD | CP |
| | | | | | | | $n = 400$ | | | | | | | |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.005 | 0.349 | 0.351 | 0.958 | 0.022 | 0.352 | 0.353 | 0.954 | 0.006 | 0.348 | 0.347 | 0.958 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | 0.005 | 0.223 | 0.223 | 0.962 | 0.023 | 0.224 | 0.228 | 0.962 | 0.005 | 0.222 | 0.223 | 0.962 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.001 | 0.050 | 0.049 | 0.956 | −0.002 | 0.050 | 0.049 | 0.962 | 0.001 | 0.050 | 0.049 | 0.964 |
| $G_{\rho 2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | −0.013 | 0.338 | 0.334 | 0.942 | −0.023 | 0.340 | 0.342 | 0.938 | −0.013 | 0.337 | 0.334 | 0.942 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.013 | 0.231 | 0.229 | 0.958 | −0.031 | 0.233 | 0.232 | 0.952 | −0.013 | 0.231 | 0.229 | 0.958 |
| | $\Lambda_2(\tau/4)$ | 0.555 | −0.002 | 0.050 | 0.052 | 0.934 | −0.005 | 0.051 | 0.052 | 0.940 | −0.002 | 0.050 | 0.052 | 0.930 |
| | $\theta$ | 1.000 | 0.003 | 0.182 | 0.180 | 0.960 | 0.007 | 0.182 | 0.184 | 0.956 | 0.003 | 0.180 | 0.180 | 0.960 |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.003 | 0.350 | 0.343 | 0.954 | 0.017 | 0.352 | 0.349 | 0.958 | 0.002 | 0.349 | 0.343 | 0.954 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | 0.005 | 0.226 | 0.227 | 0.962 | 0.022 | 0.228 | 0.232 | 0.964 | 0.005 | 0.225 | 0.227 | 0.962 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.001 | 0.050 | 0.049 | 0.958 | −0.002 | 0.050 | 0.050 | 0.952 | 0.001 | 0.050 | 0.049 | 0.962 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | −0.010 | 0.447 | 0.464 | 0.954 | −0.031 | 0.453 | 0.471 | 0.948 | −0.012 | 0.447 | 0.459 | 0.954 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.011 | 0.295 | 0.300 | 0.944 | −0.033 | 0.298 | 0.304 | 0.948 | −0.011 | 0.294 | 0.300 | 0.944 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.003 | 0.060 | 0.065 | 0.920 | −0.001 | 0.060 | 0.065 | 0.926 | 0.003 | 0.060 | 0.065 | 0.922 |
| | $\theta$ | 1.000 | −0.001 | 0.193 | 0.190 | 0.962 | 0.002 | 0.193 | 0.193 | 0.966 | −0.001 | 0.191 | 0.190 | 0.960 |
| $G_{\rho 1}(x; 0.5)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.007 | 0.398 | 0.398 | 0.944 | 0.027 | 0.402 | 0.401 | 0.946 | 0.008 | 0.398 | 0.395 | 0.944 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | 0.005 | 0.253 | 0.255 | 0.958 | 0.025 | 0.256 | 0.263 | 0.954 | 0.005 | 0.253 | 0.256 | 0.958 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.002 | 0.053 | 0.053 | 0.954 | −0.001 | 0.053 | 0.053 | 0.944 | 0.002 | 0.053 | 0.053 | 0.946 |
| $G_{r2}(x; 0.5)$ | $\boldsymbol{\beta}_{21}$ | 2.000 | 0.047 | 0.334 | 0.324 | 0.958 | 0.068 | 0.396 | 0.397 | 0.946 | 0.047 | 0.327 | 0.324 | 0.958 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.010 | 0.270 | 0.267 | 0.946 | −0.032 | 0.273 | 0.272 | 0.956 | −0.010 | 0.269 | 0.267 | 0.946 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.001 | 0.057 | 0.060 | 0.938 | −0.002 | 0.057 | 0.060 | 0.928 | 0.001 | 0.057 | 0.060 | 0.932 |
| | $\theta$ | 1.000 | −0.001 | 0.191 | 0.189 | 0.962 | 0.003 | 0.191 | 0.193 | 0.958 | −0.001 | 0.189 | 0.189 | 0.960 |
| $G_{r1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.012 | 0.451 | 0.454 | 0.946 | 0.037 | 0.457 | 0.460 | 0.944 | 0.013 | 0.451 | 0.450 | 0.946 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | 0.008 | 0.292 | 0.289 | 0.962 | 0.033 | 0.295 | 0.299 | 0.952 | 0.008 | 0.291 | 0.289 | 0.962 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.000 | 0.058 | 0.059 | 0.952 | −0.004 | 0.059 | 0.059 | 0.946 | 0.000 | 0.058 | 0.059 | 0.950 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{21}$ | −1.500 | −0.012 | 0.446 | 0.466 | 0.950 | −0.033 | 0.453 | 0.474 | 0.944 | −0.013 | 0.447 | 0.461 | 0.950 |
| | $\boldsymbol{\beta}_{22}$ | −2.000 | −0.009 | 0.297 | 0.299 | 0.950 | −0.032 | 0.300 | 0.304 | 0.948 | −0.009 | 0.296 | 0.299 | 0.948 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.004 | 0.060 | 0.065 | 0.930 | 0.000 | 0.061 | 0.065 | 0.926 | 0.004 | 0.060 | 0.065 | 0.924 |
| | $\theta$ | 1.000 | −0.006 | 0.203 | 0.196 | 0.962 | −0.002 | 0.203 | 0.199 | 0.966 | −0.006 | 0.202 | 0.196 | 0.962 |

**Example 2.** *In this example we consider the case where the two event types have different sets of relevant predictors. This case is more common in practice and is also shown in the real data analysis of Section 4. The transformation model has a cumulative hazard function of the form*

$$\Lambda_{ik}(t|w_i, \mathbf{X}_i) = G_k\left\{\Lambda_k(t)\exp(\beta_{k0} + \sum_{j=1}^{9}\beta_{kj}X_{ij})\right\}w_i, \quad i = 1, \ldots, n;\ k = 1, 2, \qquad (6)$$

*where the covariates* $\mathbf{X}_i^* = (X_{i1}, \ldots, X_{i9})^\top$*, in which* $X_{i1}$ *and* $X_{i3}$ *are generated from a uniform distribution on* $(0, 1)$*;* $X_{i2}$*,* $X_{i5}$*, and* $X_{i8}$ *are generated from a Bernoulli distribution*

TABLE 4: Relative model errors for Example 2 ($n = 400$).

| | RME | Zeros | | RME | Zeros | | RME | Zeros | |
|---|---|---|---|---|---|---|---|---|---|
| Method | Median | C | In | Median | C | In | Median | C | In |
| | $G_{\rho 1}(x; 1)$ | | | $G_{\rho 2}(x; 1)$ | | | Combination | | |
| PMLE | 0.442 (0.267) | 6.000 | 0.032 | 0.450 (0.226) | 6.000 | 0.000 | 0.910 (0.363) | 12.000 | 0.032 |
| Oracle | 0.441 (0.223) | 6.000 | 0.000 | 0.446 (0.226) | 6.000 | 0.000 | 0.906 (0.336) | 12.000 | 0.000 |
| | $G_{\rho 1}(x; 1)$ | | | $G_{r2}(x; 1)$ | | | Combination | | |
| PMLE | 0.460 (0.231) | 6.000 | 0.012 | 0.432 (0.221) | 6.000 | 0.000 | 0.935 (0.330) | 12.000 | 0.012 |
| Oracle | 0.458 (0.220) | 6.000 | 0.000 | 0.432 (0.221) | 6.000 | 0.000 | 0.934 (0.323) | 12.000 | 0.000 |
| | $G_{\rho 1}(x; 0.5)$ | | | $G_{r2}(x; 0.5)$ | | | Combination | | |
| PMLE | 0.444 (0.228) | 6.000 | 0.034 | 0.439 (0.224) | 6.000 | 0.000 | 0.891 (0.327) | 12.000 | 0.034 |
| Oracle | 0.432 (0.208) | 6.000 | 0.000 | 0.439 (0.224) | 6.000 | 0.000 | 0.886 (0.317) | 12.000 | 0.000 |
| | $G_{r1}(x; 1)$ | | | $G_{r2}(x; 1)$ | | | Combination | | |
| PMLE | 0.432 (0.222) | 6.000 | 0.054 | 0.429 (0.220) | 6.000 | 0.000 | 0.876 (0.315) | 12.000 | 0.054 |
| Oracle | 0.428 (0.206) | 6.000 | 0.000 | 0.429 (0.220) | 6.000 | 0.000 | 0.873 (0.303) | 12.000 | 0.000 |

*with success probability $P = 0.4$, and the others are generated from a standard normal distribution. Here, $w_i$, $F_k(t)$, and the censoring times are based on the same setup as Example 1. The true population values of parameters are $\boldsymbol{\beta}_1^* = (-1.5, 1.5, 2.0, 1.0, 0, 0, 0, 0, 0, 0)$, $\boldsymbol{\beta}_2^* = (2.0, 0, -1.5, -2.0, 1.0, 0, 0, 0, 0, 0)$, and $\theta = 0.5$. The simulation is based on 500 replications with sample size $n = 400$. Table 4 reports the median and standard deviation (in parentheses) of RME. Table 5 reports the estimates of parameters using PMLE, MLE, and oracle methods. The results presented in Tables 4 and 5 are very similar to those shown in Tables 1 and 2. Again the proposed method can correctly identify the significant variables and improve the efficiency of parameter estimators.*

## 4. ANALYSIS OF RISK FACTORS FOR COMPLICATIONS OF DIABETES

In this section we apply the proposed method to a case study based on the Hong Kong Diabetes Registry, which was established in 1995 as part of a continuous quality improvement program at the Prince of Wales Hospital in Hong Kong. A 10-year prospective cohort of 2,781 Chinese type 2 diabetic patients, who may experience three types of diabetic complications (stroke, heart failure, and kidney failure) was analyzed in this study. The main objective is to investigate the potential factors that may affect the development of the aforementioned diabetic complications. A total of 16 variables relevant to patients' baseline characteristics (Table 6), such as age, sex, duration of diabetes, obesity, blood pressure, glycemic control, lipid control, and renal function, were considered. The possible risk factors include age at enrolment (Age), age of diagnosis of type 2 diabetes (Age of diagnosis), sex (female or male), duration of diabetes (DMAGE) (i.e., Age—Age of diagnosis), waist circumference, body mass index (BMI), systolic blood pressure

TABLE 5: Estimation of nonzero parameters for Example 2 ($n = 400$).

| | | | PMLE | | | | MLE | | | | Oracle | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tran | Par | True | Bias | SE | SD | CP | Bias | SE | SD | CP | Bias | SE | SD | CP |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.023 | 0.280 | 0.289 | 0.948 | 0.039 | 0.287 | 0.301 | 0.934 | 0.023 | 0.278 | 0.289 | 0.948 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.006 | 0.189 | 0.186 | 0.950 | 0.015 | 0.190 | 0.190 | 0.942 | −0.006 | 0.188 | 0.186 | 0.954 |
| | $\boldsymbol{\beta}_{13}$ | 1.000 | −0.008 | 0.272 | 0.308 | 0.942 | 0.016 | 0.280 | 0.280 | 0.944 | 0.005 | 0.277 | 0.274 | 0.952 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.003 | 0.047 | 0.047 | 0.944 | −0.000 | 0.047 | 0.048 | 0.942 | 0.003 | 0.047 | 0.047 | 0.938 |
| $G_{\rho 2}(x; 1)$ | $\boldsymbol{\beta}_{22}$ | −1.500 | 0.002 | 0.269 | 0.258 | 0.952 | −0.028 | 0.273 | 0.268 | 0.954 | 0.003 | 0.266 | 0.258 | 0.950 |
| | $\boldsymbol{\beta}_{23}$ | −2.000 | 0.016 | 0.299 | 0.303 | 0.950 | 0.001 | 0.301 | 0.311 | 0.950 | 0.019 | 0.298 | 0.302 | 0.948 |
| | $\boldsymbol{\beta}_{24}$ | 1.000 | 0.004 | 0.098 | 0.092 | 0.966 | 0.014 | 0.100 | 0.096 | 0.962 | 0.004 | 0.097 | 0.092 | 0.962 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.005 | 0.048 | 0.048 | 0.944 | 0.002 | 0.048 | 0.048 | 0.950 | 0.005 | 0.048 | 0.048 | 0.950 |
| | $\theta$ | 0.500 | −0.009 | 0.113 | 0.102 | 0.966 | −0.011 | 0.113 | 0.106 | 0.964 | −0.009 | 0.112 | 0.102 | 0.966 |
| $G_{\rho 1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.023 | 0.282 | 0.293 | 0.946 | 0.040 | 0.288 | 0.303 | 0.940 | 0.023 | 0.280 | 0.293 | 0.940 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.005 | 0.193 | 0.188 | 0.956 | 0.018 | 0.194 | 0.192 | 0.956 | −0.004 | 0.192 | 0.188 | 0.956 |
| | $\boldsymbol{\beta}_{13}$ | 1.000 | 0.003 | 0.277 | 0.284 | 0.950 | 0.016 | 0.281 | 0.281 | 0.948 | 0.006 | 0.278 | 0.275 | 0.950 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.003 | 0.047 | 0.048 | 0.940 | −0.001 | 0.048 | 0.048 | 0.946 | 0.003 | 0.047 | 0.048 | 0.942 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{22}$ | −1.500 | −0.018 | 0.254 | 0.265 | 0.946 | −0.039 | 0.257 | 0.273 | 0.940 | −0.018 | 0.253 | 0.265 | 0.944 |
| | $\boldsymbol{\beta}_{23}$ | −2.000 | 0.019 | 0.413 | 0.418 | 0.958 | −0.009 | 0.418 | 0.436 | 0.944 | 0.019 | 0.412 | 0.417 | 0.958 |
| | $\boldsymbol{\beta}_{24}$ | 1.000 | 0.004 | 0.132 | 0.127 | 0.958 | 0.018 | 0.134 | 0.133 | 0.948 | 0.004 | 0.130 | 0.127 | 0.958 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.008 | 0.056 | 0.056 | 0.940 | 0.003 | 0.056 | 0.056 | 0.942 | 0.008 | 0.056 | 0.056 | 0.940 |
| | $\theta$ | 0.500 | −0.007 | 0.122 | 0.107 | 0.968 | −0.007 | 0.121 | 0.112 | 0.970 | −0.007 | 0.120 | 0.108 | 0.968 |
| $G_{\rho 1}(x; 0.5)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.020 | 0.330 | 0.347 | 0.946 | 0.038 | 0.338 | 0.358 | 0.940 | 0.021 | 0.328 | 0.347 | 0.942 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | −0.011 | 0.217 | 0.212 | 0.950 | 0.012 | 0.219 | 0.217 | 0.948 | −0.011 | 0.216 | 0.212 | 0.948 |
| | $\boldsymbol{\beta}_{13}$ | 1.000 | −0.006 | 0.320 | 0.351 | 0.942 | 0.016 | 0.331 | 0.336 | 0.944 | 0.004 | 0.327 | 0.327 | 0.956 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.003 | 0.050 | 0.049 | 0.946 | −0.001 | 0.050 | 0.050 | 0.946 | 0.003 | 0.050 | 0.049 | 0.946 |
| $G_{r2}(x; 0.5)$ | $\boldsymbol{\beta}_{22}$ | −1.500 | −0.012 | 0.227 | 0.234 | 0.938 | −0.031 | 0.229 | 0.241 | 0.938 | −0.012 | 0.226 | 0.235 | 0.938 |
| | $\boldsymbol{\beta}_{23}$ | −2.000 | 0.022 | 0.365 | 0.370 | 0.956 | −0.000 | 0.369 | 0.385 | 0.946 | 0.024 | 0.364 | 0.371 | 0.956 |
| | $\boldsymbol{\beta}_{24}$ | 1.000 | 0.003 | 0.118 | 0.113 | 0.964 | 0.016 | 0.120 | 0.117 | 0.964 | 0.003 | 0.117 | 0.113 | 0.962 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.008 | 0.053 | 0.052 | 0.950 | 0.004 | 0.053 | 0.052 | 0.954 | 0.008 | 0.053 | 0.052 | 0.950 |
| | $\theta$ | 0.500 | −0.008 | 0.120 | 0.107 | 0.968 | −0.008 | 0.120 | 0.112 | 0.962 | −0.008 | 0.119 | 0.107 | 0.968 |
| $G_{r1}(x; 1)$ | $\boldsymbol{\beta}_{11}$ | 1.500 | 0.025 | 0.392 | 0.408 | 0.948 | 0.048 | 0.402 | 0.416 | 0.948 | 0.026 | 0.391 | 0.406 | 0.948 |
| | $\boldsymbol{\beta}_{12}$ | 2.000 | 0.013 | 0.367 | 0.373 | 0.948 | 0.042 | 0.376 | 0.390 | 0.952 | 0.013 | 0.364 | 0.373 | 0.946 |
| | $\boldsymbol{\beta}_{13}$ | 1.000 | −0.005 | 0.373 | 0.412 | 0.930 | 0.019 | 0.394 | 0.405 | 0.952 | 0.006 | 0.388 | 0.391 | 0.952 |
| | $\Lambda_1(\tau/4)$ | 0.555 | 0.004 | 0.054 | 0.056 | 0.944 | −0.000 | 0.055 | 0.057 | 0.934 | 0.004 | 0.054 | 0.056 | 0.938 |
| $G_{r2}(x; 1)$ | $\boldsymbol{\beta}_{22}$ | −1.500 | −0.019 | 0.255 | 0.267 | 0.938 | −0.041 | 0.258 | 0.275 | 0.934 | −0.019 | 0.254 | 0.267 | 0.938 |
| | $\boldsymbol{\beta}_{23}$ | −2.000 | 0.016 | 0.414 | 0.418 | 0.958 | −0.012 | 0.419 | 0.437 | 0.948 | 0.017 | 0.413 | 0.418 | 0.958 |
| | $\boldsymbol{\beta}_{24}$ | 1.000 | 0.003 | 0.132 | 0.126 | 0.962 | 0.018 | 0.135 | 0.132 | 0.956 | 0.003 | 0.131 | 0.126 | 0.962 |
| | $\Lambda_2(\tau/4)$ | 0.555 | 0.008 | 0.056 | 0.056 | 0.944 | 0.003 | 0.056 | 0.057 | 0.936 | 0.008 | 0.056 | 0.056 | 0.936 |
| | $\theta$ | 0.500 | −0.006 | 0.131 | 0.118 | 0.956 | −0.004 | 0.131 | 0.122 | 0.944 | −0.006 | 0.130 | 0.118 | 0.956 |

TABLE 6: Baseline characteristics of 2,870 type 2 diabetic patients.

| Complications<br>Number of patients | Stroke<br>($n = 403$) | Heart failure<br>($n = 132$) | Kidney failure<br>($n = 679$) |
|---|---|---|---|
| Age (years) | | | |
| Age at enrolment | $61.17 \pm 10.96$ | $67.01 \pm 10.24$ | $61.19 \pm 11.30$ |
| Age of diagnosis | $52.58 \pm 12.13$ | $57.19 \pm 12.29$ | $52.39 \pm 12.23$ |
| Sex (female = 1) | 47.0% | 62.1% | 55.2% |
| DMAGE: duration of diabetes (years) | $8.59 \pm 7.01$ | $9.82 \pm 8.35$ | $8.79 \pm 7.13$ |
| Obesity | | | |
| Waist circumference (cm) | $86.77 \pm 8.87$ | $87.20 \pm 10.15$ | $86.90 \pm 9.93$ |
| BMI: body mass index (kg/m$^2$) | $25.07 \pm 3.48$ | $25.22 \pm 4.37$ | $25.13 \pm 3.95$ |
| Blood pressure (mmHg) | | | |
| SBP: systolic BP | $139.85 \pm 19.53$ | $146.01 \pm 21.57$ | $141.50 \pm 20.39$ |
| DBP: diastolic BP | $78.32 \pm 11.17$ | $77.95 \pm 13.17$ | $78.10 \pm 11.50$ |
| Glycemic control | | | |
| HbA1c (%) | $8.17 \pm 1.85$ | $8.22 \pm 2.03$ | $8.25 \pm 2.07$ |
| FBG: fasting glucose (mmol/l) | $9.42 \pm 3.60$ | $9.58 \pm 3.60$ | $9.41 \pm 3.82$ |
| Lipid control (mmol/l) | | | |
| TC: total cholesterol | $5.46 \pm 1.13$ | $5.57 \pm 1.17$ | $5.38 \pm 1.17$ |
| HDL: HDL-cholesterol | $1.23 \pm 0.31$ | $1.23 \pm 0.35$ | $1.27 \pm 0.36$ |
| LDL: LDL-cholesterol | $3.50 \pm 0.98$ | $3.54 \pm 1.05$ | $3.35 \pm 1.03$ |
| TG: triglyceride | $1.59 \pm 0.82$ | $1.71 \pm 0.83$ | $1.65 \pm 0.84$ |
| Renal function | | | |
| ACR (mg/mmol) | $39.89 \pm 108.42$ | $104.47 \pm 199.84$ | $66.52 \pm 142.20$ |
| eGFR (ml/min per 1.73 m$^2$) | $104.13 \pm 28.37$ | $95.39 \pm 30.31$ | $95.159 \pm 26.32$ |

Values are mean ± SD.

(SBP), diastolic blood pressure (DBP), glycated hemoglobin (HbA1c), fasting plasma glucose (FBG), total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglyceride (TG), urinary albumin creatinine (ACR), and estimated glomerular filtration rate (eGFR). All the measurements are continuous except for sex. In this article, 403 out of 2,871 (14%), 132 out of 2,871 (4.6%), and 679 out of 2,871 (23.7%) type 2 diabetic patients were diagnosed as having stroke, heart failure, and kidney failure, respectively. Sixty-three (2.20%) patients experienced both stroke and heart failure, 187 (6.52%) patients experienced both stroke and kidney failure, 102 (3.55%) patients experienced both heart failure and kidney failure, and 52 (1.81%) patients experienced three complications. We denote $T_1$ as the time to stroke, $T_2$ as the time to heart failure, and $T_3$ as the time to kidney failure. The failure time to each complication was calculated as the period from onset to the date of the first clinical endpoint of this complication or January 31, 2009, whichever came first. Thus, $(T_1, T_2, T_3)$ are censored at the last follow-up with censoring rates (86, 76.3, and 95.4%). We applied the proposed frailty transformation model to conduct the analysis. Following the idea of Zeng, Chen, & Ibrahim (2009) we calculated the likelihood function at a number of equally spaced grid points of $(\rho, r)$ under three event types (Eq. 4). The transformations were chosen as $G_1(x) = G_2(x) = G_3(x) = \log(1 + x)$ to maximize the likelihood function. Given that a high correlation exists between Age at diagnosis and Age

TABLE 7: Parameter estimates in the analysis of diabetes data.

| | Time to stroke | | Time to heart failure | | Time to kidney failure | |
|---|---|---|---|---|---|---|
| Covariates | MLE | PMLE | MLE | PMLE | MLE | PMLE |
| Constant | 1.920 (1.486) | 1.788 (0.816) | −1.372 (0.655) | −1.533 (0.609) | 4.779 (0.398) | 4.457 (1.037) |
| Sex | –0.641 (0.137) | –0.633 (0.133) | 0.160 (0.223) | – | –0.216 (0.126) | – |
| Age | 0.642 (0.078) | 0.585 (0.073) | 1.678 (0.193) | 1.664 (0.192) | 0.903 (0.081) | 0.871 (0.078) |
| BMI | 0.110 (0.077) | – | 0.204 (0.123) | – | 0.111 (0.069) | 0.118 (0.065) |
| DBP | 0.127 (0.068) | 0.137 (0.065) | 0.131 (0.105) | – | 0.106 (0.062) | 0.098 (0.040) |
| HBA1C | 0.075 (0.071) | – | 0.235 (0.140) | 0.233 (0.135) | 0.188 (0.063) | 0.176 (0.061) |
| TC | 0.156 (0.080) | 0.178 (0.068) | 0.099 (0.156) | – | −0.246 (0.074) | – |
| HDL | −0.278 (0.086) | −0.295 (0.074) | −0.277 (0.138) | −0.289 (0.117) | 0.087 (0.074) | −0.109 (0.055) |
| TG | −0.035 (0.084) | – | 0.079 (0.135) | – | 0.336 (0.076) | – |
| Log-ACR | 0.202 (0.069) | 0.179 (0.063) | 0.873 (0.142) | 0.923 (0.131) | 0.872 (0.064) | 0.880 (0.061) |
| Age×HBA1C | −0.010 (0.071) | – | −0.150 (0.122) | – | −0.133 (0.064) | −0.119 (0.062) |
| Age×TC | −0.004 (0.075) | – | −0.047 (0.129) | – | 0.027 (0.066) | – |
| Age×Log-ACR | −0.103 (0.068) | – | −0.436 (0.119) | −0.431 (0.110) | −0.388 (0.064) | −0.361 (0.059) |
| BMI×DBP | −0.009 (0.069) | – | −0.092 (0.107) | – | −0.039 (0.056) | – |
| BMI × TC | 0.018 (0.067) | – | −0.104 (0.101) | – | −0.044 (0.063) | – |
| DBP×HBA1C | 0.006 (0.064) | – | −0.063 (0.101) | – | −0.021 (0.057) | – |
| TC×TG | 0.156 (0.062) | 0.121 (0.056) | 0.016 (0.089) | – | 0.052 (0.056) | – |
| $\theta$ | 1.144 (0.027) | 1.022 (0.024) | | | | |

(the correlated coefficient is 0.875), TC and LDL (0.924), Waist and BMI (0.831), DBP and SBP (0.6327), HBA1c and FBG (0.690), as well as eGFR and Age (0.543), we excluded the redundant variables Age at diagnosis, LDL, Waist, SBP, FBG, and eGFR to avoid multicollinearity. We further took a logarithmic transformation of ACR to reduce high nonnormality and standardized the continuous variables before the analysis. We also considered pairwise interactions between the remaining variables. After removing those of the highly correlated variables (where correction coefficients were above 0.20), seven interaction terms were selected for consideration.

## 4.1. Assessment of Risk Factors for Three Complications

This section aims to identify risk factors for the three complications of type 2 diabetes, especially to examine the risk factors that have effects on only one or two of the complications, and the common risk factors that have effects on all the three complications.

Table 7 presents the important risk factors and their estimated coefficients with standard errors shown in parentheses. The implications of the identified risk factors for cardiovascular and renal complications are as follows. First men with type 2 diabetes are more likely to develop stroke than women. However the progressions of heart and kidney failures are not substantially different between sexes. Second age has significantly positive effects on the three endpoints. Older patients tend to have higher hazard rates of stroke, heart failure, and kidney failure. Third HbA1c, a measure primarily reflecting the average plasma glucose concentration, has significantly positive effects on the complications of heart and kidney failures. A higher level of HbA1c would have resulted in poorer control of blood glucose and thus higher risk of heart failure and kidney failure. Fourth lipid control, in particular an increase in HDL and a decrease in TC, is beneficial to the prevention of the three complications. Finally a poor profile of renal function (with high ACR) has bad effects on the three complications. An increase in ACR would significantly increase the risk of developing the three complications. We also find that a few interaction terms, including Age×HBA1C, Age×Log-ACR, and TC×TG, have significant effects on at least one of the complications, implying that a model with only the main effects of potential risk factors is not adequate for the prediction of multiple complications.

## 4.2. Comparisons of Predicted Survival Probabilities

In this section we use the proposed model for the three types of complication of diabetes to predict the survival distribution for three groups: young patients (age $\leq 45$, 759 patients), middle-aged patients ($45 <$ age $\leq 64$, 1,424 patients), and old patients (age $> 64$, 688 patients), given that the patients had experienced stroke by year $t_1$. From the proposed model we can easily obtain the survival probability of time to heart failure, $P(T_2 > t | T_1 = t_1, T_2 > t_1, \mathbf{X})$, which is given by

$$\frac{\int_w \exp(-w[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_2\{F_2(t)\exp(\alpha_2 + \boldsymbol{\beta}_2^{*\top}\mathbf{X}_2^*)\}])w dw}{\int_w \exp(-w[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_2\{F_2(t_1)\exp(\alpha_2 + \boldsymbol{\beta}_2^{\top}\mathbf{X}_2^*)\}])w dw}$$

$$= \left(\frac{1 + \theta[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_2\{F_2(t_1)\exp(\alpha_2 + \boldsymbol{\beta}_2^{*\top}\mathbf{X}_2^*)\}]}{1 + \theta[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_2\{F_2(t)\exp(\alpha_2 + \boldsymbol{\beta}_2^{*\top}\mathbf{X}_2^*)\}]}\right)^{\frac{1}{\theta}+1},$$

and the survival probability of time to kidney failure, $P(T_3 > t | T_1 = t_1, T_3 > t_1, \mathbf{X})$, which is given by

$$\left(\frac{1 + \theta[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_3\{F_3(t_1)\exp(\alpha_3 + \boldsymbol{\beta}_3^{*\top}\mathbf{X}_3^*)\}]}{1 + \theta[G_1\{F_1(t_1)\exp(\alpha_1 + \boldsymbol{\beta}_1^{*\top}\mathbf{X}_1^*)\} + G_3\{F_3(t)\exp(\alpha_3 + \boldsymbol{\beta}_3^{*\top}\mathbf{X}_3^*)\}]}\right)^{\frac{1}{\theta}+1},$$

where $\mathbf{X}_k^*$ and $\boldsymbol{\beta}_k^*$, $k = 1, 2, 3$, are the selected risk factors and the corresponding estimated coefficients displayed in Table 7. To predict the survival probabilities of time to heart failure and time to kidney failure for a patient from one group, we can first estimate the above conditional probabilities for each $\mathbf{X}_k^*$ in this group by substituting the estimated coefficients into the conditional probabilities. We then take the average over all the patients in this group.

Figure 1a and b displays the survival probabilities of time to heart failure and time to kidney failure for four groups of patients stratified by age, given that patients experienced a stroke at 7 years of follow-up. The figures show that older patients have a higher risk of developing heart and kidney failures than young patients. Moreover kidney failure is a more risky complication for type 2 diabetic patients than heart failure, given that the patients in the same group experienced a stroke at 7 years of follow-up.

## 5. CONCLUDING REMARKS

This article considered the variable selection problem for gamma frailty transformation models with multivariate failure time data. In this article we adopted the penalized nonparametric maximum likelihood estimation procedure with nonconcave penalty function proposed by Fan & Li (2002). The proposed method can be regarded as a generalization of Zeng, Chen, & Ibrahim (2009). A nonconcave penalty function has the advantage of selecting covariate variables and estimating coefficients simultaneously. We have shown that the proposed procedure performs as well as the oracle one in which the true model is assumed to be known. Simulation studies demonstrate that the proposed estimation procedure provides asymptotically efficient estimates and yields good inferential properties for finite sample sizes.

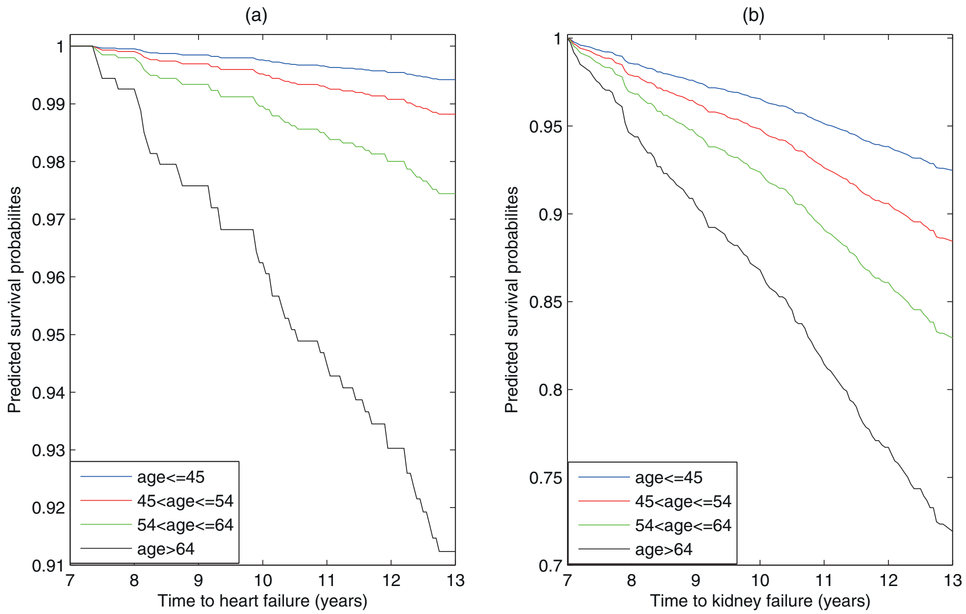


FIGURE 1: (a) Predicted survival probabilities of time to heart failure and (b) predicted survival probabilities of time to kidney failure for lower ages (age $\leq$ 45), medium ages (45 < age $\leq$ 54 or 54 < age $\leq$ 64), and higher ages (age > 64), given that the patient experiences a stroke event at year $t_1 = 7$ of follow-up.

Due to the rapid development of modern computation and massive growth of data in substantive research, dealing with ultrahigh-dimensional data is of great importance. However directly applying our method to handle such kinds of data is difficult. Many statistical methods, such as LARS (Efron et al., 2004) with LASSO penalty and coordinate descent algorithm (Breheny & Huang, 2011), have been developed to analyze ultrahigh-dimensional data in the context of linear models, generalized linear models, and Cox models. To the best of our knowledge, no existing research has ever considered variable selection for gamma frailty transformation models with high-dimensional data. This important problem may be a subject of future research. Meanwhile as the Associate Editor pointed out, the proposed methodology has a limitation when it comes to left-truncated data, as displayed in the real data. Our proposed method may be extended to left-truncated and right-censored data in future work. Recently there has been much work on the analysis of right-censored length-biased data which incorporates left-truncation and right-censored data as a special example. The related literature include, but is not limited to, for example, Ning et al. (2009), Shen et al. (2009), Tsai (2009), Qin & Shen (2010), and Chen & Zhou (2012). Further research will extend the proposed approach to deal with right-censored length-biased data. The real data analysis of this situation can also be improved in further work.

## APPENDIX

In this section we provide the computational algorithm and the estimated covariance matrix of the parameter estimator. Inspired by Zeng, Chen, & Ibrahim (2009) we employ a EM algorithm to obtain the nonparametric penalized maximum likelihood estimator. Because the penalty function is nondifferentiable we use the MM algorithm to construct a function to approximate the penalty function. Fan & Li (2001) suggested a local quadratic approximation (LQA) for the penalty function, which is a special case of the MM algorithm.

We use a discrete distribution to approximate the distribution $F$ as in Zeng, Chen, & Ibrahim (2009). To avoid the difficulty caused by the other nuisance parameter $\boldsymbol{\alpha}$ and frailty parameter $\theta$ we employ profile likelihood in the use of the Newton–Raphson algorithm. In the $(k+1)$th iteration, given $\boldsymbol{\beta}^{(k)}$, we can obtain the maxima $(\boldsymbol{\alpha}^{(k)}, \theta^{(k)})$ by maximizing $p\ell_n(\boldsymbol{\alpha}, \theta, \boldsymbol{\beta}^{(k)})$ with respect to $\boldsymbol{\alpha}$ and $\theta$, and then for fixed $(\boldsymbol{\alpha}^{(k+1)}, \theta^{(k+1)})$, we maximize $Q_n(\boldsymbol{\alpha}^{(k+1)}, \theta^{(k+1)}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ to obtain $\boldsymbol{\beta}^{(k+1)}$.

Let $E_k = \mathrm{diag}(\mathbf{0}_1, B_k)$, where $B_k$ denotes the diagonal matrix with $(j, j)$th entry $p'_{\eta_n}(|\tilde{\beta}_j^{(k)}|+)/(\varepsilon + |\tilde{\beta}_j^{(k)}|)$ and $(\tilde{\beta}_1^{(k)}, \ldots, \tilde{\beta}_s^{(k)})^\top = \boldsymbol{\beta}_{d1}^{(k)}$, and $\mathbf{0}_1$ denotes the $d \times d$ matrix with all zero elements as before. For the gamma frailty transformation model, the solution in the Newton–Raphson algorithm is updated by

$$\xi^{(k+1)} = \xi^{(k)} - \nu_k[I_0(\xi^{(k)}) - nE_k]^{-1}[\ell_{n0}(\xi^{(k)}) - nE_k\xi^{(k)}], \tag{7}$$

where $\nu_k$ is some positive scalar, and

$$\ell_{n0}(\xi^{(k)}) = \sum_{i=1}^{n} \ell_0^{(k)}\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\},$$

in which $\ell_0^{(k)}\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\}$ denotes the *efficient score function for* $\xi^{(k)}$.

Let $\hat{\xi}$ and $\hat{F}_k$ be the estimators of $\xi$ and $F_k$, $k = 1, \ldots, K$, respectively. The efficient score $\ell_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\}$ is the partial derivative of likelihood $\ell_n(\xi, F)$ with respect to $\xi$. Thus, it can be estimated by

$$\begin{aligned}
&\hat{\ell}_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\} \\
&= \frac{\partial \ell_n(\hat{\xi}, \hat{F})}{\partial \xi} \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} \Delta_{ik}\left[\log \frac{G_k''\{\hat{F}_k(Y_{ik})\exp(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{X}_i)\}}{G_k'\{\hat{F}_k(Y_{ik})\exp(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{X}_i)\}}\hat{F}_k(Y_{ik})\exp(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{X}_i) + 1\right](1, 0, X_i^\top)^\top \\
&\quad - \sum_{i=1}^{n}\hat{E}(w_i)\sum_{k=1}^{K} G_k'\{\hat{F}_k(Y_{ik})\exp(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{X}_i)\}(1, 0, X_i^\top)^\top \\
&\quad - \left[\frac{n}{\hat{\theta}^2}(\log\hat{\theta} - 1) + n\frac{\Gamma'(1/\hat{\theta})}{\hat{\theta}^2\Gamma(1/\hat{\theta})} - \frac{1}{\hat{\theta}^2}\sum_{i=1}^{n}\hat{E}(\log w_i) + \frac{1}{\hat{\theta}^2}\sum_{i=1}^{n}\hat{E}(w_i)\right](0, 1, \mathbf{0}^\top)^\top,
\end{aligned}$$

where $\hat{E}(w_i) = \hat{a}_i/\hat{b}_i$, $\hat{E}(\log w_i) = \psi(\hat{a}_i) + \log \hat{b}_i$, $\hat{a}_i = \hat{\theta}^{-1} + \sum_{k=1}^{K}\Delta_{ik}$, $\hat{b}_i = \hat{\theta}^{-1} + \sum_{k=1}^{K} G_k\{\hat{F}_k(Y_{ik})\exp(\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{X}_i)\}$, and $\psi(x) = \frac{d}{dx}\log\Gamma(x)$.

The efficient Fisher information matrix $I_0(\xi)$ is the covariance matrix of the efficient score $\ell_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\}$. Thus, it can be estimated by plugging in the estimator $\hat{\ell}_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\}$ to the covariance matrix, that is,

$$\hat{I}_0 = \widehat{\mathrm{cov}}(\hat{\ell}_{n0}(\hat{\xi})) = \frac{1}{n}\sum_{i=1}^{n}[\hat{\ell}_{n0}(\hat{\xi})][\hat{\ell}_{n0}(\hat{\xi})]^\top - \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\ell}_{n0}(\hat{\xi})\right]\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\ell}_{n0}(\hat{\xi})\right]^\top,$$

where $\hat{\ell}_{n0}(\hat{\xi}) = \sum_{i=1}^{n} \hat{\ell}_0\{(Y_{i1}, \Delta_{i1}, \mathbf{X}_{i1}), \ldots, (Y_{iK}, \Delta_{iK}, \mathbf{X}_{iK})\}$. The corresponding sandwich formula for estimating the covariance of the estimator of parameter $\hat{\xi}$ is

$$\widehat{\text{cov}}(\hat{\xi}) = [\hat{I}_0(\hat{\xi}) - n\hat{E}_\eta(\hat{\xi})]^{-1}\hat{I}_0(\hat{\xi})[\hat{I}_0(\hat{\xi}) - n\hat{E}_\eta(\hat{\xi})]^{-1},$$

where $\hat{E}_\eta(\hat{\xi})$ is $E_k$ with tuning parameter $\eta$ and $\boldsymbol{\beta}_{d1}^{(k)}$ replaced by $\hat{\boldsymbol{\beta}}_{d1}$, $\hat{I}_0(\hat{\xi}) = \widehat{\text{cov}}(\ell_{n0}(\hat{\xi}))$ and

$$\widehat{\text{cov}}(\ell_{n0}(\hat{\xi})) = \frac{1}{n}\sum_{i=1}^{n}[\ell_{n0}(\hat{\xi})][\ell_{n0}(\hat{\xi})]^\top - \left[\frac{1}{n}\sum_{i=1}^{n}\ell_{n0}(\hat{\xi})\right]\left[\frac{1}{n}\sum_{i=1}^{n}\ell_{n0}(\hat{\xi})\right]^\top.$$

Define $e(\eta) = \text{tr}[(\hat{I}_0(\hat{\xi}) - \hat{E}_\eta(\hat{\xi}))^{-1}\hat{I}_0(\hat{\xi})]$, where $tr(A)$ denotes the trace of matrix $A$. We can select the threshold parameter $\eta$ through the generalized cross-validation method similar to that proposed by Fan & Li (2002):

$$\text{GCV}(\eta) = \frac{p\ell_n(\hat{\xi})}{n\{1 - e(\eta)/n\}^2},$$

and $\hat{\eta} = \text{argmin}_\eta\{\text{GCV}(\eta)\}$ is selected.

## ACKNOWLEDGEMENTS

The authors wish to express their thanks to the Prince of Wales Hospital of Chinese University of Hong Kong for providing the data in the real example, Donglin Zeng, the Editor, the Associate Editor, and two anonymous referees for helpful suggestions. Xinyuan Song's research was supported by the National Natural Science Foundation of China and grants GRF from Hong Kong Special Administration Region. Shangyu Xie's research is supported by the National Natural Science Foundation of China and the State Key Program of National Natural Science Foundation of China, and the Fundamental Research Funds for the Central Universities in UIBE. Yong Zhou's research was supported by the National Natural Science Foundation of China, the State Key Program of National Natural Science Foundation of China, Key Laboratory of RCSDS, AMSS, CAS and Shanghai First-class Discipline A and IRTSHUFE, PCSIRT.

## BIBLIOGRAPHY

American Diabetes Association (ADA). (2008). Complications of diabetes in the United States. http://www.diabetes.org/xdiabetesstatistics/complications.jsp.

Androulakis, E., Koukouvinos, C., & Vonta, F. (2012). Estimation and variable selection via frailty models with penalized likelihood. *Statistics in Madicine*, 31, 2223–2239.

Asgharian, M. (2014). On the singularities of the information matrix and multipath change-point problems. *Theory of Probability and Its Application*, 58, 546–561.

Bhulai, M. A. J., Boomsma, D. I., Ligthart, R. S. L., Posthuma, D., & van der Vaart, A. W. (2009). Gamma frailty model for linkage analysis with application to interval-censored migraine data. *Biometrika*, 10, 187–200.

Breheny, P. & Huang, J. (2011). Coordinate descent algorithm for nonconvex penalized regression, with application to biological feature selection. *Annals of Applied Statistics*, 5, 232–253.

Cai, J., Fan, J., Li, R., & Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, 92, 303–316.

Centers for Disease Control and Prevention. (2008). National diabetes fact sheet. http://www.cdc.gov/diabetes/pubs/factsheet07

Chan, J. C., Malik, V., Jia, W., Kadowaki, T., Yajnik, C. S., Yoon, K. H., & Hu, F. B. (2009). Diabetes in Asia: Epidemiology, risk factors, and pathophysiology. *The Journal of the American Medical Association*, 301, 2129–2140.

Chen, X. & Zhou, Y. (2012). Quantile regression for right-censored and length-biased data. *Acta Mathematicae Applicatae Sinica, English Series*, 28, 443–462.

Clayton, D. G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A*, 148, 82–117.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, Series B*, 34, 187–220.

Effron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32, 407–499.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360.

Fan, J. & Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics*, 30, 74–99.

Frank, I. E. & Friedman, J. H. (1993). A Statistical view of some chemometrics regression tools. *Technometrics*, 35, 109–148.

Huang, J., Ma, S., Xie, H., & Zhang, C. (2009). A group bridge approach for variable selection. *Biometrika*, 96, 339–355.

Hunter, D. R. & Li, R. (2005). Variable selection using MM algorithms. *Annals of Statistics*, 33, 1617–1642.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, Hoboken, New Jersey.

Liu, X. & Zeng, D. (2013). Variable selection in semiparametric transformation models for right-censored data. *Biometrika*, 100, 859–876.

Murphy, S. A. & van der Vaart, A. W. (2000). On the profile likelihood. *Journal of the American Statistical Association*, 95, 449–465.

Nielsen, G. G., Gill, R. D., Andersen, P. K., & Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19, 25–43.

Ning, J., Qin, J., & Shen, Y. (2011). Buckley–James-type estimator with right-censored and length-biased data. *Biometrics*, 67, 1369–1378.

Oakes, D. (1991). Frailty models for multiple event times. In *Survival Analysis: State of the Art*, Klein, J. P. & Goel, P. K., editors. Kluwer, Dordrecht, pp. 371–379.

Qin, J. & Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*, 66, 382–392.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Tsai, W. Y. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika*, 96, 601–615.

Wang, S., Nan, B., Zhou, N., & Zhu, J. (2009). Hierarchically penalized Cox regression with grouped variables. *Biometrika*, 96, 307–322.

Wienke, A., Lichtenstein, P., & Yashin, A. I. (2003). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, 59, 1178–1183.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68, 1–49.

Zeng, D., Chen, Q., & Ibrahim, J. (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika*, 96, 277–291.

Zeng, D. & Lin, D. Y. (2007). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, 102, 167–180.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67, 301–320.

---

*Received 6 July 2014*
*Accepted 4 June 2016*