Biometrics 61, 208–216
March 2005

A Class of Bayesian Shared Gamma Frailty Models with Multivariate Failure Time Data

Guosheng Yin

Department of Biostatistics, M. D. Anderson Cancer Center, The University of Texas, Houston, Texas 77030, U.S.A.
email: gsyin@mdanderson.org

and

Joseph G. Ibrahim

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

Summary. For multivariate failure time data, we propose a new class of shared gamma frailty models by imposing the Box–Cox transformation on the hazard function, and the product of the baseline hazard and the frailty. This novel class of models allows for a very broad range of shapes and relationships between the hazard and baseline hazard functions. It includes the well-known Cox gamma frailty model and a new additive gamma frailty model as two special cases. Due to the nonnegative hazard constraint, this shared gamma frailty model is computationally challenging in the Bayesian paradigm. The joint priors are constructed through a conditional–marginal specification, in which the conditional distribution is univariate, and it absorbs the nonlinear parameter constraints. The marginal part of the prior specification is free of constraints. The prior distributions allow us to easily compute the full conditionals needed for Gibbs sampling, while incorporating the constraints. This class of shared gamma frailty models is illustrated with a real dataset.

Key words: Additive hazards; Bayesian inference; Box–Cox transformation; Constrained parameter; Correlated survival data; Frailty model; Gibbs sampling.

1. Introduction

1.1 *Cox Frailty Model*

Multivariate failure time data often arise in biomedical research due to natural or artificial clustering effects. For example, in family-based studies of genetic diseases, dental research, or litter-matched mice experiments, failure times in the same cluster may be correlated. The intracluster correlation needs to be taken into account to ensure the validity of estimation and inference. A well-known example is from the Diabetic Retinopathy Study (DRS; Diabetic Retinopathy Study Research Group, 1985), which was conducted to assess the effectiveness of laser photocoagulation in delaying severe visual loss (blindness) among patients with diabetic retinopathy. For each patient, one eye was randomly selected to receive the laser treatment, while the other eye was used as a control, for ethical reasons. The failure time of interest is the time to blindness, as measured by visual acuity less than 5/200. Clearly, the survival times from the two eyes on the same patient may not be independent.

A common approach to accommodating the intraclass correlation is to incorporate an unobserved random effect, or a frailty, into the Cox (1972) proportional hazards model. Let $\mathbf{Z}_{ik}(t)$ $(i = 1, \ldots, n;\ k = 1, \ldots, K_i)$ be a possibly external time-dependent covariate vector for the kth subject in the ith cluster. The usual Cox frailty model is defined as

$$\lambda(t \mid \mathbf{Z}_{ik}, W_i) = \lambda_0(t) W_i \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ik}(t)\}, \tag{1}$$

where $\lambda_0(t)$ is the unknown and unspecified baseline hazard function, $\boldsymbol{\beta}$ is the unknown $p \times 1$ regression coefficient vector, and W_i is the unobservable frailty induced by cluster i. Conditional on the W_i's, the failure times are assumed to be independent. A common parametric distribution for W_i is a gamma distribution with mean 1 (Clayton, 1978; Clayton and Cuzick, 1985). A widely used alternative is to assume that W_i has a positive stable distribution (Hougaard, 2000), which preserves the proportional hazards structure unconditionally (after integrating W_i out). Semiparametric Bayesian methods for multivariate failure time data have been studied extensively in various contexts (Clayton, 1991; Sinha, 1993; Sahu et al., 1997; Aslanidou, Dey, and Sinha, 1998; Sargent, 1998; Qiou, Ravishanker, and Dey, 1999; among others).

1.2 *Additive Frailty Model*

The proportional hazards formulation might not be a valid modeling structure in many situations. The underlying true relation of the hazards could be parallel as opposed to

proportional. By imposing a linear relationship of the covariate to the hazard function, Lin and Ying (1994) proposed the additive hazards model for independent failure time data. Here, we propose a new additive frailty model where an unobservable frailty has a multiplicative effect on the baseline hazard,

$$\lambda(t \mid \mathbf{Z}_{ik}, W_i) = \lambda_0(t)W_i + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t). \tag{2}$$

This model is particularly interesting because it assumes that, given the frailty, the hazard for each survival time follows an additive hazards model. Note that in (1) and (2), the frailty W_i is always linked to $\lambda_0(t)$ in a multiplicative fashion, which indicates that the baseline hazard function for cluster i becomes $\lambda_0(t)\,W_i$ due to the heterogeneity arising from the ith clustering effect. The multiplicative form of $\lambda_0(t)\,W_i$ also brings in mathematical convenience for parameter estimation and inference. One main difficulty arising in (2) is the nonnegative hazard constraint, i.e., $\lambda_0(t)\,W_i + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t) \geq 0$ for all i, k, and t. As will be seen in Section 1.4, model (2) is in fact a special case of a more general class of shared frailty models, which includes model (1) as well.

1.3 *Box–Cox Transformation*

In the traditional linear regression model, $Y_i = \boldsymbol{\beta}'\mathbf{Z}_i + \epsilon_i$, the error terms ϵ_i $(i = 1, \ldots, n)$ are usually assumed to be i.i.d. from a zero-mean normal distribution. When the normality assumption of ϵ_i does not hold, the Box–Cox transformation (Box and Cox, 1964) may be applied to the response variable,

$$Y^{(\gamma)} = \begin{cases} (Y^\gamma - 1)/\gamma, & \gamma \neq 0, \\ \log(Y), & \gamma = 0, \end{cases}$$

where γ is the transformation parameter and $\gamma \in R^1$ (the real line).

The Box–Cox transformation has also been applied to independent failure time data. Breslow and Storer (1985) and Barlow (1985) applied this family of power transformations to the covariate structure to model the relative risk $R(\mathbf{Z}_i)$,

$$\log R(\mathbf{Z}_i) = \begin{cases} \{(1 + \boldsymbol{\beta}'\mathbf{Z}_i)^\gamma - 1\}/\gamma, & \gamma \neq 0, \\ \log(1 + \boldsymbol{\beta}'\mathbf{Z}_i), & \gamma = 0. \end{cases}$$

Aranda-Ordaz (1983) proposed a regression model by imposing the Box–Cox transformation on the conditional probability, $-\log\{1 - \Pr(t_{j-1} < T \leq t_j \mid T > t_{j-1})\}$ where $0 = t_0 < t_1 < \cdots < t_m$ is the partition of the time scale. However, the focus was only on grouped data with the multiplicative $(\gamma = 0)$ and additive $(\gamma = 1)$ cases because it appeared to be intractable for a general power parameter γ. A critical assumption of the aforementioned applications of the Box–Cox transformation in survival analysis is the independence of the failure times, which may not be satisfied in the DRS data.

1.4 *A New Class of Frailty Models*

In biomedical studies involving clustered time-to-event data, it might not be reasonable to assume independence of failure times. In this article, we propose a class of shared gamma frailty models based on the Box–Cox transformation for clustered survival data,

$$\frac{\lambda(t \mid \mathbf{Z}_{ik}, W_i)^\gamma - 1}{\gamma} = \frac{\{\lambda_0(t)W_i\}^\gamma - 1}{\gamma} + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t), \quad \gamma \in [0, 1]. \tag{3}$$

The frailty has a multiplicative effect on the common baseline hazard function, implying heterogeneity due to clustering. The W_i's allow members within the same cluster to share a common baseline hazard function, and different hazards across clusters. In linear models, a traditional Box–Cox transformation is applied to the response variable Y, which is observed. However, the transformation proposed here is simultaneously applied to the unknown hazard, and the product of the baseline hazard and the unobservable frailty, as in (3). Throughout this article, we consider γ as known, while treating γ as an unknown parameter might cause numerical difficulty and instability. It is easy to see that as $\gamma \to 0$, (3) becomes

$$\log\{\lambda(t \mid \mathbf{Z}_{ik}, W_i)\} = \log\{\lambda_0(t)W_i\} + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t),$$

which reduces to the Cox frailty model (1), and when $\gamma = 1$, (3) reduces to our proposed additive frailty model (2). Our primary interest in γ lies in [0, 1], which contains broad modeling structures between the proportional $(\gamma = 0)$ and the additive $(\gamma = 1)$ hazards, though γ could mathematically take any value on the real line. The relationship between the hazard functions and the modeling structure change as γ varies from 0 to 1. The interpretations of the β's are conditional on the random effects, which change with respect to γ in (3). For $\gamma = 0$, model (3) yields the hazard ratios, i.e., $\exp(\beta)$, while for $\gamma = 1$, (3) gives the hazard differences. However, for $0 < \gamma < 1$, it is difficult to interpret the β's, which are some intermediate statistical quantities between hazard ratios and differences. We assume that $W_i \sim \text{Gamma}(\eta, \eta)$, where W_i induces the heterogeneity on the baseline hazard due to the cluster-specific effect.

This family of shared gamma frailty models is general and flexible, which allows for a very rich class of hazard patterns. In many multivariate survival applications where the hazards are neither proportional nor parallel, our new gamma frailty model provides a very unified methodology. As will be shown in Section 5 where we apply model (3) to the DRS example, the best fitting model is indeed neither when $\gamma = 0$ nor when $\gamma = 1$, based on a suitable model selection criterion.

The rest of this article is organized as follows. In Section 2.1, we introduce notation and quantify the underlying failure time correlation using the dependence measure, Kendall's τ. In Section 2.2, we derive the likelihood function for the proposed frailty model within the Bayesian paradigm. In Section 2.3, we study the prior distributions under the constrained parameter space. In Section 3, we derive the full conditional distributions needed for Gibbs sampling and specify the constrained sampling support range for each parameter. In Section 4, we propose a model selection technique based on the conditional predictive ordinate (CPO; Geisser, 1993). We illustrate the proposed methods with the DRS example in Section 5, and provide concluding remarks in Section 6.

2. Dependence, Likelihood, and Priors

2.1 *Dependence Measure*

Suppose that there are n clusters, and within cluster i, there are K_i subjects. Let T_{ik} $(i = 1, \ldots, n;\ k = 1, \ldots, K_i)$ be the failure time for the kth subject in the ith cluster, and $\mathbf{Z}_{ik}(t)$ be the corresponding $p \times 1$ vector of bounded and possibly time-dependent covariates. Let C_{ik} be the censoring variable, $Y_{ik} = \min(T_{ik},\ C_{ik})$ be the observed time, and $\nu_{ik} = I(T_{ik} \le C_{ik})$ be the failure time indicator, where $I(\cdot)$ is the indicator function. Assume that T_{ik} and C_{ik} are conditionally independent given $\mathbf{Z}_{ik}(t)$. Within cluster i, $\{(T_{ik},\ C_{ik},\ \mathbf{Z}_{ik}(t)),\ k = 1, \ldots, K_i\}$ may be dependent but exchangeable.

Due to the parameter γ in (3), the β's are intertwined together with the frailty term W_i and the baseline hazard $\lambda_0(t)$, and thus $\lambda_0(t)$ cannot be factored out in the frailty model, leading to

$$\lambda(t \,|\, \mathbf{Z}_{ik}, W_i) = \left[\{\lambda_0(t)W_i\}^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)\right]^{1/\gamma}, \qquad (4)$$

where $W_i \sim \text{Gamma}(\eta,\ \eta)$, with mean 1 and variance η^{-1}.

Without loss of generality, considering the bivariate distribution of $(T_1,\ T_2)$, the joint survival function given time-independent covariate vectors $(\mathbf{Z}_1,\ \mathbf{Z}_2)$ is

$$\begin{aligned} &S(t_1, t_2 \,|\, \mathbf{Z}_1, \mathbf{Z}_2) \\ &= \int_0^{\infty} \exp\left[-\int_0^{t_1} \left\{\lambda_0(u)^{\gamma} W^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_1\right\}^{1/\gamma} du \right. \\ &\qquad \left. -\int_0^{t_2} \left\{\lambda_0(u)^{\gamma} W^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_2\right\}^{1/\gamma} du \right] \\ &\quad \times \frac{\eta^{\eta}}{\Gamma(\eta)} W^{\eta-1} \exp(-\eta W)\, dW. \end{aligned}$$

Apparently, there is no explicit closed form for this bivariate survival function for a general γ. For $\gamma = 0$, that is the Cox shared gamma frailty model,

$$\begin{aligned} &S(t_1, t_2 \,|\, \mathbf{Z}_1, \mathbf{Z}_2) \\ &= \left\{1 + \frac{\Lambda_0(t_1)\exp(\boldsymbol{\beta}'\mathbf{Z}_1) + \Lambda_0(t_2)\exp(\boldsymbol{\beta}'\mathbf{Z}_2)}{\eta}\right\}^{-\eta}, \end{aligned}$$

where $\Lambda_0(\cdot)$ is the cumulative hazard function, and for $\gamma = 1$ corresponding to model (2),

$$\begin{aligned} &S(t_1, t_2 \,|\, \mathbf{Z}_1, \mathbf{Z}_2) \\ &= \left\{1 + \frac{\Lambda_0(t_1) + \Lambda_0(t_2)}{\eta}\right\}^{-\eta} \exp\{-(\boldsymbol{\beta}'\mathbf{Z}_1 t_1 + \boldsymbol{\beta}'\mathbf{Z}_2 t_2)\}. \end{aligned}$$

As $\eta \to \infty$, the bivariate failure times $(T_1,\ T_2)$ become independent, i.e., $S(t_1,\ t_2 \,|\, \mathbf{Z}_1,\ \mathbf{Z}_2) = S(t_1 \,|\, \mathbf{Z}_1)S(t_2 \,|\, \mathbf{Z}_2)$.

Kendall's τ (Kendall, 1938) is a global measure of dependence, which can be estimated by enumeration of the concordant and discordant pairs of bivariate observations (Hougaard, 2000). For independent clusters of a and b with paired data, $(T_{a1},\ T_{a2})$ and $(T_{b1},\ T_{b2})$,

$$\tau = \text{E}[\text{sign}\{(T_{a1} - T_{b1})(T_{a2} - T_{b2})\}].$$

An alternative definition is given by the integration of the bivariate baseline survival function,

$$\tau = 4\int_0^{\infty}\int_0^{\infty} f(t_1, t_2)S(t_1, t_2)\, dt_1\, dt_2 - 1,$$

where $f(t_1,\ t_2)$ is the joint density function. We evaluate Kendall's τ without any covariates, in which case model (4) reduces to the proportional hazards structure no matter what value γ takes. Therefore, Kendall's τ is invariant with respect to γ and $\tau = (2\eta + 1)^{-1}$. For the Cox shared gamma frailty model ($\gamma = 0$), τ takes the same form with or without covariates. To quantify the short- or long-term dependence, a local correlation measure $\theta(t_1,\ t_2)$ (Clayton, 1978; Oakes, 1989) is defined as

$$\theta(t_1, t_2) = \frac{S(t_1, t_2)\Delta_1\Delta_2 S(t_1, t_2)}{\Delta_1 S(t_1, t_2)\Delta_2 S(t_1, t_2)},$$

where Δ_i denotes the operator $-\partial/\partial t_i$, $i = 1, 2$. Without any covariates, $\theta = 1 + \eta^{-1}$ for all γ in (4).

2.2 *Likelihood Function*

The following complex nonlinear constraints in $\{\boldsymbol{\beta}, \lambda_0(t)W_i, \gamma\}$ need to be satisfied,

$$\{\lambda_0(t)W_i\}^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik} \ge 0 \quad (i = 1, \ldots, n;\ k = 1, \ldots, K_i). \qquad (5)$$

Here, we propose to carry out estimation and inference using a Bayesian approach. We assume a piecewise exponential distribution for the baseline hazard function $\lambda_0(t)$ (Ibrahim, Chen, and Sinha, 2001). Let J denote the number of partitions of the time axis, i.e., $0 < s_1 < \cdots < s_J$. When $J = 1$, namely with no partition, the baseline hazard reduces to that of an exponential distribution. The piecewise constant hazards model assumes that $\lambda_0(y) = \lambda_j$ for $y \in (s_{j-1},\ s_j]$, $j = 1, \ldots, J$. By increasing J, i.e., with finer partitions of the time scale, we obtain a more flexible model for the underlying baseline hazard. The piecewise exponential model is useful and simple for modeling survival data, which serves as a benchmark for comparisons with other semiparametric and fully parametric models. The likelihood function is constructed as follows. Let y_{ik} be the observed time for the kth subject in the ith cluster, $\mathbf{y} = (y_{11}, \ldots, y_{1K_1}; \ldots; y_{n1}, \ldots, y_{nK_n})'$, $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{1K_1}; \ldots; \nu_{n1}, \ldots, \nu_{nK_n})'$, and $\mathbf{Z}(t)$ be the $(N \times p)$-dimensional covariate matrix where N is the total sample size, i.e., $N = \sum_{i=1}^{n} K_i$. Define $\delta_{ikj} = 1$ if subject k in cluster i fails or is censored in interval j, and 0 otherwise. The hazard function in the jth interval for the shared gamma frailty model is then given by

$$\lambda_j(t \,|\, \mathbf{Z}_{ik}, W_i) = \left\{(\lambda_j W_i)^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)\right\}^{\delta_{ikj}/\gamma},$$

where γ is assumed fixed (known) in the range $[0, 1]$. Let $D = (N, \mathbf{y}, \mathbf{Z}(t), \boldsymbol{\nu})$ denote the observed data, $\mathbf{W} = (W_1, \ldots, W_n)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$. For ease of exposition, let $\mathbf{Z}_{ik} \equiv \mathbf{Z}_{ik}(t)$. The likelihood function is given by

$$\begin{aligned} &L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W} \,|\, D) \\ &= \prod_{i=1}^{n} \left(\prod_{k=1}^{K_i} \prod_{j=1}^{J} \left(\lambda_j^{\gamma} W_i^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik}\right)^{\delta_{ikj}\nu_{ik}/\gamma} \right. \\ &\qquad \times \exp\left[-\delta_{ikj} \left\{ \left(\lambda_j^{\gamma} W_i^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik}\right)^{1/\gamma} (y_{ik} - s_{j-1}) \right.\right. \\ &\qquad \left.\left. \left. + \sum_{q=1}^{j-1} \left(\lambda_q^{\gamma} W_i^{\gamma} + \gamma\boldsymbol{\beta}'\mathbf{Z}_{ik}\right)^{1/\gamma} (s_q - s_{q-1}) \right\} \right] \right) \pi(W_i), \end{aligned} \qquad (6)$$

where $\pi(W_i)$ is the probability density function (p.d.f.) of a Gamma$(\eta,\ \eta)$ random variable. For a general γ, the marginal

likelihood involves a complicated integral over W_i, which does not have a closed form. Therefore, instead of integrating W_i out, we treat $(W_1, \ldots, W_n)$ as "parameters" that need to be sampled in the Gibbs iterations as shown in Section 3.

2.3 *Prior Specification with Constrained Parameters*

The joint prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W})$ needs to accommodate the nonnegativity constraint for the hazard function,

$$\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}' \mathbf{Z}_{ik} \geq 0 \quad (i = 1, \ldots, n;\ k = 1, \ldots, K_i;\ j = 1, \ldots, J). \tag{7}$$

Bayesian computation and analysis become complicated in the constrained parameter space (Gelfand, Smith, and Lee, 1992; Chen and Shao, 1998; Chen, Shao, and Ibrahim, 2000, Chapter 6). For example, the order constraints on a set of parameters is very common in Bayesian modeling, where one can incorporate monotonicity constraints to improve testing efficiency (Dunson and Neelon, 2003). However, (7) is very different from the usual order constrained parameter problems. If the nonlinear constraint (7) is violated, the likelihood function and the posterior density are not well defined.

To accommodate the constraint (7), we bypass the multidimensional inequality and reduce the prior specification to a one-dimensional truncated distribution by absorbing (7) into *one* parameter β_g. Hence, we are able to obtain the normalizing constant required for the full conditional distributions of $\boldsymbol{\lambda}$, $\mathbf{W}$, and the rest of β's in a closed form. Without loss of generality and for ease of exposition, we assume that all the covariates are positive. Let $\mathbf{Z}_{(-g)}$ denote the $N \times (p-1)$-dimensional covariate matrix with the gth covariate (column) deleted, and $\boldsymbol{\beta}_{(-g)}$ denote the $(p-1)$-dimensional parameter vector with the gth component of $\boldsymbol{\beta}$ removed; thus $\mathbf{Z}_{(-g)} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{g-1}, \mathbf{Z}_{g+1}, \ldots, \mathbf{Z}_p)$ and $\boldsymbol{\beta}_{(-g)} = (\beta_1, \ldots, \beta_{g-1}, \beta_{g+1}, \ldots, \beta_p)'$. We propose a joint "prior" for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W})$ of the form

$$\begin{aligned}
&\pi(\beta_1, \ldots, \beta_p; \lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n) \\
&\quad = \pi(\beta_g \,|\, \beta_1, \ldots, \beta_{g-1}, \beta_{g+1}, \ldots, \beta_p; \lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n) \\
&\qquad \times I\Bigg(\beta_g \geq -\frac{\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}'_{(-g)} \mathbf{Z}_{(-g),ik}}{\gamma Z_{g,ik}}, \\
&\qquad\qquad i = 1, \ldots, n;\ k = 1, \ldots, K_i;\ j = 1, \ldots, J\Bigg) \\
&\qquad \times \pi(\beta_1, \ldots, \beta_{g-1}, \beta_{g+1}, \ldots, \beta_p; \lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n).
\end{aligned} \tag{8}$$

From (8), we see that β_g and $(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W})$ are *not* independent a priori due to the nonlinear constraint. This joint prior specification only involves *one* parameter β_g in the constraints and makes all the other parameters $(\beta_1, \ldots, \beta_{g-1}, \beta_{g+1}, \ldots, \beta_p; \lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n; \eta)$ free of constraints. It gives great flexibility in the choice of joint prior and is computationally attractive. Informative priors via historical data or expert opinion can be easily specified through the free parameters $\boldsymbol{\beta}_{(-g)}$, $\boldsymbol{\lambda}$, and η.

Let $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution. Specifically, we take $(\beta_g \,|\, \boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W})$ to have a truncated normal prior distribution given by

$$\begin{aligned}
&\pi\big(\beta_g \,|\, \boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\big) \\
&\quad = c^{-1}\big(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\big) \exp\left\{-\frac{\beta_g^2}{2\sigma_g^2}\right\} \\
&\qquad \times I\left(\beta_g \geq -\min_{i,k,j}\left\{\frac{\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}'_{(-g)} \mathbf{Z}_{(-g),ik}}{\gamma Z_{g,ik}}\right\}\right),
\end{aligned} \tag{9}$$

where the normalizing constant depends on $\boldsymbol{\beta}_{(-g)}$, $\boldsymbol{\lambda}$, and $\mathbf{W}$,

$$\begin{aligned}
&c\big(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\big) \\
&\quad = \int_{-\min_{i,k,j}\left\{\frac{\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}'_{(-g)} \mathbf{Z}_{(-g),ik}}{\gamma Z_{g,ik}}\right\}}^{\infty} \exp\left\{-\frac{\beta_g^2}{2\sigma_g^2}\right\} d\beta_g \\
&\quad = \sqrt{2\pi}\sigma_g \left[1 - \Phi\left(-\min_{i,k,j}\left\{\frac{\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}'_{(-g),ik} \mathbf{Z}_{(-g),ik}}{\gamma Z_{g,ik} \sigma_g}\right\}\right)\right].
\end{aligned} \tag{10}$$

Although not required for the development, we can take $\boldsymbol{\beta}_{(-g)}$, $\boldsymbol{\lambda}$, and $\mathbf{W}$ to be independent a priori in (8), that is,

$$\pi(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}) = \pi(\boldsymbol{\beta}_{(-g)}) \pi(\boldsymbol{\lambda}) \pi(\mathbf{W} \,|\, \eta) \pi(\eta).$$

The components of $\boldsymbol{\lambda}$ are assumed to be independent a priori, and each λ_j has a Gamma$(\alpha_\lambda, \xi_\lambda)$ distribution. One can easily construct priors to make the λ_j's dependent a priori by considering first-order autoregressive structures or Markovian relations on the λ_j's, as in Arjas and Gasbarra (1994) and Ibrahim et al. (2001). We assume that $W_i \sim$ Gamma(η, η), and $\eta \sim$ Gamma(a, b) where the shape parameter a and the scale parameter b are specified to yield a large prior unconditional variance of W_i, with var$(W_i) = b/(a-1)$.

3. Gibbs Sampling

For $\gamma = 0$, (4) is the Cox shared gamma frailty model, and the likelihood function is log-concave for all the parameters $(\beta_1, \ldots, \beta_p; \lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n)$ (for details, see Ibrahim et al., 2001, p. 100–112). For $0 < \gamma \leq 1$, because the priors for the β's are log-concave, without loss of generality, it suffices to examine only the likelihood function in (6) for the log-concave properties of the parameters. It can be shown that for $\gamma \in (0, 1]$, the second derivative $\partial^2 \log L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W} \,|\, D)/\partial \beta_l^2 \leq 0$, $(l = 1, \ldots, p)$, that is the full conditional of β_l is log-concave in β_l. Therefore, we only need to use the adaptive rejection sampling algorithm proposed by Gilks and Wild (1992) to sample from the full conditional distributions of β_l $(l = 1, \ldots, p)$. However, the full conditionals of $(\lambda_1, \ldots, \lambda_J; W_1, \ldots, W_n)$ are not log-concave, thus a metropolis step needs to be implemented within the Gibbs steps (Gilks, Best, and Tan, 1995).

Suppose that the gth component of $\boldsymbol{\beta}$ has a truncated normal prior as given in (9), to satisfy the nonlinear constraint (5), and the rest of the parameters are unconstrained. We define the conditional likelihood function as

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D) = \prod_{i=1}^{n} L_i(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D),$$

where $L_i(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D)$ is the likelihood contribution for the ith cluster. Let $[U \,|\, V]$ denote the posterior distribution of U given V. The full conditionals of the parameters are given as follows:

$$\left[\beta_g \,\middle|\, \boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}, D\right] \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D)\pi(\beta_g),$$

$$\left[\beta_l \,\middle|\, \boldsymbol{\beta}_{(-l)}, \boldsymbol{\lambda}, \mathbf{W}, D\right] \propto \frac{L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D)\pi(\beta_l)}{c\left(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\right)} \quad (l \neq g, l = 1, \ldots, p),$$

$$\left[\lambda_j \,\middle|\, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-j)}, \mathbf{W}, D\right] \propto \frac{L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D)\pi(\lambda_j)}{c\left(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\right)} \quad (j = 1, \ldots, J),$$

$$\left[W_i \,\middle|\, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W}_{(-i)}, \eta, D\right] \propto \frac{L_i(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D) W_i^{\eta-1} \exp(-\eta W_i)}{c\left(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W}\right)} \quad (i = 1, \ldots, n),$$

$$[\eta \,|\, \mathbf{W}, D] \propto \frac{\eta^{n\eta+a-1}\left(\prod_{i=1}^{n} W_i\right)^{\eta-1} \exp\left\{-\eta\left(\sum_{i=1}^{n} W_i + b\right)\right\}}{\{\Gamma(\eta)\}^n},$$

where $\pi(\beta_l)$ is the p.d.f. of an $N(\mu_l, \sigma_l^2)$ distribution, $\pi(\lambda_j)$ is that of a Gamma$(\alpha_\lambda, \xi_\lambda)$ distribution, and $c(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W})$ is given by (10). These full conditionals have nice tractable structures with our proposed prior specification because $c(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W})$ has an analytic closed form. The posterior estimates are very robust with respect to the choice of g in (8), which is an attractive and remarkable feature as demonstrated in Section 5. In the usual random effects model, the full conditional distribution of the ith random effect, W_i, only involves the data corresponding to the ith cluster. However, in our proposed model, the full conditional of W_i involves the data from all clusters through the normalizing constant $c(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{W})$. In each Gibbs sampling step, the support for each parameter needs to be set to accommodate the nonlinear constraint, such that the likelihood function is always well defined within the sampling boundaries. For $i = 1, \ldots, n$; $k = 1, \ldots, K_i$; $j = 1, \ldots, J$; $l = 1, \ldots, p$, the following inequalities need to be satisfied:

$$\beta_l \geq -\frac{\lambda_j^\gamma W_i^\gamma + \gamma \boldsymbol{\beta}'_{(-l)} \mathbf{Z}_{(-l),ik}}{\gamma Z_{l,ik}},$$

$$\lambda_j \geq \left\{-\min\left(\frac{\gamma \boldsymbol{\beta}' \mathbf{Z}_{ik}}{W_i^\gamma}, 0\right)\right\}^{1/\gamma},$$

$$W_i \geq \left\{-\min\left(\frac{\gamma \boldsymbol{\beta}' \mathbf{Z}_{ik}}{\lambda_j^\gamma}, 0\right)\right\}^{1/\gamma}.$$

4. Model Adequacy Evaluation

Model checking plays an important role in regression models. It is critical to compare a class of competing models for a given dataset and select the one that best fits the data. The CPO statistic cross-validates the conditional predictive distribution from single observation deletion against the observed responses (Gelfand, Dey, and Chang, 1992; Geisser, 1993; Dey, Chen, and Chang, 1997).

We derive the cluster-based CPO statistic as follows. For the ith cluster, $\mathbf{y}_i = (y_{i1}, \ldots, y_{iK_i})'$ are assumed to be independent, conditional on the unobservable random effect W_i. Let $\mathbf{y}^{(-i)}$ denote the $(N - K_i) \times 1$ response vector with $\mathbf{y}_i$ deleted, and thus $\mathbf{y} = \{\mathbf{y}_i', \mathbf{y}^{(-i)\prime}\}'$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \mathbf{W}')'$. The CPO_i for the ith cluster is defined as the joint posterior predictive density of $\mathbf{y}_i$ given $\mathbf{y}^{(-i)}$, which can be written as

$$\begin{aligned}\mathrm{CPO}_i &= f\left(\mathbf{y}_i \,\middle|\, \mathbf{y}^{(-i)}\right) = \frac{f(\mathbf{y})}{f\left(\mathbf{y}^{(-i)}\right)} \\ &= \frac{\displaystyle\int \frac{f(\mathbf{y} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) f(\mathbf{y})} \pi(\boldsymbol{\theta} \,|\, \mathbf{y})\, d\boldsymbol{\theta}}{\displaystyle\int \frac{f\left(\mathbf{y}^{(-i)} \,\middle|\, \boldsymbol{\theta}\right)\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) f(\mathbf{y})} \pi(\boldsymbol{\theta} \,|\, \mathbf{y})\, d\boldsymbol{\theta}} \\ &= \left\{\int \frac{\pi(\boldsymbol{\theta} \,|\, \mathbf{y})}{\displaystyle\prod_{k=1}^{K_i} f(y_{ik} \,|\, \boldsymbol{\theta})}\, d\boldsymbol{\theta}\right\}^{-1},\end{aligned}$$

where $\pi(\boldsymbol{\theta} \,|\, \mathbf{y})$ is the joint posterior density function and $f(y_{ik} \,|\, \boldsymbol{\theta})$ is the conditional density function of y_{ik}. For model (6), a Monte Carlo approximation of CPO_i is given by

$$\widehat{\mathrm{CPO}}_i = \left\{\frac{1}{M}\sum_{m=1}^{M} \frac{1}{L_i\left(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,\middle|\, W_{i,[m]}, \mathbf{y}_i, \mathbf{Z}_i\right)}\right\}^{-1},$$

where M is the number of Gibbs samples after burn-in, and

$$\begin{aligned}&L_i\left(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,\middle|\, W_{i,[m]}, \mathbf{y}_i, \mathbf{Z}_i\right) \\ &= \prod_{k=1}^{K_i}\prod_{j=1}^{J}\left(\lambda_{j,[m]}^\gamma W_{i,[m]}^\gamma + \gamma \boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik}\right)^{\delta_{ikj}\nu_{ik}/\gamma} \\ &\quad \times \exp\left[-\delta_{ikj}\left\{\left(\lambda_{j,[m]}^\gamma W_{i,[m]}^\gamma + \gamma \boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik}\right)^{1/\gamma}(y_{ik} - s_{j-1}) \right.\right. \\ &\qquad \left.\left. + \sum_{q=1}^{j-1}\left(\lambda_{q,[m]}^\gamma W_{i,[m]}^\gamma + \gamma \boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik}\right)^{1/\gamma}(s_q - s_{q-1})\right\}\right],\end{aligned}$$

$\boldsymbol{\beta}_{[m]} = (\beta_{1,[m]}, \ldots, \beta_{p,[m]})'$, $\boldsymbol{\lambda}_{[m]} = (\lambda_{1,[m]}, \ldots, \lambda_{J,[m]})'$, and $W_{i,[m]}$ $(i = 1, \ldots, n)$ are the Markov chain Monte Carlo (MCMC) samples of the mth Gibbs iteration. For correlated data, we propose the summary statistic $B = \sum_{i=1}^{n} \log(\mathrm{CPO}_i)$, where a larger value of B indicates a better fit.

5. Application

As an illustration, we applied the proposed methods to data from the DRS example. There were 197 high-risk patients in the dataset, and we considered three covariates: treatment, type of diabetes (adult-onset or juvenile-onset diabetes), and age (ranging from 1 to 58 years with a mean of around 21 years). The response variable was time to blindness (in months), which might be right censored. Figure 1 shows the Kaplan–Meier survival curves stratified by treatment and type of diabetes groups.

In the analysis, we constrained the regression coefficient for laser treatment (β_1) to have a truncated normal prior, and set $\gamma = (0, 0.25, 0.5, 0.75, 1)$. The priors for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$ were all taken to be noninformative, where $(\beta_1 \,|\, \beta_2, \beta_3, \boldsymbol{\lambda}, \mathbf{W})$ had a truncated $N(0, 10,000)$ prior as defined in (9), β_2 and β_3 were independent a priori and

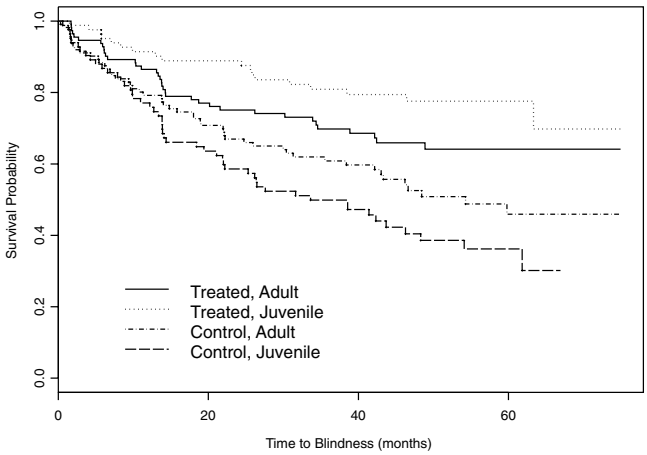

Figure 1. Kaplan–Meier curves stratified by treatment and type of diabetes in the DRS.

were taken to have $N(0,\ 10,\ 000)$ distributions. For $\gamma = 0$, we took $\lambda_j \sim \text{Gamma}(0.001,\ 0.001)$, and independent for $j = 1, \ldots, J$, and $W_i \sim \text{Gamma}(\eta,\ \eta)$, $\eta \sim \text{Gamma}(a,\ b)$ with $a = b = 0.001$ to yield a noninformative "prior" on W_i. For $\gamma \in (0,\ 1]$, we took $\lambda_j \sim \text{Gamma}(2,\ 0.001)$, and $a = 2$ and $b = 1000$ in order to have a large variance for W_i. Under different modeling structures, we accordingly chose more suitable priors that would lead to better model convergence. We specified noninformative priors such that the likelihood dominated the posterior distribution. We took $J = 1$ (no partition) and $J = 5$ because the shape and flexibility of the baseline hazard function is controlled by J. The optimal J usually lies around 5, seldom exceeds 10, and the partition points should be chosen such that we would obtain an approximately equal number of failures in each interval. For the proposed shared gamma frailty model, γ also directly affects the shapes of the hazard functions. The parameter estimates are usually very robust with respect to J, whereas γ has a critical influence on the estimates of the β's. Therefore, we look for the best γ, which yields the largest B-statistic for a given J.

After taking a burn-in of 5000 samples, we then ran 30,000 Gibbs iterations and kept every fifth iteration (thinning = 5). Table 1 summarizes the B-statistic, posterior mean, standard deviation, and the 95% highest posterior density (HPD) interval for the β's and η, using $J = 1$ (the parametric model), and Table 2 corresponds to $J = 5$. The parameter estimates are close between $J = 1$ and $J = 5$ for a given γ, but are different among the set of γ's for a given J. Across the 10 competing models, the laser treatment effect is significant and neither the type of diabetes nor age has significant effects on survival. Therefore, the treatment effect is well confirmed under a variety of shared gamma frailty models with different γ's. By appropriately varying γ from 0 to 1, we can obtain a comprehensive evaluation of covariate effects under a series of modeling structures. Using $J = 5$, the model with $\gamma = 0$ concludes that the untreated eye of a subject is about 2.6 times at more risk of blindness than his or her treated eye, $\exp(\hat{\beta}_{\text{treat},\gamma=0}) = 2.6$, while the model with $\gamma = 1$ states that the risk difference per month between the untreated and treated eyes of a subject is around 0.006, $\hat{\beta}_{\text{treat},\gamma=1} = 0.0063$. Based on the B-statistics, the model with $J = 5$ and $\gamma = 0.75$ is deemed as the best fitting model, from which all of the following analyses will be based. To quantify the underlying failure time dependence, we estimated [Kendall's $\hat{\tau} = 0.848$ and Oakes' $\hat{\theta} = 12.186$], using $J = 5$, where $\tau \to 0$ or $\theta \to 1$ indicates independence of the bivariate failure times.

Model convergence was monitored by running two parallel MCMC chains with diverse starting values and different

Table 1
Posterior means, standard deviations, and 95% *HPD intervals for the DRS data, using* $J = 1$

γ	B	Covariate	Mean	Std. dev.	95% HPD interval
0	−830.4	Treatment	0.9974	0.1790	(0.6434, 1.3506)
		Type of diabetes	−0.3796	0.4193	(−1.2307, 0.4287)
		Age	0.2416	0.2134	(−0.1771, 0.6608)
		Frailty η	0.9162	0.2693	(0.5501, 1.6018)
0.25	−825.4	Treatment	0.3856	0.0509	(0.2879, 0.4852)
		Type of diabetes	0.1166	0.1210	(−0.1277, 0.3419)
		Age	−0.0082	0.0802	(−0.1643, 0.1500)
		Frailty η	0.0947	0.0072	(0.0815, 0.1091)
0.5	−824.4	Treatment	0.0927	0.0145	(0.0642, 0.1206)
		Type of diabetes	−0.0130	0.0266	(−0.0663, 0.0395)
		Age	0.0057	0.0177	(−0.0286, 0.0412)
		Frailty η	0.0933	0.0073	(0.0794, 0.1081)
0.75	−826.7	Treatment	0.0241	0.0043	(0.0156, 0.0324)
		Type of diabetes	−0.0083	0.0064	(−0.0198, 0.0049)
		Age	0.0022	0.0044	(−0.0061, 0.0113)
		Frailty η	0.0946	0.0072	(0.0813, 0.1095)
1	−825.9	Treatment	0.0065	0.0013	(0.0039, 0.0089)
		Type of diabetes	−0.0028	0.0015	(−0.0053, 0.0006)
		Age	0.0006	0.0010	(−0.0013, 0.0028)
		Frailty η	0.0938	0.0073	(0.0797, 0.1084)

Table 2
Posterior means, standard deviations, and 95% *HPD intervals for the DRS data, using* $J = 5$

γ	B	Covariate	Mean	Std. dev.	95% HPD interval
0	−835.3	Treatment	0.9449	0.1765	(0.5996, 1.3003)
		Type of diabetes	−0.3220	0.4089	(−1.1283, 0.4693)
		Age	0.2158	0.2014	(−0.1684, 0.6037)
		Frailty η	1.2487	0.6928	(0.6123, 2.9588)
0.25	−827.9	Treatment	0.3817	0.0531	(0.2778, 0.4844)
		Type of diabetes	0.0100	0.1344	(−0.2646, 0.2624)
		Age	0.0221	0.0870	(−0.1454, 0.1893)
		Frailty η	0.0933	0.0073	(0.0798, 0.1080)
0.5	−822.7	Treatment	0.0902	0.0147	(0.0609, 0.1187)
		Type of diabetes	−0.0152	0.0274	(−0.0675, 0.0380)
		Age	0.0053	0.0176	(−0.0286, 0.0396)
		Frailty η	0.0882	0.0069	(0.0753, 0.1021)
0.75	−819.4	Treatment	0.0233	0.0041	(0.0152, 0.0314)
		Type of diabetes	−0.0085	0.0061	(−0.0191, 0.0040)
		Age	0.0023	0.0043	(−0.0055, 0.0114)
		Frailty η	0.0894	0.0072	(0.0759, 0.1039)
1	−821.1	Treatment	0.0063	0.0012	(0.0039, 0.0086)
		Type of diabetes	−0.0027	0.0014	(−0.0052, 0.0005)
		Age	0.0006	0.0010	(−0.0011, 0.0029)
		Frailty η	0.0880	0.0068	(0.0753, 0.1019)

seeds, and these converged to the same range of values for each parameter. All of the model parameters appeared to mix satisfactorily from the trace plots. The convergence test proposed by Gelman and Rubin (1992) is analogous to a classical analysis of variance, which compares the within-chain and between-chain variances. We estimated the factor by which the scale parameter might shrink if the chain were run to infinity. The median and 97.5% quantiles of the sampling distribution for the shrink factor reported by the Gelman and Rubin (1992) diagnostics were close to 1 for all the β's, λ's, and η's. This indicated convergence of the Markov chains.

We carried out two sets of sensitivity analyses to investigate the robustness of the posterior inference with respect to the prior distributions and the choice of β_g assigned to have the truncated normal prior. Table 3 shows that the parameter estimates obtained from the proposed model are very robust with respect to noninformative prior distributions

Table 3
Sensitivity analysis with noninformative prior specifications for the DRS data using $J = 5$ *and* $\gamma = 0.75$

σ_β	ξ_λ	b	Covariate	Mean	Std. dev.	95% HPD interval
10	0.001	1000	Treatment	0.0236	0.0043	(0.0152, 0.0320)
			Type of diabetes	−0.0072	0.0062	(−0.0181, 0.0061)
			Age	0.0013	0.0041	(−0.0066, 0.0095)
1000	0.001	1000	Treatment	0.0231	0.0043	(0.0147, 0.0316)
			Type of diabetes	−0.0076	0.0066	(−0.0200, 0.0057)
			Age	0.0026	0.0046	(−0.0058, 0.0122)
100	0.01	1000	Treatment	0.0239	0.0041	(0.0157, 0.0318)
			Type of diabetes	−0.0089	0.0060	(−0.0194, 0.0035)
			Age	0.0022	0.0041	(−0.0050, 0.0107)
100	0.0001	1000	Treatment	0.0236	0.0042	(0.0154, 0.0317)
			Type of diabetes	−0.0092	0.0063	(−0.0202, 0.0037)
			Age	0.0034	0.0043	(−0.0048, 0.0119)
100	0.001	100	Treatment	0.0225	0.0041	(0.0146, 0.0306)
			Type of diabetes	−0.0096	0.0055	(−0.0189, 0.0022)
			Age	0.0027	0.0039	(−0.0045, 0.0110)
100	0.001	10,000	Treatment	0.0245	0.0042	(0.0163, 0.0326)
			Type of diabetes	−0.0070	0.0063	(−0.0184, 0.0062)
			Age	0.0012	0.0042	(−0.0066, 0.0096)

Table 4
Analysis of the DRS data with different regression parameters having the truncated normal priors using $J = 5$ and $\gamma = 0.75$

Truncated coefficient	Covariate	Mean	Std. dev.	95% HPD interval
Type of diabetes	Treatment	0.0234	0.0042	(0.0152, 0.0316)
	Type of diabetes	−0.0084	0.0061	(−0.0190, 0.0047)
	Age	0.0021	0.0041	(−0.0055, 0.0105)
Age	Treatment	0.0237	0.0041	(0.0157, 0.0316)
	Type of diabetes	−0.0084	0.0059	(−0.0186, 0.0040)
	Age	0.0020	0.0040	(−0.0054, 0.0103)

Table 5
Frequentist analysis of the DRS data under the Cox shared gamma frailty model

Covariate	$\hat{\beta}$	SE	p-Value
Treatment	0.919	0.175	<0.001
Type of diabetes	−0.304	0.401	0.450
Age	0.208	0.201	0.300

under a wide range of hyperparameters. Table 4 summarizes the results regarding the constraint, where β_2 (type of diabetes) or β_3 (age) has the truncated normal prior. Clearly, the posterior estimates are very robust with respect to the choice of the constrained parameter. This demonstrates the flexibility and robustness of our proposed prior specification. In the frequentist paradigm, model (1) has been studied extensively and estimation is usually based on the EM algorithm. For example, Klein (1992) proposed a profile likelihood construction while assuming a nonparametric baseline intensity function as opposed to the piecewise exponential rate. For comparison, we applied the frequentist estimation procedure to the DRS data under model (1) and summarized the results in Table 5. The estimate of the frailty parameter is $\hat{\eta} = 1.149$, which is statistically significant at the $\alpha = 0.05$ level. Under the Cox shared gamma frailty model, the frequentist results are very close to those corresponding to the Bayesian model with $\gamma = 0$, especially for $J = 5$.

6. Discussion

We have proposed a class of shared gamma frailty models based on the Box–Cox transformation for multivariate failure time data. This family of frailty models makes the hazard modeling scheme more flexible, general, and versatile compared to other methods, and facilitates a wide variety of relationships between the baseline hazard and hazard function. Proportional or additive modeling structures are now unified into a general class. Due to the complexity of the model, we have proposed a general joint prior specification by absorbing the nonlinear constraint into one parameter while leaving all other parameters free of constraints. This prior specification can be applied to other constrained parameter problems arising from other types of regression models. The proposed methods would have great potential in applications where correlated survival data arise.

An alternative form is to include the random effect as an additive factor. For example, a possible model is given by

$$\frac{\lambda(t \mid \mathbf{Z}_{ik}, W_i)^{\gamma} - 1}{\gamma} = \frac{\{\lambda_0(t) + W_i\}^{\gamma} - 1}{\gamma} + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t). \quad (11)$$

Assuming a zero-mean normal distribution for W_i, the nonnegative hazard constraint imposed by (11) is rarely achieved in practice and thus (11) is not an attractive from of the model. As $\gamma \to 0$, model (11) reduces to

$$\lambda(t \mid \mathbf{Z}_{ik}, W_i) = \{\lambda_0(t) + W_i\} \exp\{\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)\},$$

which is not the well-known Cox shared gamma frailty model.

By specifying a multiplicative random effect on the baseline hazard in (3), i.e., $\lambda_0(t)\, W_i$, we are able to assume $W_i \sim$ Gamma(η, η) for every γ in [0, 1]. The frailty may be incorporated into the model in a different form or fashion, which heavily depends on the motivation from the scientific application as well as mathematical convenience.

Acknowledgements

We would like to thank the editor, an associate editor, and a referee for their critical reading of the manuscript and constructive comments, which led to a great improvement of an earlier version. We thank Donald Berry and Peter Müller for very helpful discussions. This research was partially supported by National Cancer Institute grants 70101 and 74015.

Résumé

Nous proposons ici, dans le cadre des données de survie, une nouvelle classe de modèles à fragilité gamma partagée en appliquant la transformation de Box-Cox à la fonction de hasard, ainsi qu'au produit du risque de base et de la fragilité. Cette nouvelle classe de modèles offre un large éventail de formes et relations entre la fonction de hasard et le risque de base, et inclut, en tant que cas particuliers, le modèle de Cox usuel à fragilité gamma ainsi qu'un nouveau modèle additif à fragilité gamma. De par la contrainte que le risque soit non négatif, ce modèle à fragilité gamma partagée présente un défi calculatoire dans le paradigme bayésien. Les distributions conjointes a priori sont définies par une distribution conditionnelle marginale, où la distribution conditionnelle est univariée: cela permet d'intégrer les contraintes non linéaires sur les paramètres. La partie marginale de la distribution a priori, quant à elle, est libre de toute contrainte. Les distributions a priori permettent de programmer facilement l'intégralité des distributions conditionnelles requises pour l'échantillonnage de Gibbs, tout en intégrant les contraintes. Un jeu réel de

données permet d'illustrer cette classe de modèles à fragilité gamma partagée.

References

Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics* **39,** 109–117.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4,** 505–524.

Aslanidou, H., Dey, D. K., and Sinha, D. (1998). Bayesian analysis of multivariate survival data using Monte Carlo methods. *Canadian Journal of Statistics* **26,** 38–48.

Barlow, W. E. (1985). General relative risk models in stratified epidemiologic studies. *Applied Statistics* **34,** 246–257.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B* **26,** 211–252.

Breslow, N. E. and Storer, B. E. (1985). General relative risk functions for case-control studies. *American Journal of Epidemiology* **122,** 149–162.

Chen, M. and Shao, Q. (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika* **85,** 73–87.

Chen, M., Shao, Q., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familiar tendency in chronic disease incidence. *Biometrika* **65,** 141–151.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47,** 467–485.

Clayton, D. G. and 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 the Royal Statistical Society, Series B* **34,** 187–220.

Dey, D. K., Chen, M., and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53,** 1239–1252.

Diabetic Retinopathy Study Research Group. (1985). Diabetic Retinopathy Study. *Investigative Ophthalmology and Visual Science* **21**(part 2), 149–226.

Dunson, D. B. and Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics* **59,** 286–295.

Geisser, S. (1993). *Predictive Inference: An Introduction*. London: Chapman and Hall.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion). In *Bayesian Statistics*, Volume 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 147–167. Oxford: Oxford University Press.

Gelfand, A. E., Smith, A. F. M., and Lee, T. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87,** 523–532.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7,** 457–511.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41,** 337–348.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* **44,** 455–472.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Ibrahim, J. G., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* **30,** 81–93.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48,** 795–806.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81,** 61–71.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84,** 487–493.

Qiou, Z., Ravishanker, N., and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics* **55,** 637–644.

Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D. (1997). A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis* **3,** 123–137.

Sargent, D. J. (1998). A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics* **54,** 1486–1497.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association* **88,** 979–998.

Received August 2003. *Revised February* 2004.
Accepted April 2004.