Published in final edited form as:

Scand Stat Theory Appl. 2014 March ; 41(1): 187–199. doi:10.1111/sjos.12010.

Bayesian Transformation Models for Multivariate Survival Data

MÁRIO DE CASTRO,

Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo

MING-HUI CHEN,

Department of Statistics, University of Connecticut

JOSEPH G. IBRAHIM, and

Department of Biostatistics, University of North Carolina

JOHN P. KLEIN

Division of Biostatistics, Medical College of Wisconsin

Abstract

In this paper we propose a general class of gamma frailty transformation models for multivariate survival data. The transformation class includes the commonly used proportional hazards and proportional odds models. The proposed class also includes a family of cure rate models. Under an improper prior for the parameters, we establish propriety of the posterior distribution. A novel Gibbs sampling algorithm is developed for sampling from the observed data posterior distribution. A simulation study is conducted to examine the properties of the proposed methodology. An application to a data set from a cord blood transplantation study is also reported.

Keywords

cure rate; gamma frailty; Gibbs sampler; piecewise exponential model; proportional hazards model; proportional odds model

1 Introduction

In the analysis of univariate censored survival data, the Cox proportional hazards model (Cox, 1972) plays a prominent role. Even when more complex models are fit to a given data set, the Cox model may often serve as a starting point. As an alternative, under the proportional odds model (Bennett, 1983), there is proportionality between the odds computed with different covariate values. However, in many situations, proportionality does not seem to be a tenable assumption. To overcome such problems, many authors consider linear models relating the covariates to some transformation of the survival function.

Transformation models for univariate survival data were studied by Cheng *et al.* (1995),

Joseph G. Ibrahim, Department of Biostatistics, University of North Carolina, 3109 McGavran-Greenberg Hall, Campus Box 7420, 27599 Chapel Hill, NC, USA ibrahim@bios.unc.edu.

Supporting information

Additional information for this article is available online including:

Appendix A. Proof of Theorem 1 and

Appendix B. Computational development for the Appendices that are included.

Scharfstein *et al.* (1998), Kosorok *et al.* (2004), and Banerjee *et al.* (2007), among others. The so-called generalized odds-rate hazards (GORH) models (Scharfstein *et al.*, 1998) contain as special cases both the proportional hazards and the proportional odds models.

In this paper, our goal is to carry out Bayesian inference for a class of gamma shared frailty transformation models for multivariate survival data. Frailty models allow one to model the association between survival times from subjects within subgroups. The correlation structure for any pair of times is the same and depends on the frailty parameter. An account of Bayesian methods for frailty models is provided by Ibrahim *et al.* (2001, Chapter 4). For transformation models for multivariate survival data, we mention the recent work by Zeng & Lin (2007) and Zeng *et al.* (2009) dealing with maximum likelihood estimation, whereas Yin (2008) develops Bayesian inference for frailty cure rate models. Unlike Zeng & Lin (2007) and Zeng *et al.* (2009), in our approach, the transformation parameters are estimated within the Bayesian framework. Our paper follows the development of Zeng *et al.* (2009) and can be also seen as a multivariate extension of the work by Banerjee *et al.* (2007). The proposed model formulation is also quite different than the one of Yin (2008). Under the proposed model, the frailty multiplies the transformed cumulative hazard function and, using the Laplace transform, the marginal survival function is available in a closed form. Thus, as a by-product, the closed form expression of the observed data likelihood function can be easily derived, which facilitates convenient computation of Bayesian model selection criteria such as the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) and the Conditional Predictive Ordinate (CPO) (Ibrahim *et al.*, 2001, Chapter 6). In addition, the proposed model formulation allows us to establish propriety of the posterior distribution with improper priors for the model parameters under mild conditions and to obtain flexible survival functions which can be improper or proper.

Multivariate survival models are quite useful in modeling patient and donor factors which influence outcome in Hematopoietic Stem Cell Transplantation (HSCT), a common therapy for leukemia. Joint modeling the effects of patient characteristics on outcomes such as the time-to-relapse of the primary disease and the time-to-first-incidence of chronic graft-versus-host disease (CGVHD), a major complication of HSCT, is of clinical interest. In this paper, we illustrate the proposed methodology using a dataset of 859 patients reported in Ballen *et al.* (2012).

The remainder of the paper is organized as follows. Section 2 is dedicated to the formulation of the model. The prior distribution is specified and the propriety of the posterior distribution is established in Section 3. In Appendix B, we present the full conditional distributions for the Gibbs sampler. Properties of the proposed methodology are assessed through a simulation study conducted in Section 4. In Section 5, we present an illustrative example using a dataset from a cord blood transplantation study. Some general concluding remarks are given in Section 6.

2 The model

2.1 The model formulation and likelihood function

Let T_{ik} denote the failure time of an event of type k for the i -th subject, for $k = 1, \dots, K$ and $i = 1, \dots, n$. We assume that the study ends at some finite τ . Let w_i , $i = 1, \dots, n$ be unobserved independent frailty variables, which follow a gamma distribution (G) with parameters $(1/\theta, 1/\theta)$, where $\theta > 0$ is the variance of w_i . We assume that $T_{i1}, \dots, T_{iK}$ are independent conditional on w_i . For the gamma frailty model, some dependence measures are presented in Hougaard (2000, Section 7.3). For example, Kendall's τ and the median concordance are given by $1/(1 + 2/\theta)$ and $4(2^{\theta+1} - 1)^{-1/\theta} - 1$, respectively. We consider the transformation $G_k(v) = \log(1 + r_k v)/r_k$ if $r_k > 0$ and $G_k(v) = v$ if $r_k = 0$, where r_k is the transformation parameter. Banerjee *et al.* (2007), Yin (2008) and Zeng *et al.* (2009) discuss some properties of this transformation, which are not repeated here.

Let $\mathbf{x}_i$ denote a $p \times 1$ vector of covariates which does not include an intercept. Given w_i , we assume that the cumulative hazard function for T_{ik} is given by

$$H_{ik}(t|w_i, \mathbf{x}_i) = w_i G_k \{H_k(t) e^{\mathbf{x}_i' \beta_k}\} = w_i \log\{1 + r_k H_k(t) e^{\mathbf{x}_i' \beta_k}\} / r_k, \quad (1)$$

where $H_k(\cdot)$ is a cumulative baseline hazard function and $\beta_k = (\beta_{k1}, \dots, \beta_{kp})'$ denotes the regression coefficient vector, $k = 1, \dots, K$. For $r_k \rightarrow 0$, $G_k(v) = v$, which is the identity transformation, and if $r_k = 1$, we obtain $G_k(v) = \log(1 + v)$. With these two special cases, we recover the well-known proportional hazards (PH) and proportional odds (PO) models, respectively. When $0 < r_k < 1$ or $r_k > 1$, the model does not have an interpretation as clear as the PH and PO models.

Under the specification in (1), the conditional survival function and density function are given, respectively, by

$$S_{ik}(t|w_i, \mathbf{x}_i) = \{1 + r_k H_k(t) e^{\mathbf{x}_i' \beta_k}\}^{-w_i/r_k} \quad (2)$$

and $f_{ik}(t|w_i, \mathbf{x}_i) = h_k(t) e^{\mathbf{x}_i' \beta_k} w_i \{1 + r_k H_k(t) e^{\mathbf{x}_i' \beta_k}\}^{-w_i/r_k - 1}$, where $h_k(t)$ is the baseline hazard function such that $H_k(t) = \int_0^t h_k(u) du$.

Let C_{ik} denote the censoring time. We assume that C_{ik} is independent of T_{ik} and w_i given the covariates $\mathbf{x}_i$. We observe $y_{ik} = \min(T_{ik}, C_{ik})$ and $v_{ik} = I(T_{ik} < C_{ik})$, where v_{ik} is the failure indicator such that $v_{ik} = 1$ if y_{ik} is a failure time and $v_{ik} = 0$ if y_{ik} is right censored.

Henceforth, $\mathbf{D}_c = (n, K, \mathbf{y}, \mathbf{X}, \mathbf{v}, \mathbf{w})$ denotes the complete data, where $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{y}_i = (y_{i1}, \dots, y_{iK})'$, $i = 1, \dots, n$, are independent, $\mathbf{X}$ is the $n \times p$ matrix of covariates with i -th row $\mathbf{x}_i'$, $\mathbf{w} = (w_1, \dots, w_n)'$ and $\mathbf{v}$ follows the notation for $\mathbf{y}$. The complete data likelihood function is given by

$$L(\beta, \lambda, \mathbf{r} | \mathbf{D}_c) \propto \prod_{i=1}^n \prod_{k=1}^K \{h_k(y_{ik}) e^{\mathbf{x}_i' \beta_k} w_i\}^{v_{ik}} \{1 + r_k H_k(y_{ik}) e^{\mathbf{x}_i' \beta_k}\}^{-v_{ik} - w_i/r_k}. \quad (3)$$

As an additional assumption, the baseline hazard function $h_k(\cdot)$ is represented by a piecewise function. First, we construct K sets of finite partitions of the time axis with cut points $0 = s_{k0} < s_{k1} < \dots < s_{k,J_k-1} < s_{k,J_k}$, where $s_{k,J_k-1} = \max(y_{ik} : v_{ik} = 1)$ and $s_{k,J_k} = \tau$, that is, s_{k,J_k-1} is the maximum failure time, $k = 1, \dots, K$. In the results reported in Sections 4 and 5, for $j = 1, \dots, J_k - 1$, we chose the intervals $(s_{k,j-1}, s_{k,j}]$ with approximately the same number of failure times (choice based on the percentiles of the failure times). In the first $J_k - 1$ intervals, we specify a constant hazard λ_{kj} , whereas in the last interval, we assume an exponential decay hazard function $\lambda_{k,J_k} e^{-(t-s_{k,J_k-1})}$ so that the cumulative hazard function is bounded (we will return to this point in Section 2.2). The choice of s_{k,J_k-1} implies that no failures occur in the last interval, meaning that the jumps in the baseline hazard function $(\lambda_{k1}, \dots, \lambda_{k,J_k-1})$ correspond to the failure times. The form of the hazard function in the last interval can be different from exponential decay. We also define an interval indicator δ_{ikj} such that $\delta_{ikj} = 1$ if for the k -th event type, the i -th subject failed or was censored in the j -th interval, and $\delta_{ikj} = 0$ otherwise. Therefore, from (3), it follows that

$$\begin{aligned} L(\beta, \lambda, r | D_c) &\propto \prod_{i=1}^n \prod_{k=1}^K \prod_{j=1}^{J_k} \{ \lambda_{kj} e^{\mathbf{x}'_i \beta_k} w_i e^{-\delta_{ikj} J_k (y_{ik} - s_{k,J_k-1})} \}^{\delta_{ikj} \nu_{ik}} \\ &\times \{ 1 + r_k \delta_{ikj} [\lambda_{kj} (y_{ik} - s_{k,j-1})^{1-\delta_{ikj} J_k} (1 - e^{-(y_{ik} - s_{k,J_k-1})})^{\delta_{ikj} J_k} \\ &+ \sum_{g=1}^{j-1} \lambda_{kg} (s_{kg} - s_{k,g-1})] e^{\mathbf{x}'_i \beta_k} \}^{-\delta_{ikj} \nu_{ik} - w_i / r_k}. \end{aligned} \quad (4)$$

After integrating out $\mathbf{w}$ in (4) with respect to the density $\pi(\mathbf{w} | \theta)$, we obtain the likelihood function for the observed data $D_o = (n, K, \mathbf{y}, \mathbf{X}, \mathbf{v})$ as

$$\begin{aligned} L(\beta, \lambda, \theta, r | D_o) &\propto \prod_{i=1}^n \left\{ \prod_{k=1}^K \left[\prod_{j=1}^{J_k} \{ \lambda_{kj} e^{\mathbf{x}'_i \beta_k} e^{-\delta_{ikj} J_k (y_{ik} - s_{k,J_k-1})} \}^{\delta_{ikj} \nu_{ik}} \right. \right. \\ &\times \{ 1 + r_k \delta_{ikj} [\lambda_{kj} (y_{ik} - s_{k,j-1})^{1-\delta_{ikj} J_k} (1 - e^{-(y_{ik} - s_{k,J_k-1})})^{\delta_{ikj} J_k} \\ &+ \sum_{g=1}^{j-1} \lambda_{kg} (s_{kg} - s_{k,g-1})] e^{\mathbf{x}'_i \beta_k} \}^{-\delta_{ikj} \nu_{ik}} \left. \right] (\theta^{-1})^{\theta^{-1} \frac{\Gamma(\theta^{-1} + \nu_{i.})}{\Gamma(\theta^{-1})}} \\ &\times \left[\theta^{-1} + \sum_{k=1}^K \sum_{j=1}^{J_k} \log \{ 1 + r_k \delta_{ikj} [\lambda_{kj} (y_{ik} - s_{k,j-1})^{1-\delta_{ikj} J_k} (1 - e^{-(y_{ik} - s_{k,J_k-1})})^{\delta_{ikj} J_k} \right. \right. \\ &\left. \left. + \sum_{g=1}^{j-1} \lambda_{kg} (s_{kg} - s_{k,g-1})] e^{\mathbf{x}'_i \beta_k} \} / r_k \right]^{-\theta^{-1} - \nu_{i.}} \right\}, \end{aligned} \quad (5)$$

where $\nu_{i.} = \sum_{k=1}^K \nu_{ik}$, $i = 1, \dots, n$.

The piecewise specification for the baseline hazard function can accommodate different shapes of the underlying hazard function. The degree of complexity of the model represented by (5) is controlled by the number of intervals J_k . The larger the number of intervals, the more complex the model is. In practice, we suggest doing analyses for some values of $(J_1, \dots, J_K)$ in order to assess the sensitivity of the posterior estimates. Moreover, the selection of $(J_1, \dots, J_K)$ may be guided by measures of model adequacy as we discuss in Sections 4 and 5.

2.2 Cure rate

In many clinical and epidemiological studies, there are subjects who are not susceptible to the occurrence of the event of interest. The proportion of such subjects is often called the cure rate or cure fraction. Models for survival data with a cure rate, also known as long-term survival models, have attracted the attention of an increasing number of researchers and practitioners in survival analysis, as exemplified by the work of Maller & Zhou (1996), Yakovlev & Tsodikov (1996), Ibrahim *et al.* (2001), Tsodikov *et al.* (2003), Nieto-Barajas & Yin (2008), Yin (2008), and Kim *et al.* (2011), to name just a few. As in Zeng *et al.* (2009), by allowing τ to be infinite, our general model encompasses cure rate models. The argument can be seen as follows. We rewrite (2) as

$S_{ik}(y|w_i, \mathbf{x}_i) = \{1 + r_k F_k(y) e^{\beta_{0k}(\tau) + \mathbf{x}_i' \beta_k}\}^{-w_i/r_k}$, where $F_k(y) = H_k(y)/H_k(\tau)$ and $\beta_{0k}(\tau) = \log\{H_k(\tau)\}$, so that $F_k(\cdot)$ acts like a cumulative distribution function (cdf). After integrating out w_i , the marginal survival function for the k -th event type is

$S_{ik}(y|\mathbf{x}_i) = [1 + \theta \log\{1 + r_k F_k(y) e^{\beta_{0k}(\tau) + \mathbf{x}_i' \beta_k}\}/r_k]^{-\theta^{-1}}$. Under the piecewise baseline hazard function, we have

$$\beta_{0k}(\tau) = \log \left\{ \sum_{j=1}^{J_k-1} \lambda_{kj} (s_{kj} - s_{k,j-1}) + \lambda_{k J_k} (1 - e^{-(\tau - s_{k, J_k-1})}) \right\}.$$

Therefore, for each event type k and given $\mathbf{x}$, the cure rate, denoted by ρ_k , is given by

$$\rho_k = \lim_{\tau \rightarrow \infty} S_k(\tau|\mathbf{x}) = \frac{1}{[1 + \theta \log\{1 + r_k e^{\beta_{0k}(\infty) + \mathbf{x}' \beta_k}\}/r_k]^{-\theta^{-1}}}, \quad (6)$$

where $\beta_{0k}(\infty) = \log\{\sum_{j=1}^{J_k-1} \lambda_{kj} (s_{kj} - s_{k,j-1}) + \lambda_{k J_k}\}$, meaning that the marginal survival function is improper. The following proposition summarizes some properties of the cure rate.

Proposition 1—The cure rate in (6) has the following properties:

- i. The cure rates for the PH and PO models are $\rho_k = \{1 + \theta e^{\beta_{0k}(\infty) + \mathbf{x}' \beta_k}\}^{-\theta^{-1}}$ and $\rho_k = [1 + \theta \log\{1 + e^{\beta_{0k}(\infty) + \mathbf{x}' \beta_k}\}]^{-\theta^{-1}}$, respectively.
- ii. If $\lambda_{kj} \rightarrow \infty$ for some j , then $\rho_k \rightarrow 0$.
- iii. If $\lambda_{kj} \rightarrow 0$ for all j , then $\rho_k \rightarrow 1$.

From (6), the proof of Proposition 1 is straightforward. In Proposition 1, (ii) and (iii) imply that high (low) hazard ratios correspond to low (high) cure rates, as expected.

In the frequentist framework of Zeng *et al.* (2009), the estimator for $F_k(\cdot)$ is a cdf with jumps only at the failure times. Nonparametric maximum likelihood estimation is carried out with a constraint on the summation of the estimates of the density function. In our formulation, we do not have this kind of constraint on the parameters of the piecewise hazard ratio function.

The exponential decay expression of the baseline hazard ratio function in the last interval of each partition of the time axis implies that the cumulative baseline hazard function is bounded. Thus, the survival function is improper, as seen above. On the other hand, if $\lambda_{kJ_k} \rightarrow \infty$, the cumulative hazard function is unbounded and the survival function is proper. So, under our approach, different from Yin (2008), we can deal with improper and proper survival functions as well.

3 Prior and posterior

The Bayesian framework requires specification of a prior distribution for $(\beta, \lambda, \theta, r)$. We postulate an improper joint prior distribution in which β, λ, θ and r are independent and β has an improper uniform prior, so that

$$\pi(\beta, \lambda, \theta, r) \propto \prod_{k=1}^K \left\{ \pi(r_k) \prod_{j=1}^{J_k} \pi(\lambda_{kj}) \right\} \pi(\theta), \quad (7)$$

where θ and r_k follow inverse gamma distributions (*IG*) with parameters $(\gamma_{01}, \gamma_{02})$ and $(\gamma_{03k}, \gamma_{04k})$, respectively, and $\lambda_{kj} \sim G(\gamma_{05kj}, \gamma_{06kj})$, $j = 1, \dots, J_k$ and $k = 1, \dots, K$. If $\gamma_{05kj} = \gamma_{06kj} = 0$, we get a Jeffreys-type prior for λ_{kj} . The specification of an improper uniform prior for β is commonly found in the literature. We specify the priors for θ and r_k , $k = 1, \dots, K$, as inverse gamma distributions. Since the frailty follows a gamma distribution in our model, this choice facilitates the computational development in Appendix B. By combining (7) with the likelihood function in (5), the posterior distribution based on the observed data is given by

$$\pi(\beta, \lambda, \theta, r | D_o) \propto L(\beta, \lambda, \theta, r | D_o) \pi(\beta, \lambda, \theta, r). \quad (8)$$

Now, under very mild conditions, we provide a theorem characterizing the propriety of the posterior distribution of $(\beta, \lambda, \theta, r)$ given the data D_o .

Theorem 1

Consider the posterior distribution in (8) with $\gamma_{05kj} = \gamma_{06kj} = 0$ in the prior distribution (7), $j = 1, \dots, J_k - 1$, $k = 1, \dots, K$. Let $\mathbf{X}_k^$ be an $n \times (p + J_k - 1)$ matrix with its i -th row equal to $\nu_{ik} (\delta_{ik1}, \delta_{ik2}, \dots, \delta_{ik, J_k-1}, \mathbf{x}_i')$. Assume that (i) when $\nu_{ik} = 1$, $y_{ik} > 0$, (ii) $\mathbf{X}_k^*$ is of full rank $p + J_k - 1$, $k = 1, \dots, K$, (iii) $\pi(\theta)$ is proper, i.e., $\gamma_{01} > 0$ and $\gamma_{02} > 0$, (iv) $\pi(r_k)$ is proper, i.e., $\gamma_{03k} > 0$ and $\gamma_{04k} > 0$, and (v) $\pi(\lambda_{kJ_k})$ is proper, i.e., $\gamma_{05kJ_k} > 0$ and $\gamma_{06kJ_k} > 0$, $k = 1, \dots, K$. Then, the posterior distribution in (8) is proper.*

The proof of Theorem 1 is given in Appendix A. If the prior distribution of λ_{kj} , $j = 1, \dots, J_k - 1$, is proper, i.e., $\gamma_{05kj} > 0$ and $\gamma_{06kj} > 0$, the steps in the proof are similar. The conditions (i) and (ii) require that all event times be strictly positive, at least one event occurs in each chosen interval $(s_{k,j-1}, s_{kj}]$, and the design matrix $\mathbf{X}$ is of full rank. Such conditions are very often satisfied and are quite easy to check.

In Appendix B, we provide the full conditional distributions needed to carry out the MCMC simulations. The Gibbs sampler is based on the distribution (8), but to ease the

computations, we resort to parameter transformations and introduction of latent variables, as the reason for this is discussed in details in Appendix B.

4 A simulation study

In this section, we conduct a simulation study to compare the fits of the PH and transformation models to data sets created under different conditions. Moreover, we also assess some frequentist properties of the Bayesian model. In the data generation, we first generate n independent $x_{i2} \sim N(0, 1)$ and given x_{i2} , we sample $x_{i1} \sim \text{Bernoulli}(p(x_{i2}))$, where $p(x_{i2}) = \exp(0.5 + 0.3x_{i2}) / \{1 + \exp(0.5 + 0.3x_{i2})\}$, $i = 1, \dots, n$. These values remain fixed throughout the 500 repetitions of the simulations. The true values of the parameters are $\beta_{11} = 1$, $\beta_{12} = -1.5$, $\beta_{21} = -0.5$, $\beta_{22} = 0.5$, and $\theta = 0$ or $\theta = 1$. The failure times are generated from (2) with $H_k(y) = y$, whereas the censoring variable has a uniform distribution on $(0.5\tau, 1.5\tau)$ with $\tau = 10$. We chose $K = 2$ event types. The mean percentage of right censoring in the simulations was about 10% and 26% for event types 1 and 2, respectively. The time intervals for the piecewise baseline hazard functions are computed as described in Section 2.1. In (7), for the priors of θ , $\mathbf{r}$, and $\boldsymbol{\lambda}$, we put $\gamma_{01} = \gamma_{02} = 0.1$, $\gamma_{03} = \gamma_{04} = 1$, $\gamma_{05kj} = \gamma_{06kj} = 0$, $j = 1, \dots, J_k - 1$, and $\gamma_{05kJ_k} = \gamma_{06kJ_k} = 0.001$, $k = 1, 2$. For these priors, both the mean and variance do not exist. After discarding the first 2,000 iterations of the Gibbs sampler, from the next 5,000 iterations, we took a spacing of size 5 leading to 1,000 samples for each parameter.

We first compare the fitting of the proposed model to simulated data sets created under two combinations of (r_1, r_2) . There are many proposed methods for the comparison of models fitted to a given data set and for selecting the one that best fits the data. Two commonly used criteria are based on the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) and the Conditional Predictive Ordinate (CPO) (Ibrahim *et al.*, 2001, Chapter 6). The DIC is built upon the deviance $D(\boldsymbol{\vartheta}) = -2 \log\{L(\boldsymbol{\vartheta}|\mathbf{D}_o)\}$, where $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \theta, \mathbf{r}')'$ and $L(\boldsymbol{\vartheta}|\mathbf{D}_o)$ comes from (5). Using S samples $\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_S$ from the Gibbs sampler output, the DIC is

computed as $\text{DIC} = D(\bar{\boldsymbol{\vartheta}}) + 2p_D$, where $\bar{\boldsymbol{\vartheta}} = \sum_{j=1}^S \boldsymbol{\vartheta}_j / S$ and $p_D = D(\bar{\boldsymbol{\vartheta}}) - D(\bar{\boldsymbol{\vartheta}})$ represents the

effective number of parameters, with $\bar{D}(\boldsymbol{\vartheta}) = \sum_{j=1}^S D(\boldsymbol{\vartheta}_j) / S$. Given a set of candidate models, the model yielding the smallest value of DIC is the one that best fits the data. For

each observation, CPO_i can be approximated as $C\hat{P}O_i = [\{\sum_{j=1}^S 1/L(\boldsymbol{\vartheta}_j|\mathbf{D}_{oi})\} / S]^{-1}$,

where $\mathbf{D}_{oi} = (K, \mathbf{y}_i, \mathbf{x}_i, \mathbf{v}_i)$. An omnibus measure of fit based on the $C\hat{P}O_i$'s is given by the

log-pseudomarginal likelihood (LPML) computed as $LPML = \sum_{i=1}^n \log(C\hat{P}O_i)$. The larger the value of LPML, the better the fit of the model.

In our simulation study, for each model used in the generation of the simulated data sets (true model), we fit the PH, PO, and transformation models. The averages of LPML and DIC over the 500 repetitions of the simulations, as well as the percentage of data sets in which the true model was selected (correctness), are shown in Table 1. Since the baseline hazard function is the same for both event types, we take the same number of intervals in the partition of the time axes. Two values of θ ($\theta = 0$ corresponds to the independence case) and

two sample sizes were considered. When the PH model is the true one, the fit provided by the PH model is better than the fit with the transformation model. Since the PH model is a limiting case of the transformation model, this result is not unexpected. Since in these simulations the results from the PO and transformation models are similar when PO is the true model, correctness rates were computed by comparing the PO and PH models. The results in Table 1 indicate that the data favor the transformation model over PO when PH is the true model. When PO is the true model, PH yields the poorest fit, as expected, and the differences between the PO and transformation models are remarkably small. Since PO corresponds to a model in the interior of the parameter space, whereas PH is on the boundary, this result is not surprising. For most of the cases, as expected, the correctness rates increase with the sample size.

The second part of our simulation study involves point and interval estimation. The transformation model was fitted. The results corresponding to $(J_1, J_2) = (2, 2)$ intervals are reported in Table 2. For both sample sizes the biases are small and, in general, the averages of the posterior standard deviations (SD) are close to the root mean squared error of the posterior means (RMSE). As expected, both SD and RMSE decrease with the sample size. The coverage of the 95% highest posterior density (HPD) intervals lies between 91% and 97%. When $\theta = 0$, the averages of the posterior means of θ are small for both sample sizes. Moreover, the SD and RMSE of β are smaller. Taking into account that the transformation parameters (r_1, r_2) are not kept fixed in our approach, these results suggest that the Bayesian procedure has good properties since the posterior summaries in Table 2 are close to the results of Zeng *et al.* (2009).

Additionally, we also fitted the three models under a misspecified frailty distribution. The data were generated under the PO model and a lognormal frailty with mean = 1 and variance = $\theta = 1$. For the sake of space, the results are omitted here. As in Table 1, the DIC and LPML provide good discrimination between the two models fitted with a wrong distribution for the frailty (gamma distribution). When $(J_1, J_2) = (5, 5)$, the correctness rates with (DIC, LPML) are (96.8, 96.8) and (99.6, 99.8) for $n = 200$ and $n = 400$, respectively. With respect to Table 2, the most remarkable differences are in the estimates of (r_1, r_2) and θ . Under the misspecified frailty distribution and $n = 400$, the estimates of (r_1, r_2) are biased (with the averages of the posterior means of 1.14 and 1.25, respectively) and have greater variability than in Table 2. The large bias in the estimates of θ inflates its RMSE. The estimates of $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$ are unbiased.

5 Analysis of the transplant data

We apply the proposed methodology to a study of 859 patients receiving an umbilical cord blood transplant for leukemia (AML, ALL, or CML) or myelodysplastic syndromes reported in Ballen *et al.* (2012). According to the study protocol, the population consists of White, African-American, Asian/Pacific Islander, and Hispanic patients with ALL, AML, MDS, and CML who received unrelated umbilical cord blood transplants at U.S.A. transplant centers and were reported to the CIBMTR between 1995 and 2006. Two events are considered; namely, the time from transplant to first evidence of chronic graft-versus-host disease (CGVHD) which occurred in 22.0% of the patients and the time from transplant to

relapse or recurrence of the primary disease which occurred in 26.7% of the cases. These events correspond to two complications of Hematopoietic Stem Cell Transplantation (HSCT). The ranges of the failure times from transplant to CGVHD and to relapse are 0.093–3.035 and 0.002–9.893 years, respectively. For both the events, the censoring times range from 0.003 to 12.41 years. In the dataset, 33 patients (3.8%) had both CGVHD and leukemia relapse. Also, it is of clinical interest to know the likelihood that a patient will never relapse or will never develop CGVHD. These are the “cure” fractions arising from a cure model. Figure 1(a) displays the Kaplan-Meier estimates of the marginal survival functions.

We consider the following covariates collected at the time of transplant: Disease (AML $n = 33$, ALL $n = 361$, CML $n = 53$, and MDS $n = 100$); disease status (early $n = 252$, intermediate $n = 386$, advanced $n = 211$); patient age (range 0.27–78.21 years, mean 13.57, $sd=15$); and recipient weight (range 5–181 kg, mean 37 kg, $sd=26$ kg). Our goal is to identify risk factors for either or both of the event times. The transformation and PH models were fitted to this data set. In (7), the priors of θ and λ are as in Section 4. For the transformation model, we consider two types of priors for (r_1, r_2) ; namely, (1) moderate prior: $r_k \sim IG(10, 10)$ and (2) informative prior: $r_k \sim IG(100, 100)$. With the vague prior adopted in Section 4, we observed a lack of convergence of the MCMC chains for some parameters. Results obtained with the PH model are also presented for the sake of comparison. After discarding the first 10,000 iterations of the Gibbs sampler, we used 600,000 iterations with spacing of size 20, thus obtaining 30,000 samples for each parameter. Convergence of the chains was monitored using graphical displays and the test statistic proposed by Geweke (1992). In Table 3, we show the LPML and DIC values for some combinations of the number of intervals. We point out that the transformation model outperforms the PH model regardless of the number of intervals. When the number of intervals increases, the two models yield a similar fit. However, when this number is small, the fit from the transformation model is markedly better. For both models, the best fit is obtained at $(J_1, J_2) = (10, 25)$ intervals. With moderate priors for the transformation parameters and $(J_1, J_2) = (10, 25)$, Figure 1(b) displays the CPO_i 's for the PH and transformation models. We can see that there is support of the transformation model over the PH model.

Table 4 displays some posterior summaries for the parameters with moderate priors for (r_1, r_2) . Under the transformation model, patients with ALL and CML are at the same risk level as patients with AML (reference category for disease) for both time-to-relapse and time-to-CGVHD. Patients having intermediate status disease at transplant are at the same risk level as patients having early status disease (reference category) for both time-to-relapse and time-to-CGVHD. Patients with MDS are at a lower risk level than patients with AML for time-to-relapse. The difference between AML and MDS is not significant for time-to-CGVHD. Older patients have higher risk for time-to-relapse. The effects of the recipient weight at transplant on the risks are significant and in opposite directions. The point estimate and the HPD interval for the frailty variance θ suggest a small variability in the frailty variable. Posterior summaries for the bivariate PH model are also shown in Table 4. For both event times, the estimates of most of the coefficients are smaller (in absolute value)

under the PH model. While the effect of ALL is not significant under the transformation model, by fitting the PH model, we see a significant effect on time-to-relapse. Since the best fit is achieved with the transformation model, this point illustrates that the results from the PH model can be misleading.

By plugging the posterior means of the parameters into (6), we obtain the estimates of the cure rates for all patients. The dot plots of these estimates in Figure 2 highlight more variability in the estimates for time-to-relapse. Since patients with advanced disease status at transplant have a significantly higher risk for time-to-relapse, the cure rates for these patients tend to be smaller, as depicted in Figure 2(a). Results from the transformation model fitted with $(J_1, J_2) = (5, 5)$ intervals show larger standard deviations than the entries in Table 4 for all the parameters. There are also changes in the posterior mean estimates. For example, the estimates of (r_1, r_2) jump from (1.731, 1.278) to (4.833, 3.480). To gain an understanding of the role of the prior distributions for the transformation parameters in this example, we fit a transformation model with $(J_1, J_2) = (10, 25)$ intervals (best model in Table 3) and $IG(100, 100)$ priors for (r_1, r_2) . For this model, DIC and LPML are equal to 1682.5 and -843.1, respectively, so that the differences do not seem to be important when compared to the entries in Table 3.

In all the results reported thus far, for $j = 1, \dots, J_k - 1$ we chose the intervals $(s_{k,j-1}, s_{k,j}]$ as described in Section 2.1. With this setup, condition (ii) in Theorem 1 is satisfied and the Gibbs sampler is stable. Of course, there are other constructions, such as intervals having approximately equal lengths subject to the restriction that at least one failure occurs in each interval. For this data set, most of the failures occur at earlier times (see Figure 1(a)), and therefore such a construction does not seem to be sensible. Indeed, under this construction, the fit is worse when compared to the DIC and LPML values in Table 3.

6 Concluding remarks

In this paper, we develop a Bayesian framework for inference in a general class of frailty models for multivariate survival data. The commonly used proportional hazards and proportional odds models arise as special cases of the logarithmic transformation in our proposed transformation model. Although we mainly focus on the gamma frailty model, one could entertain other types of frailty models as well as other types of censored data settings and transformations. In Section 2, we can envision a more general model in which the some event times may not be available for all the subjects. For example, if for a given subject i , only the times $y_{i2}, \dots, y_{iK}$ are observed, all we have to do is put $y_{i1} = 0$ and $v_{i1} = 0$. Furthermore, we can extend our shared frailty model to a correlated frailty model. As stated by Wienke (2011, Chapter 5), this class of models can be seen as a natural extension of the shared frailty models. In a transformation correlated frailty model, identifiability and efficient Bayesian computation constitute topics of interesting future research.

The interpretation of the regression parameter is an issue pervading transformation models. Quoting Zeng & Lin (2007, p. 559): “we should not confine ourselves to a hazard interpretation, especially when the hazards are not proportional and alternative formulations

lead to more parsimonious models.” When the proportionality of the hazards is not a justifiable assumption, the meaning of the regression parameter is not so important.

We emphasize that in the papers by Yin (2008) and Zeng *et al.* (2009), the transformation parameters in their multivariate models are fixed and model selection is based on some criterion (for example, LPML, DIC, or likelihood function) computed over a grid of values of the transformation parameters. These approaches pose questions on the choice of the grid and on the extra variability due to the estimation of these parameters. In our approach, we estimate the transformation parameters within a Bayesian framework, so that the variation due to their estimation is accounted for. The results given in Sections 4 and 5 illustrate the feasibility of our proposed model.

The dataset in our example has 859 observations from two event times and four covariates. To obtain 50,000 samples from the posterior distribution, the computing time for the transformation model in Table 3 with $J_1 = 10$ and $J_2 = 25$ was about 12 minutes. We ran the program on an Intel i7 processor machine with 8 GB of RAM memory using a GNU/Linux operating system. We believe that the program is scalable and is possible to fit the model to datasets with hundreds of thousands of observations.

Acknowledgments

We would like to thank the Editor, an Associate Editor, and two referees for their valuable comments, which have led to an improved version of the paper. Dr. de Castro’s research was partially supported from CNPq, Brazil. Drs. Chen and Ibrahim’s research was partially supported by the U.S. National Institutes of Health grants. The CIBMTR (Center for International Blood and Marrow Transplant Research), which provided the data, is supported by grants from the U.S. National Cancer Institute (NCI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Institute of Allergy and Infectious Diseases (NIAID). Dr. Klein’s work was also partially supported by these agreements.

References

- Ballen KK, Klein JP, Pedersen TL, Bhatla D, Duerst R, Kurtzberg J, Lazarus HM, LeMaistre CF, McCarthy P, Mehta P, Palmer J, Setterholm M, Wingard JR, Joffe S, Parsons SK, Switzer GE, Lee SJ, Rizzo JD, Majhail NS. Relationship of race/ethnicity and survival after single umbilical cord blood transplantation for adults and children with leukemia and myelodysplastic syndromes. *Biology of Blood and Marrow Transplantation*. 2012; 18:903–912. [PubMed: 22062801]
- Banerjee T, Chen M-H, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Anal.* 2007; 13:241–260. [PubMed: 17401683]
- Bennett S. Analysis of survival data by the proportional odds model. *Stat Med.* 1983; 2:273–277. [PubMed: 6648142]
- Chen, M-H.; Shao, Q-M.; Ibrahim, JG. Monte Carlo methods in Bayesian computation. Springer; New York: 2000.
- Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82:835–845.
- Cox DR. Regression models and life tables (with discussion). *J R Stat Soc Ser B Stat Methodol.* 1972; 34:187–220.
- Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Proceedings of the Fourth Valencia International Meeting*, Peñíscola, 1991; Oxford: Clarendon Press; 1992. p. 169–193.
- Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *J R Stat Soc Ser C Appl Stat.* 1992; 41:337–348.

- Hougaard, P. Analysis of multivariate survival data. Springer; New York: 2000.
- Ibrahim, JG.; Chen, M-H.; Sinha, D. Bayesian survival analysis. Springer; New York: 2001.
- Kim S, Chen M-H, Dey DK. A new threshold regression model for survival data with a cure fraction. *Lifetime Data Anal.* 2011; 17:101–122. [PubMed: 20414804]
- Kosorok MR, Lee BL, Fine JP. Robust inference for proportional hazards univariate frailty regression models. *Ann Statist.* 2004; 32:1448–1491.
- Maller, RA.; Zhou, X. Survival analysis with long-term survivors. Wiley; New York: 1996.
- Nieto-Barajas LE, Yin G. Bayesian semiparametric cure rate model with an unknown threshold. *Scand J Stat.* 2008; 35:540–556.
- Scharfstein DO, Tsiatis AA, Gilbert PB. Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Anal.* 1998; 4:355–391. [PubMed: 9880995]
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol.* 2002; 64:583–639.
- Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. *J Amer Statist Assoc.* 2003; 98:1063–1078.
- Wienke, A. Frailty models in survival analysis. CRC Press; Boca Raton: 2011.
- Yakovlev, AY.; Tsodikov, AD. Stochastic models of tumor latency and their biostatistical applications. World Scientific; Singapore: 1996.
- Yin G. Bayesian transformation cure frailty models with multivariate failure time data. *Stat Med.* 2008; 27:5929–5940. [PubMed: 18618427]
- Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *J R Stat Soc Ser B Stat Methodol.* 2007; 69:507–564.
- Zeng D, Chen Q, Ibrahim JG. Gamma frailty transformation models for multivariate survival times. *Biometrika.* 2009; 96:277–291.

Figure 1.

(a) Kaplan-Meier estimates for the marginal survival functions (solid line: time to relapse, dashed line: time to CGVHD) with tick marks representing the failure times (top side: time to relapse, bottom side: time to CGVHD). (b) CPO plot for the bivariate PH and transformation models with $(J_1, J_2) = (10, 25)$ intervals and moderate priors for (r_1, r_2) .

Figure 2.

Dot plots of the estimates of the cure rate according to the disease status under the bivariate transformation model with $(J_1, J_2) = (10, 25)$ intervals and moderate priors for (r_1, r_2) . (a): time-to-relapse and (b): time-to-CGVHD.

Table 1

Average of DIC, average of LPML, and correctness rates (%) with different numbers of intervals and noninformative priors for (r_1, r_2) (Tr: transformation model).

True model	(J_1, J_2)	Criteria and fitted models							
		DIC			LPML			Correctness	
		PH	PO	Tr	PH	PO	Tr	DIC	LPML
$\theta = 1, n = 200$									
PH ($r_1 \rightarrow 0, r_2 \rightarrow 0$)	(2, 2)	1093.7	1136.6	1104.8	-546.9	-568.0	-552.4	99.2	99.2
	(5, 5)	1081.0	1112.5	1091.6	-540.8	-556.1	-546.1	98.8	98.8
	(10, 10)	1088.6	1114.6	1098.1	-544.8	-557.7	-550.3	96.4	98.8
PO ($r_1 = r_2 = 1$)	(2, 2)	1137.3	1096.1	1097.8	-569.8	-548.2	-549.1	99.8	99.8
	(5, 5)	1111.8	1099.5	1100.3	-556.6	-550.1	-551.0	94.6	94.8
	(10, 10)	1112.3	1104.9	1098.9	-557.0	-553.3	-554.6	91.0	98.0
$\theta = 0, n = 200$									
PH ($r_1 \rightarrow 0, r_2 \rightarrow 0$)	(2, 2)	808.8	911.2	823.6	-404.3	-454.4	-411.1	100.0	100.0
	(5, 5)	825.3	866.4	836.1	-412.7	-432.8	-417.9	99.6	99.4
	(10, 10)	832.5	863.1	842.6	-416.6	-431.6	-421.7	99.8	99.8
PO ($r_1 = r_2 = 1$)	(2, 2)	1192.9	1145.3	1147.7	-597.8	-572.6	-574.0	100.0	100.0
	(5, 5)	1156.5	1141.5	1142.6	-579.1	-570.9	-572.1	94.8	95.4
	(10, 10)	1180.0	1172.4	1172.3	-590.9	-586.7	-587.5	89.2	91.0
$\theta = 1, n = 400$									
PH ($r_1 \rightarrow 0, r_2 \rightarrow 0$)	(2, 2)	2154.9	2239.1	2169.5	-1077.5	-1119.2	-1084.8	99.4	99.4
	(5, 5)	2134.7	2196.9	2148.3	-1067.4	-1098.2	-1074.3	99.0	99.0
	(10, 10)	2187.0	2237.4	2200.0	-1094.0	-1118.7	-1100.5	99.8	99.8
PO ($r_1 = r_2 = 1$)	(2, 2)	2253.5	2164.5	2166.5	-1127.9	-1082.3	-1083.4	100.0	100.0
	(5, 5)	2222.4	2193.9	2195.2	-1111.8	-1097.1	-1098.0	99.8	99.8
	(10, 10)	2195.9	2180.5	2178.9	-1098.5	-1090.7	-1091.6	96.2	96.2
$\theta = 0, n = 400$									
PH ($r_1 \rightarrow 0, r_2 \rightarrow 0$)	(2, 2)	1618.1	1815.9	1635.4	-809.0	-906.8	-817.1	100.0	100.0

True model	(J_1, J_2)	Criteria and fitted models								
		DIC			LPML			Correctness		
		PH	PO	Tr	PH	PO	Tr	DIC	LPML	
PO ($r_1 = r_2 = 1$)	(5, 5)	1632.6	1713.3	1645.9	-816.3	-856.1	-822.7	100.0	99.8	
	(10, 10)	1601.5	1661.2	1613.7	-800.9	-830.4	-806.9	100.0	100.0	
	(2, 2)	2377.4	2275.6	2278.3	-1190.0	-1137.8	-1139.2	100.0	100.0	
	(5, 5)	2305.6	2272.7	2274.6	-1153.6	-1136.4	-1137.7	99.8	100.0	
	(10, 10)	2300.3	2279.7	2280.1	-1150.8	-1140.1	-1141.5	98.6	99.2	

Table 2

Posterior estimates from 500 replications for the transformation model (Par: parameter to be estimated, True: true value of the parameter, Est: average of the posterior means, SD: average of the posterior standard deviations, RMSE: root mean squared error of the posterior means, and CP: coverage probability of the 95% HPD interval).

Par	True	Est	SD	RMSE	CP	Par	True	Est	SD	RMSE	CP
<i>n</i> = 200											
β_{11}	1.0	0.98	0.34	0.36	0.93	β_{21}	-1.5	-1.53	0.34	0.35	0.94
β_{12}	-0.5	-0.50	0.18	0.18	0.95	β_{22}	0.5	0.50	0.18	0.16	0.97
r_1	1	1.01	0.30	0.28	0.95	r_2	1	1.00	0.33	0.30	0.93
θ	1	1.03	0.22	0.21	0.95						
<hr/>											
β_{11}	1.0	1.03	0.29	0.30	0.93	β_{21}	-1.5	-1.58	0.33	0.33	0.96
β_{12}	-0.5	-0.51	0.14	0.14	0.95	β_{22}	0.5	0.52	0.15	0.15	0.94
r_1	1	0.98	0.31	0.29	0.94	r_2	1	1.10	0.44	0.43	0.96
θ	0	0.06	0.04	0.06	–						
<hr/>											
<i>n</i> = 400											
β_{11}	1.0	0.99	0.25	0.24	0.95	β_{21}	-1.5	-1.50	0.25	0.23	0.96
β_{12}	-0.5	-0.49	0.12	0.12	0.96	β_{22}	0.5	0.51	0.12	0.12	0.94
r_1	1	1.00	0.21	0.21	0.96	r_2	1	0.99	0.25	0.22	0.94
θ	1	1.02	0.16	0.16	0.95						
<hr/>											
β_{11}	1.0	1.00	0.20	0.20	0.95	β_{21}	-1.5	-1.52	0.23	0.23	0.95
β_{12}	-0.5	-0.50	0.10	0.10	0.92	β_{22}	0.5	0.51	0.10	0.10	0.96
r_1	1	0.93	0.22	0.23	0.91	r_2	1	1.01	0.31	0.30	0.93
θ	0	0.04	0.03	0.05	–						

Table 3

DIC and LPML for bivariate PH and transformation models with moderate priors for (r_1, r_2) and different number of intervals.

		p_D		DIC		LPML	
		PH	Transformation	PH	Transformation	PH	Transformation
(J_1, J_2)							
(2, 2)		17.0	19.4	2057.6	1912.2	-1031.3	-955.5
(5, 5)		22.5	22.9	1864.1	1815.8	-933.3	-911.1
(10, 25)		47.8	49.1	1685.8	1681.9	-845.0	-843.2
(15, 20)		47.5	49.1	1700.5	1698.2	-851.9	-851.0
(15, 30)		58.1	59.4	1697.5	1695.1	-851.6	-850.4

Table 4

Posterior summaries for the bivariate models with $(J_1, J_2) = (10, 25)$ intervals.

Transformation model with moderate priors for (r_1, r_2)						
Coefficients, r , and θ	Time to relapse			Time to CGVHD		
	Est	SD	95% HPD	Est	SD	95% HPD
Disease ALL	-0.427	0.221	(-0.866, 0.001)	-0.015	0.229	(-0.451, 0.448)
CML	-0.266	0.409	(-1.087, 0.519)	0.350	0.384	(-0.425, 1.082)
MDS	-1.212	0.353	(-1.887, -0.496)	0.164	0.324	(-0.469, 0.797)
Status Intermediate	0.065	0.204	(-0.408, 0.534)	0.204	0.240	(-0.272, 0.664)
Advanced	1.475	0.295	(0.900, 2.053)	0.249	0.282	(-0.294, 0.813)
Age	0.043	0.012	(0.021, 0.067)	-0.008	0.012	(-0.032, 0.014)
Weight	-0.029	0.008	(-0.044, -0.015)	0.013	0.006	(0.001, 0.025)
r	1.731	0.713	(0.627, 3.152)	1.278	0.515	(0.517, 2.264)
θ	0.105	0.064	(0.017, 0.233)			

PH model						
Coefficients and θ	Time to relapse			Time to CGVHD		
	Est	SD	95% HPD	Est	SD	95% HPD
Disease ALL	-0.339	0.162	(-0.660, -0.026)	-0.035	0.181	(-0.382, 0.324)
CML	-0.069	0.289	(-0.640, 0.486)	0.349	0.296	(-0.236, 0.918)
MDS	-1.083	0.268	(-1.628, -0.576)	0.133	0.255	(-0.358, 0.635)
Status Intermediate	0.041	0.182	(-0.325, 0.391)	0.188	0.191	(-0.203, 0.548)
Advanced	1.020	0.184	(0.665, 1.380)	0.186	0.223	(-0.247, 0.624)
Age	0.027	0.008	(0.012, 0.043)	-0.007	0.009	(-0.024, 0.010)
Weight	-0.018	0.005	(-0.029, -0.009)	0.009	0.004	(0.001, 0.018)
θ	0.119	0.072	(0.019, 0.263)			