Biostatistics (2010), **11**, 1, *pp.* 111–126
doi:10.1093/biostatistics/kxp041
Advance Access publication on October 14, 2009

Bayesian random-effects threshold regression with application to survival data with nonproportional hazards

MICHAEL L. PENNELL*
Division of Biostatistics, College of Public Health, The Ohio State University, 320 West 10th Avenue, Columbus, OH 43210, USA
mpennell@cph.osu.edu

G. A. WHITMORE
Department of Mathematics and Statistics, Desautels Faculty of Management, McGill University, Montreal, Canada

MEI-LING TING LEE
Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, MD, USA

SUMMARY

In epidemiological and clinical studies, time-to-event data often violate the assumptions of Cox regression due to the presence of time-dependent covariate effects and unmeasured risk factors. An alternative approach, which does not require proportional hazards, is to use a first hitting time model which treats a subject's health status as a latent stochastic process that fails when it reaches a threshold value. Although more flexible than Cox regression, existing methods do not account for unmeasured covariates in both the initial state and the rate of the process. To address this issue, we propose a Bayesian methodology that models an individual's health status as a Wiener process with subject-specific initial state and drift. Posterior inference proceeds via a Markov chain Monte Carlo methodology with data augmentation steps to sample the final health status of censored observations. We apply our method to data from melanoma patients with nonproportional hazards and find interesting differences from a similar model without random effects. In a simulation study, we show that failure to account for unmeasured covariates can lead to inaccurate estimates of survival probabilities.

Keywords: Bayesian methodology; Nonproportional hazards; Random effects; Survival analysis; Threshold regression; Wiener process.

1. INTRODUCTION

In many clinical and epidemiological studies, there is interest in identifying important predictors of time to onset of disease or death. For example, oncologists are interested in using clinical and pathological

*To whom correspondence should be addressed.

characteristics of subjects to develop a prognostic index for patients. A common approach is to fit a Cox proportional hazards model (Cox, 1972) and use a variable selection technique to find a subset of variables that adequately predict time to the health event (cf. Klein and Moeschberger, 2003, Section 8.7). However, the proportional hazards assumption may not hold due to various reasons. As explained by Hougaard (1991) and Keiding *and others* (1997), nonproportional hazards can be caused by unexplained heterogeneity in the risk or frailty. In other settings, the effects of one or more covariates may indeed vary with time making a proportional hazards model inappropriate; for instance, drugs that are slowly metabolized would not appear to be effective initially but eventually patients on the drug should have lower risk. The problem of extra-heterogeneity may be solved by fitting a shared frailty model (Vaupel *and others*, 1979), though this requires the assumption of proportional hazards conditional on the frailty value. Other common solutions to the nonproportional hazards problem include stratification (cf. Klein and Moeschberger, 2003, Section 9.3) and inclusion of time-dependent covariate effects (e.g. Hastie and Tibshirani, 1990; Tian *and others*, 2005).

An alternative approach, which also does not assume proportional hazards, would be to use a first hitting time (FHT) model. In this methodology, time-to-event data are represented by a stochastic process, which fails once it hits a boundary (Lee and Whitmore, 2006). For instance, a subject's health status may be modeled using a Wiener process resulting in an inverse Gaussian FHT (Cox, 1972). Several authors have used this stochastic process approach for biomedical data (e.g. Lee *and others*, 2000; Aalen and Gjessing, 2001; Aalen *and others*, 2008; Horrocks and Thompson, 2004) as well as engineering (e.g. Doksum and Hoyland, 1992; Whitmore *and others*, 1998; Petit and Young, 1999; Padgett and Tomlinson, 2004; Park and Padgett, 2005) and social science applications (e.g. Lancaster, 1972). In epidemiological and clinical studies, this model is conceptually appealing since death or disease is really the culmination of a series of genetic and physiological events through which a patient's health depreciates until it reaches a threshold.

In an approach known as threshold regression, covariate information is incorporated into FHT models through regression structures for the different parameters of the Wiener process. For instance, the initial state and variance of the Wiener process have been related to covariates using a log-link, and an identity link has been used for the drift (cf. Lee *and others*, 2000, 2004; Aalen and Gjessing, 2001; Aalen *and others*, 2008). An inherent characteristic of these models is that the effects on the hazard vary with time. Hence, the proportional hazards assumption is avoided in a very parsimonious fashion.

A common assumption of most regression models (including threshold regression) is that subjects with the same covariate values have the same expected trajectory. While this may be a reasonable assumption for engineering applications in which the subjects are machines, it does not make sense in human studies as there is often a great deal of heterogeneity in their unmeasured covariates (e.g. genetic factors) which likely affect the different parameters of the Wiener diffusion process. Hence, it makes sense to extend these models to include subject-specific coefficients in order to account for this between-subject heterogeneity. Whitmore (1986) proposed a normal–gamma mixture of inverse Gaussian distributions that accommodates between-subject heterogeneity in the drift and variance of the Wiener process. Similarly, Aalen and Gjessing (2001) proposed a threshold regression model with randomized drift. Although these are reasonable approaches for some applications, in many clinical and epidemiologic settings one would expect considerable heterogeneity in the initial state of the Wiener process as subjects enter the study with different inherent risks of disease. Aalen and Gjessing (2001) discuss an extension of the Wiener process to allow a random starting point, though their model does not allow for separate covariate effects for the initial state and drift.

Most authors have used frequentist methods for inference in threshold regression. Bayesian methods have many well-known advantages over frequentist approaches including the ability to incorporate historical data and to perform exact inferences through the use of Markov chain Monte Carlo (MCMC) methods. A few authors have proposed Bayesian FHT approaches. Petit and Young (1999) used a Bayesian

methodology for inverse Gaussian lifetime data in which the final state of the degradation process is observable. Shubina (2005) proposed a related method that accommodates censored data, though it was in the context of joint inference of a marker process and degradation. Saebo and Almoy (2004) used an FHT approach for competing risks data, and Saebo *and others* (2005) extended the model to account for genetic relationships in the data. In the latter paper, random effects were included in the model for the drift, but the initial state was assumed to be constant across subjects.

In this paper, we present a random-effects threshold regression model and Bayesian methodology for inference. The drift is modeled as a function of a random subject-specific intercept and covariates, while the initial state is modeled as a truncated normal random variate whose location parameter is also dependent upon covariates. Inference proceeds using a hybrid MCMC methodology consisting of Gibbs and rejection sampling steps. In addition, we improve the tractability of our model by using data augmentation to compute values of the final health status of censored observations.

In Section 2, we present our threshold regression model with random-effects implementation. In Section 3, we introduce our MCMC algorithm. In Section 4, we apply our method to mortality data from melanoma patients with nonproportional hazards, and in Section 5, we present results from a simulation study. In Section 6, we discuss our results and future areas of research.

2. Threshold regression model with random effects

2.1 *Stochastic model for health*

In many biomedical applications, the "event" in a time-to-event analysis is really the end product of a series of physiologic changes in a subject. For example, death from cancer occurs once the disease has spread to the vital organs causing them to fail. In these situations, it is natural to describe progression to the event using a latent stochastic process characterizing fluctuations in health. An event occurs once the process reaches a threshold value for a first time; for example, a patient dies when his/her health depreciates to a level from which he/she cannot recover. Hence, the event time can also be thought of as the FHT of the stochastic process. This class of models is appropriately called FHT models.

Several options exist for the stochastic process in FHT models. For example, previous articles have considered the Wiener process (e.g. Lancaster, 1972; Doksum and Hoyland, 1992; Petit and Young, 1999; Lee *and others*, 2000; Padgett and Tomlinson, 2004), the gamma process (e.g. Lawless and Crowder, 2004; Park and Padgett, 2005), and the Ornstein–Uhlenbeck process (Aalen and Gjessing, 2004). In this paper, we consider modeling fluctuations in a patient's health status using a Wiener process. The Wiener process is a realistic model for health status because although patient health will tend to decline with age, on a daily or hourly basis, improvements or depreciations are both possible and are accommodated by the bidirectional movements of the Wiener process (Lee *and others*, 2004).

Let $X_i(t)$ denote the latent health status of subject i at time t, which we model by a Wiener diffusion process with initial state x_{0i}, variance σ^2, and drift μ_i, $i = 1, \ldots, n$. The drift parameter quantifies the rate at which the subject approaches the threshold which triggers the event. In an FHT model, a subject's event time (s_i) is the time at which the Wiener process reaches some threshold level, a. In our model, we assign $a = 0$ that requires the initial state to be positive. As discussed in Chhikara and Folks (1989), the FHT for a Wiener process has an inverse Gaussian distribution with probability density function (pdf)

$$f(s_i|x_{0i}, \mu_i, \sigma^2) = \frac{x_{0i}}{\sqrt{2\pi\sigma^2 s_i^3}} \exp\left\{ -\frac{(x_{0i} + \mu_i s_i)^2}{2\sigma^2 s_i} \right\} \tag{2.1}$$

and survival function

$$\overline{F}(s_i|x_{0i}, \mu_i, \sigma^2) = \Phi\left[\frac{\mu_i s_i + x_{0i}}{\sqrt{\sigma^2 s_i}}\right] - \exp(-2x_{0i}\mu_i/\sigma^2)\Phi\left[\frac{\mu_i s_i - x_{0i}}{\sqrt{\sigma^2 s_i}}\right]. \tag{2.2}$$

If we let D_L denote the event that the threshold is eventually reached, the mean survival time conditional on D_L is

$$E(s_i|x_{0i}, \mu_i, \sigma^2, D_L) = \frac{x_{0i}}{|\mu_i|} \tag{2.3}$$

for $\mu_i \neq 0$.

Although the Wiener process has 3 parameters (x_{0i}, μ_i, and σ^2), the survival function (2.2) is only dependent upon these variables through 2 functions: μ_i/σ and x_{0i}/σ. Hence, for censored event time data, there are only 2 free parameters. Gelfand and Sahu (1999) discuss 2 practical problems caused by nonidentifiability in Bayesian analysis. First, if the prior on the nonidentifiable parameter is not informative, then the parameter will drift to extreme values leading to unstable estimates. Second, if the prior is too informative, there would not be any Bayesian learning. Thus, although a prior could be assigned to each of the 3 parameters of the Wiener process, we do not see a practical advantage over fixing 1 parameter if the prior needs to be very informative to achieve reliable results. Since the initial state and rate are more meaningful in biomedical applications, we choose to model μ_i and x_{0i} and fix σ at 1.

If $\mu_i \leqslant 0$, the Wiener process will eventually reach the threshold with probability 1 and consequently

$$\lim_{s\to\infty} \overline{F}(s|x_{0i}, \mu_i, \sigma^2) = 1.$$

Hence in applications in which death from any cause is the outcome of interest, it makes sense to restrict μ_i to be negative in order to improve efficiency. However, in many biomedical studies the event of interest may not be certain. For instance, one may be interested in death from a specific cause (e.g. cancer) or in time to relapse following treatment. Thus, we allow $-\infty < \mu_i < \infty$ in our models to accommodate a cure rate, meaning that a certain proportion of individuals may never experience the event of interest because they are "cured" (cf., Ibrahim *and others*, 2001, and references therein). For a fixed x_{0i}, σ^2, and $\mu_i > 0$, the expected proportion of cured subjects is $1 - \exp(-2x_{0i}\mu_i/\sigma^2)$.

2.2 *Incorporating covariate information and random effects*

As noted by Aalen and Gjessing (2001), one of the major advantages of threshold regression is its ability to distinguish between 2 types of covariate effects: the effects of covariates on how far the process has advanced prior to the study (i.e. the effects on x_0) and causal effects on the degradation (i.e. the effects on μ). Several authors have used the following regression structures:

$$\begin{aligned} \log(x_{0i}) &= \mathbf{u}_i'\boldsymbol{\alpha}, \\ \mu_i &= \mathbf{z}_i'\boldsymbol{\beta}, \end{aligned} \tag{2.4}$$

where $\mathbf{u}_i = (1, u_{i1}, \ldots, u_{ip})'$ and $\mathbf{z}_i = (1, z_{i1}, \ldots, z_{iq})'$ are covariate vectors and $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)'$ are the respective covariate effects (cf., Lee *and others*, 2000, 2004; Aalen and Gjessing, 2001; Aalen *and others*, 2008). Choice of which covariates to include in $\mathbf{u}_i$ and $\mathbf{z}_i$ requires careful consideration of the application. Some covariates may only have causal effects and hence should only be included in the model for μ, while others may exhibit both types of effects. For example, in a randomized clinical trial, treatment is unrelated to patient health at the time of randomization (x_0) but should affect subsequent changes in health (μ). Age, on the other hand, should be related to patient health prior to follow-up as well as the rate at which health improves or declines.

Another attractive feature of threshold regression is that it models the effects of covariates on clinically relevant parameters. Because the Wiener process represents the latent health process of a subject, the initial health level x_0 and drift μ have immediate clinical interpretations in terms of a subject's health without the need to make indirect reference to the survival or hazard function. For example, if an investigator finds that x_0 is 20% higher for case subjects than control subjects, this effect size is directly meaningful to a clinician.

Although the regression models in (2.4) are reasonable, they do not allow for the presence of unmeasured covariates, which are often present in human studies. As mentioned previously, Aalen and Gjessing (2001) and Saebo *and others* (2005) addressed this issue by incorporating random effects into the model for the drift but not the initial state. Hence, we extend these models to include random effects characterizing the between-subject heterogeneity in both parameters of the Wiener process. Our extension of the model for the drift follows the approach of Aalen and Gjessing (2001) and Saebo *and others* (2005):

$$\mu_i = b_i + \mathbf{z}_i'\boldsymbol{\beta}, \quad \text{where } b_i \sim N(0, \lambda^{-1}) \tag{2.5}$$

and $N(0, \lambda^{-1})$ is the normal distribution with mean zero and variance λ^{-1}. Conjugate priors are assigned to both the regression parameters and the precision of the random effects (λ):

$$\boldsymbol{\beta} \sim N_{q+1}(\boldsymbol{\beta}_{\text{prior}}, \boldsymbol{\Sigma}_\beta),$$

$$\lambda \sim \text{Ga}(\gamma_1, \gamma_2),$$

where $\text{Ga}(\gamma_1, \gamma_2)$ denotes a gamma distribution with mean γ_1/γ_2 and variance γ_1/γ_2^2. If the data are available, $\boldsymbol{\beta}_{\text{prior}}$ may be a point estimate from previous analyses with $\boldsymbol{\Sigma}_\beta$ chosen to reflect one's confidence in this value for the current data. The prior for λ may also reflect values estimated from previous studies, though it is likely that there is less information available on this parameter. Hence, we recommend a diffuse prior over moderate values (e.g. $\text{Ga}(2, 1/2)$).

Similar to our model for the drift, one could extend the above log-link function for x_{0i} to include a random intercept. However, there is no conjugate prior for either the random effects or the regression coefficients in this model. Hence, we use an alternate specification for x_{0i} to improve computational efficiency. Specifically, we assume

$$x_{0i} \sim N_+(\mathbf{u}_i'\boldsymbol{\alpha}, \tau^{-1}), \tag{2.6}$$

where $N_+()$ denotes a normal distribution truncated to the left at 0 and, as in (2.4), $\boldsymbol{\alpha}$ is a set of $p+1$ regression coefficients and $\mathbf{u}_i$ are the corresponding covariate values. These covariates are related to the mean and variance of x_{0i} through the following moment equations for the truncated normal (Barr and Sherrill, 1999):

$$E(x_{0i}|\boldsymbol{\alpha}, \tau) = \mathbf{u}_i'\boldsymbol{\alpha} + \frac{h(\zeta_i)}{\sqrt{\tau}}, \tag{2.7}$$

$$V(x_{0i}|\boldsymbol{\alpha}, \tau) = \tau^{-1}\{1 - h(\zeta_i)[h(\zeta_i) - \zeta_i]\}, \tag{2.8}$$

where $\zeta_i = -\tau^{\frac{1}{2}} \cdot \mathbf{u}_i'\boldsymbol{\alpha}$ and $h(\cdot)$ is the hazard function of the standard normal distribution.

As seen in Section B.2 of the supplementary material (available at *Biostatistics* online, http://www.biostatistics.oxfordjournals.org), the truncated normal prior combines neatly with the likelihood to facilitate an efficient rejection sampling algorithm for x_{0i}. In addition, by assuming

$$\boldsymbol{\alpha} \sim N_{p+1}(\boldsymbol{\alpha}_{\text{prior}}, \boldsymbol{\Sigma}_\alpha)$$

a priori, we can sample directly from the full conditional posterior of $\boldsymbol{\alpha}$ using an auxiliary variable approach; this is explained in more detail in Section B.2 of the supplementary material, available at *Biostatistics* online.

A gamma prior may also be assigned to τ, the inverse of the scale parameter in the truncated normal distribution of x_{0i}, or equivalently the inverse of the variance of the untruncated version of that distribution. However, we have found in simulation studies that the posterior of τ converges on arbitrarily large values suggesting that the data contain little, if any, information about this parameter. Hence, we recommend fixing τ to avoid identifiability problems and recommend moderately small values. Results of sensitivity analyses of τ are discussed in Sections 4 and 5.

3. BAYESIAN COMPUTATION AND INFERENCE

3.1 *Likelihood function*

When a subject's drift is positive or the observation period is relatively short, the event time (s_i) may not be observed. Let T_i denote the censoring time of a subject and d_i be a censoring indicator which equals 1 if a subject experiences the event of interest within the observation period $[0, T_i]$. The censoring time either may be determined by the study design (e.g. end of data collection) or could be due to a random process unrelated to the risk of an event (e.g. loss to follow-up or death by an unrelated event). Let t_i be the observation time for subject i, where

$$t_i = \begin{cases} s_i & \text{if } d_i = 1, \\ T_i & \text{if } d_i = 0. \end{cases}$$

Given $d_i = 1$, t_i follows an inverse Gaussian distribution. Hence, subject i's contribution to the likelihood function is given in (2.1). When $d_i = 0$, there are a couple of different ways to describe subject i's contribution to the likelihood. One approach is to use the survival function of the inverse Gaussian distribution given in (2.2). Although the approach does not cause problems for frequentist methods, in the Bayesian framework, it results in complicated posterior distributions involving the cumulative distribution function of a normal distribution. For example, Saebo and Almoy (2004) and Saebo *and others* (2005) used the survival function in their likelihoods, and as a result none of their full conditional posterior distributions had closed forms.

An alternative approach, which is more tractable in the Bayesian framework, is to use the pdf of the level of the sample path at t_i for censored observations:

$$\begin{aligned} f_0(X_i(t_i) = x_i, t_i < s_i | x_{0i}, \mu_i, \sigma^2) \\ = \frac{1}{\sqrt{2\pi\sigma^2 t_i}} \exp\left\{ -\frac{(x_i - x_{0i} - \mu_i t_i)^2}{2\sigma^2 t_i} \right\} \left[1 - \exp\left(\frac{-2x_i x_{0i}}{\sigma^2 t_i} \right) \right] 1(x_i > 0). \end{aligned} \tag{3.1}$$

A derivation of this pdf is provided by Cox and Miller (1968). Since x_i is unobserved when $d_i = 0$, we treat these values as unknown parameters and sample from their full conditional posterior distributions. This data augmentation method is described in Section 3.2. Hence, the likelihood function for our model is as follows:

$$L(\mathbf{x}_s, \mathbf{x}_0, \boldsymbol{\beta}, \mathbf{b}) = \prod_{i=1}^{n} [f(s_i | x_{0i}, \mu_i, \sigma^2)]^{d_i} [f_0(x_i, t_i < s_i | x_{0i}, \mu_i, \sigma^2)]^{1-d_i}, \tag{3.2}$$

where $\mathbf{x}_s$ is a vector containing the unobserved degradation levels at the censoring time t_i for censored observations, $\mathbf{x}_0 = (x_{01}, \ldots, x_{0n})'$, and $\mathbf{b} = (b_1, \ldots, b_n)'$.

3.2 *Data augmentation*

Shubina (2005) proposed a data augmentation approach for unobserved degradation values in a bivariate model for markers and degradation. In this article, we extend the Shubina approach to our threshold regression model with random effects.

The data augmentation method begins by assigning a prior to x_i, given $d_i = 0$. Although there are many possibilities, Shubina argues that the prior should be noninformative since "incorporating additional information about one set of subjects introduces bias into the posteriors of the parameters of interest." We not only agree with this argument but also believe that it is unlikely that there is any information *a priori* about the degradation values. Hence, we assign an improper, flat prior to the unobserved x_i:

$$\pi(x_i|d_i = 0) \propto 1(x_i > 0).$$

Under our improper prior, the full conditional posterior distribution of $x_i|d_i = 0$ is proportional to $K(x_i)$, the kernel of $f_0(x_i, t_i < s_i|x_{0i}, \mu_i, \sigma^2)$; hence, propriety is automatically assured. Note that the first exponential term in (3.1) is recognizable as a normal kernel. Hence, one could potentially use a $N_+(m_i, \sigma^2 t_i)$ candidate density in a Metropolis–Hastings (M-H) algorithm independence chain (Tierney, 1994), where $m_i = x_{0i} + \mu_i t_i$. Although a reasonable approach, convergence may be slow since, when imbedded in a Gibbs sampling algorithm, the M-H algorithm does not generate draws from the full conditional posterior at each iteration of the MCMC. Hence, we instead use a rejection sampling algorithm, which guarantees that a sample is drawn from the full conditional posterior distribution at each iteration (see Devroye, 1986). The details of our algorithm may be found in Section A of the supplementary material (available at *Biostatistics* online, http://www.biostatistics.oxfordjournals.org).

3.3 *Summary of MCMC algorithm*

Posterior computation proceeds using a hybrid MCMC algorithm comprised of Gibbs and rejection sampling steps. The full conditional posterior distributions of the random initial states (x_{0i}), drift effects (b_i), regression coefficients ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$), and the inverse of the variance of b_i (λ) are provided in Section B of the supplementary material (available at *Biostatistics* online, http://www.biostatistics.oxfordjournals.org). At each iteration of the MCMC, we use rejection sampling to update the degradation levels at censoring times (i.e. the vector $\mathbf{x}_s$) and the x_{0i} and Gibbs sampling to update b_i, $\boldsymbol{\beta}$, and λ. The regression coefficients for the initial state ($\boldsymbol{\alpha}$) are updated using an auxiliary variable approach proposed by Griffiths (2002). We sample from each full conditional posterior for M iterations following a burn-in of M_B iterations to ensure convergence. More details regarding our MCMC methodology are provided in Section B of the supplementary material (available at *Biostatistics* online).

4. Application to mortality data of melanoma patients

4.1 *Description of data*

Drzewiecki and Andersen (1982) examined prognostic factors of survival of patients with malignant melanoma. The data set consists of clinical data from 205 patients treated at the Plastic Surgery Unit in Odense from 1964 to 1973. Of these patients, 57 died from the disease prior to retrieval of their medical records. The remaining 148 observations were considered censored as these patients were either still alive (134 patients) or died from other causes (14 patients).

Using a Cox model, Andersen *and others* (1993, Examples VII.2.5, VII.3.1, and VII.3.4) found that presence of ulceration and tumor thickness (in millimeters) were the main predictors of patient survival. However, diagnostic plots indicated that the hazards of patients with and without ulceration were

converging and hence were not proportional. Andersen *and others* (1993, Example IX.4.3) and more recently Keiding *and others* (1997) fit frailty models to the data and found significant evidence of unexplained heterogeneity in the hazard. Hence, the lack of proportional hazards and the presence of unmeasured covariates provide good motivation for applying our model to these data.

4.2 *Model specification*

In our application, we modeled time to death from melanoma in years since surgery. We assumed that both the drift and the initial state are dependent upon the presence of ulceration and tumor thickness, that is, for subject i

$$\mathbf{u}_i = \mathbf{z}_i = (1, \text{ulceration}_i, \log(\text{thickness}_i))',$$

where $\text{ulceration}_i = 1$ if ulceration was present in subject i, 0 otherwise. Thus,

$$x_{0i} \sim N_+((\alpha_0, \alpha_1, \alpha_2)\mathbf{u}_i, \tau^{-1}),$$

$$\mu_i = b_i + (\beta_0, \beta_1, \beta_2)\mathbf{z}_i.$$

We assigned flat priors to the regression parameters with

$$\boldsymbol{\alpha}_{\text{prior}} = (\log(5), 0, 0)', \quad \boldsymbol{\beta}_{\text{prior}} = (-\log(5)/\bar{t}_1, 0, 0)',$$

$$\Sigma_\alpha = \Sigma_\beta = 10^4 \cdot \mathbf{I}_3,$$

where $\bar{t}_1$ is the average survival time of the 57 uncensored observations. In addition, we assigned a $\text{Ga}(2, 1/2)$ prior to λ and let $\tau = 1$. We ran the MCMC for 100 000 iterations following a 5000 iteration burn-in and saved every 10th iteration to thin the chain.

4.3 *Results*

The marginal posterior distributions for the regression coefficients are provided in Figure 1. The posteriors of α_1, α_2, β_1, and β_2 are all centered on negative values which suggests that the average initial proximity and drift toward death via melanoma are greater for patients with ulceration and thick tumors. There is also strong evidence that tumor thickness has a negative impact on both the initial state and the drift as the posterior probabilities that α_2 and β_2 are less than 0 are greater than 0.93. There is also strong evidence in favor of a negative effect of ulceration on the drift (posterior probability $= 0.95$) but only moderate evidence that it affects the initial state ($\Pr(\alpha_1 < 0|\text{Data}) = 0.812$).

Our results agree somewhat with previous analyses of the data set by Keiding *and others* (1997). In this paper, the authors fit a series of parametric and semiparametric frailty models and in each model found that the hazard increases in the presence of ulceration and thick tumors. However, these frailty models do not tell us the nature of the covariate effects; our model indicates that tumor thickness increases risk by both decreasing health immediately following surgery and increasing the rate at which health declines post-surgery. Ulceration, however, only affects the rate at which health declines.

In order to better understand how ulceration and tumor thickness affect survival, we estimated the survival curves for the average subject (i.e. $x_0 = E(x_0|\mathbf{u}, \boldsymbol{\alpha})$ and $\mu = \mathbf{z}'\boldsymbol{\beta}$) at different levels of the covariates. As seen in Figure 2, when a patient's tumor is 0.663 log(mm) thick (the median value in the population) the survival curves are relatively linear after year 4 and patients with and without ulceration both have reasonable probabilities of survival at the last observation time (0.57 and 0.83, respectively). In contrast, the expected survival curves for patients whose tumor is 1.91 log(mm) in diameter (the 90th

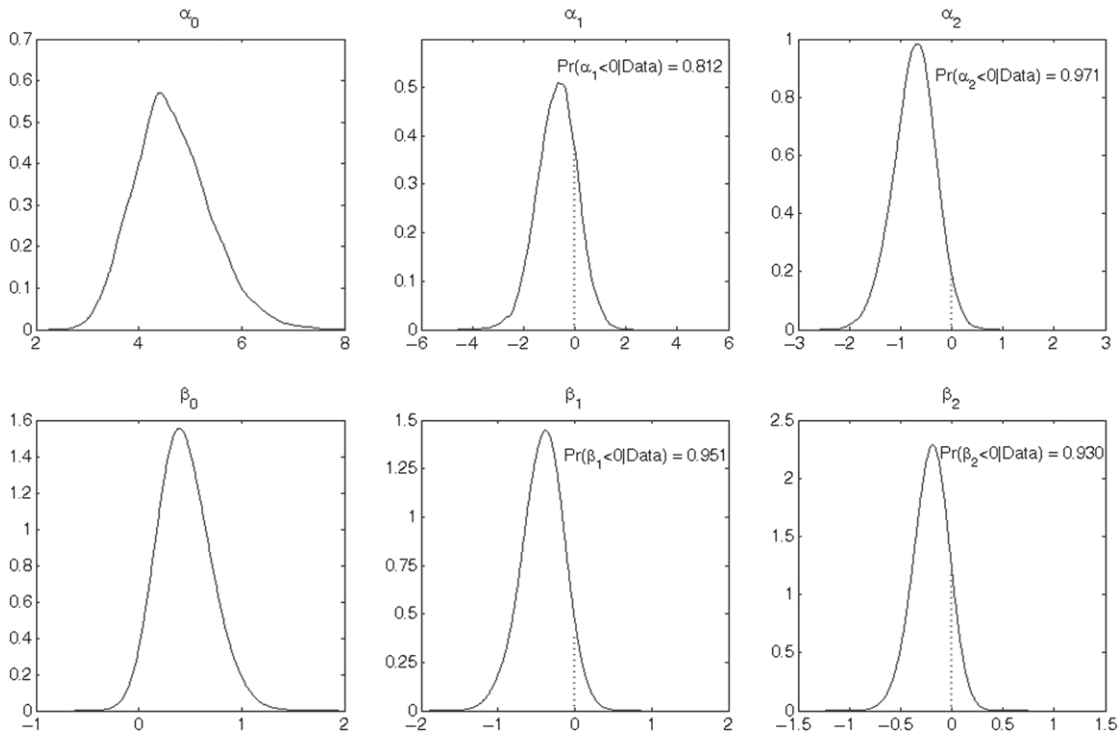

Fig. 1. Kernel-smoothed estimates of the marginal posterior distributions of regression coefficients for the melanoma mortality data.

percentile) are more S-shaped and suggest smaller survival probabilities at the end of the observation period (0.08 when ulceration is present and 0.26 when it is absent).

An assessment of the level of unexplained heterogeneity in the drift may be obtained by looking at the posterior distribution of λ. As seen in Figure 3, the posterior gives low probability to high values of λ, corresponding to low values of the variance of the b_i, which suggests that some of the variability in the drift cannot be explained by the measured covariates. Although the posterior has essentially the same mode as the prior, it is more peaked indicating that the data are informative about λ.

We repeated our analysis assuming $\tau = 0.5$ and $\tau = 2$ to evaluate the effects of this arbitrarily chosen parameter on the results; the results may be found in Section D of the supplementary material (available at *Biostatistics* online, http://www.biostatistics.oxfordjournals.org). To summarize, we found that the posterior probabilities reported in Figure 2 are relatively insensitive to a 2-fold increase or decrease in τ. The survival curves are also very similar across τ, with the differences in survival probabilities never exceeding 0.08 for the cases considered in Figure 3.

We also compared our methodology to the results from the standard maximum likelihood approach for threshold regression assuming no random effects, which we will denote as MLE-NRE (Lee *and others*, 2000, 2004). This approach uses the link functions given in (2.4). As with our model, the MLE-NRE approach suggests that tumor thickness has a significant negative effect on the initial state (p-value $=$ 0.045), while ulceration only has a weak effect (p-value $= 0.178$). In contrast to our model, the MLE-NRE approach found a highly insignificant effect on the drift of both tumor thickness (p-value $= 0.707$) and ulceration (p-value $= 0.495$).

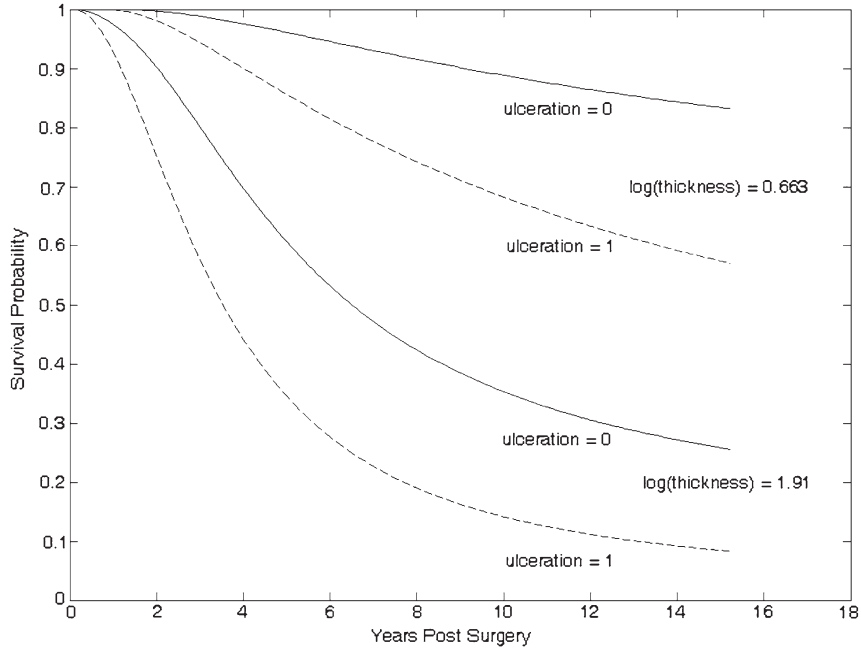

Fig. 2. Posterior means of the survival probabilities of the average patient at each set of covariate values.

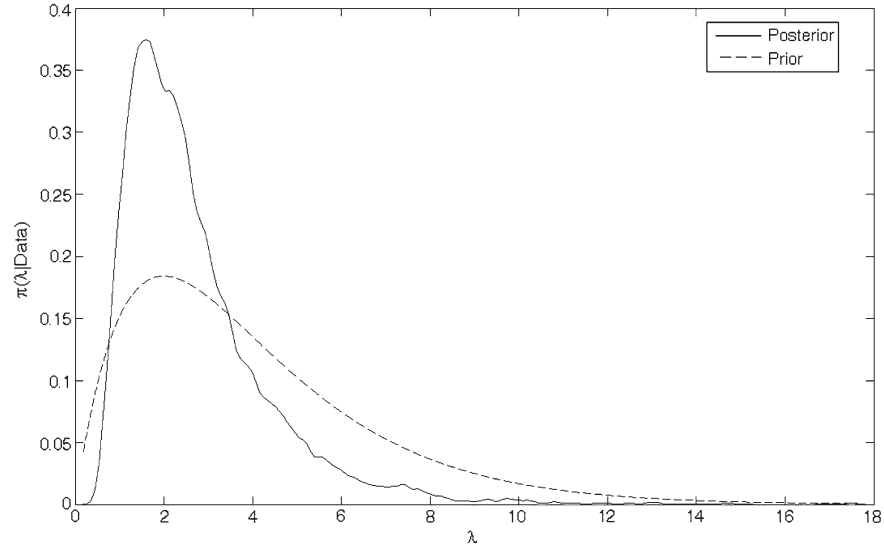

Fig. 3. Kernel-smoothed estimate of the posterior distribution of λ for the melanoma mortality data.

Discernment of the appropriate model (random effects or no random effects) may be made by comparing survival estimates to Kaplan–Meier (KM) plots. However, we cannot restrict our comparisons to patients with specific tumor thicknesses since few are likely to have this value. Thus, in order to compare

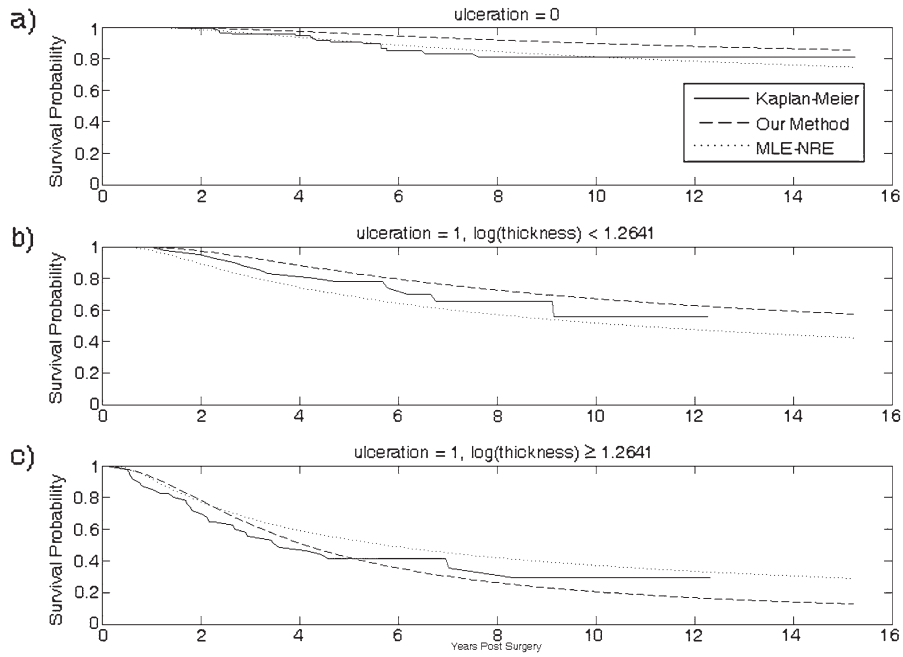

Fig. 4. Comparison of survival estimates of our method, MLE-NRE, and KM for the melanoma mortality data.

the 2 models to the observed data, we created 3 subgroups: patients without ulceration, patients with ulceration and tumor thickness <1.2641 log(mm) (the median for patients with ulceration), and patients with ulceration and tumor thickness $\geqslant 1.2641$ log(mm). Within each group, we averaged the survival curves computed at the 10th, 25th, 50th, 75th, and 90th percentiles of tumor thickness and compared them to the KM estimates; a similar approach was used by Lee *and others* (2000). As seen in Figure 4, both methods appear to fit the data fairly well; MLE-NRE provides a slightly better fit for patients without ulceration, while our method does slightly better for patients with ulcerations and thick tumors.

The observed differences between our approach and MLE-NRE may be explained by extra-heterogeneity in the drift and/or initial status. Arguments may be made for both. Although patients were all treated at the same hospital, admissions were over a 10-year period and patients had a broad range of demographics including sex (61% female, 39% male) and age (range $=$ 4.5–95 years old). Hence, at the time of surgery, the initial health status of subjects probably varied greatly due to differences in genetic susceptibility and other unmeasured risk factors (e.g. presence of latent tumors). In addition, baseline information cannot account for differences in exposures following surgery such as level of UV exposure. Thus, we would also expect there to be substantial heterogeneity in the drift, even after accounting for measured predictors.

In this application, it appears that ignoring between-subject heterogeneity may have a larger effect on the assessment of covariate effects than on prediction of survival. The fact that the significance of the coefficients changed with the addition of random effects is evidence of unmeasured confounders. Ignoring the random effects, the MLEs for the drift coefficients are -0.11 for ulceration and -0.03 for log(tumor thickness). In contrast, the posterior means from our model are -0.42 for the drift coefficient of ulceration and -0.20 for the coefficient of log(tumor thickness). Hence, the random effects appear to be accounting

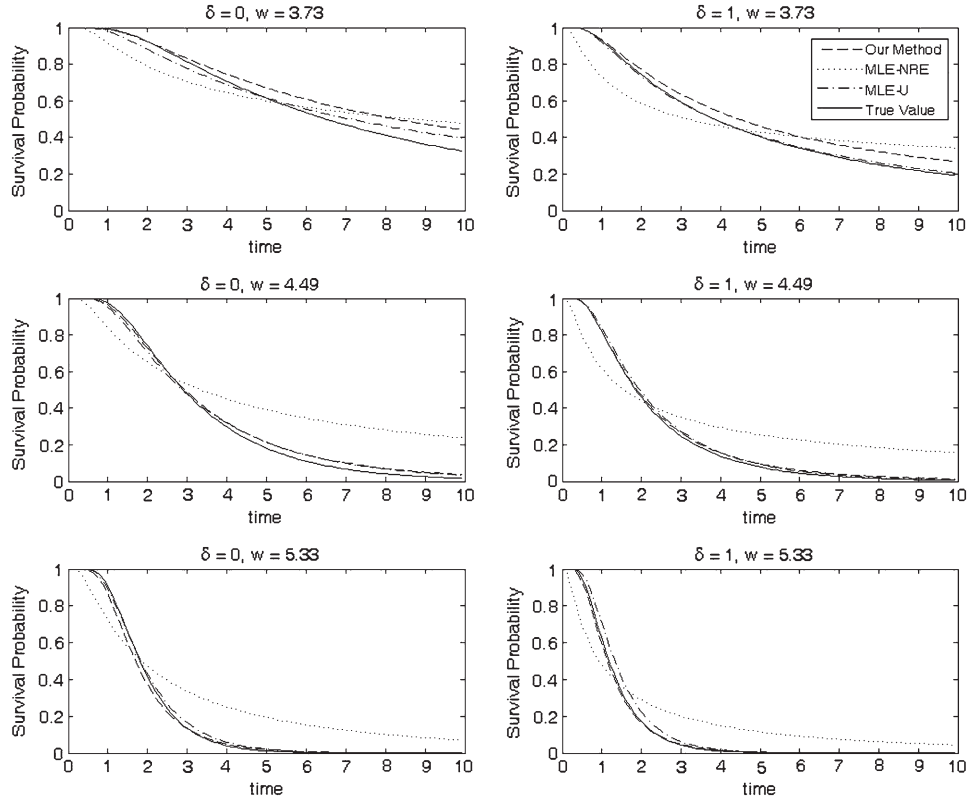

Fig. 5. Comparison of average survival estimates of our method and frequentist methods for the simulated data. Results reported for our method are posterior means for the average patient for each set of covariate values.

for negative confounding: adjusting for the random effects pulls the coefficients further away from zero thereby increasing their significance.

5. SIMULATION STUDY

5.1 *Generation of simulated data*

To further evaluate our method, we applied it to data simulated from a Wiener process. The initial state of each subject was simulated as

$$x_{0i} \sim N_+(3 - \delta_i, 1), \quad \text{where } \delta_i \sim \text{Bernoulli}(0.5),$$

and the drift of each subject was

$$\mu_i = b_i + 2.5 - 0.75w_i, \quad \text{where } b_i \sim N(0, 1) \text{ and } w_i \sim N(4.5, 0.36),$$

where δ_i and w_i are measured covariates. Two hundred sample paths were generated by running steps with small time increments ($\mathrm{d}t = 0.01$) until the cumulative sums were below 0 or the number of steps reached 1000, which is equivalent to truncating at $t = 10$. The resulting data set had 46 censored observations.

5.2 *Model specification*

In our analysis, we assumed that both the drift and the initial state were dependent upon δ and w, that is,

$$\mathbf{u}_i = \mathbf{z}_i = (1, \delta_i, w_i)'.$$

Since the data are similar in structure to the melanoma data, we used the same priors as described in Section 4.2. We ran our MCMC for 50 000 iterations following a 5000 iteration burn-in, otherwise implementing as in Section 4.

5.3 *Simulation results*

The trace plots of $\boldsymbol{\beta}, \boldsymbol{\alpha}$, and λ are provided in Section E of the supplementary material (available at *Biostatistics* online, http://www. biostatistics.oxfordjournals.org). It appeared as if we achieved convergence for each parameter, with λ mixing the poorest. The posteriors of the regression parameters were approximately centered on the true values. The posterior mean of λ (1.4) was greater than the true value, although the true value lay just outside the interquartile range (1.05–1.74). This result is not surprising considering the relatively large prior mean of λ (4) and our moderate sample size.

We also estimated the survival function of the average subject (i.e. $x_0 = E(x_0|\mathbf{u}, \boldsymbol{\alpha})$ and $\mu = \mathbf{z}'\boldsymbol{\beta}$) for different values of δ and w. As seen in Figure 5, our posterior means were close to the true values for the larger values of w. When $w = 3.73$ (the 10th percentile in the data), our approach tended to overestimate survival. This result is also not surprising since more than half the observations with $w \leqslant 3.73$ were censored.

As in our analysis of the melanoma data, we performed a sensitivity analysis on τ. The results are presented in Section E of the supplementary material, available at *Biostatistics* online. We found that posterior probabilities of effect estimates are relatively insensitive to the choice of τ. Survival curves were more sensitive, though the differences from the implementation under $\tau = 1$ never exceeded 0.06.

We next compared our method to MLE-NRE. Assessment of the significance of the effect estimates was consistent with our approach, as seen in Section E of the supplementary material, available at *Biostatistics* online. The survival estimates, however, were quite different. These results are presented in Figure 5. The MLE-NRE approach provided many survival estimates that differed from the true value by 0.2 or more, while our method never differed by more than 0.11 from the true value. These results suggest that failure to adequately account for between-subject heterogeneity in threshold regression can result in highly inaccurate estimates of survival.

As an intermediate between our approach and the MLE-NRE, we also used maximum likelihood to fit a threshold regression model containing the unmeasured covariates (MLE-U), that is, we included b_i as a covariate for the drift and $a_i = \log(x_{0i}) - \log(3 - \delta_i)$ as a covariate for the log of the initial state. Survival estimates from this model are provided in Figure 5. The estimates from this model are closer than our method to the true value when $w = 3.73$. However, the results are very similar for the larger values of w. These findings reflect the dependence of our estimates on the prior; for low values of w, there is a high level of censoring, and due to the limited information, our method relies heavily on the prior to generate random-effects estimates. At larger values of w, there is less censoring and hence the prior is less influential. It should be noted that the 3 methods we considered here form a natural hierarchy. Ideally, one would like the missing covariates a_i and b_i to be known and taken into account with either a frequentist or a Bayesian approach. Without knowledge of these missing covariates, the next best option is to model them as random effects. Finally, an unacceptable solution is to ignore the missing covariates and fit a misspecified model which depends only on the known covariate values (δ and w in this example).

6. DISCUSSION

Threshold regression is an attractive approach for time-to-event data from biomedical studies because it is a conceptually appealing model that does not require proportional hazards. However, as with any regression model, special care is needed when it is applied to heterogeneous populations as there may be unexplained variability in the different parameters of the latent health process. In this paper, we addressed this issue by allowing the initial state and drift to vary across subjects with the same measured covariate values using a tractable Bayesian approach. Our simulation study revealed that failure to account for these random effects can lead to poor survival estimates. In addition, our analysis of the melanoma data suggests that ignoring between-subject heterogeneity may also affect the assessment of covariate effects. We found that survival estimates are somewhat sensitive to the choice of τ. In order to address this sensitivity in applications, one could compare survival estimates under different values of τ and report those which agree best with KM plots.

Computation time is always a concern with Bayesian methods leading many applied statisticians to perform simpler frequentist approaches. Though our method took longer than the MLE approaches described in the paper, we were able to run 105 000 iterations of the MCMC in Matlab (Mathworks, Inc., Natick, MA) in under 3 h using a Dell Latitude D620 laptop (Dell, Austin, TX) with an Intel Centrino Duo processor (Intel, Santa Clara, CA). Hence, computation is certainly tractable and, in many cases, the payoff in terms of flexibility should outweigh time and monetary costs. In addition, sample Matlab code is available upon request from the authors.

Our work opens the door to several areas of future research. For instance, very little work has been done with regard to model selection in threshold regression. In addition, the introduction of random effects to these models raises some interesting questions about whether fixed and/or random effects should be used in modeling the initial state and drift. Potentially, a stochastic search variable selection method (George and McCulloch, 1993) could be developed to answer this question. Another interesting extension of our method would be a nonparametric treatment of the random effects. For example, we could assign Dirichlet process priors (Ferguson, 1973) to the unknown distributions of the random initial state and drift. This work would be related to Bayesian semiparametric frailty models proposed by Walker and Mallick (1997) and Pennell and Dunson (2006).

Another area of future work would be to extend our models to allow for time-varying covariates. Time-varying covariates fit into threshold regression models, such as ours, in a conceptually elegant way. Lee and Whitmore (2006) provide a general framework for modeling the latent health process and covarying processes whereby the health process could be either jointly modeled with the covarying processes or modeled conditional on the covariate processes forming a linear regression structure. In the case of covariate data that are collected longitudinally, Whitmore and Lee also outline an "uncoupling" approach for threshold regression which unlinks the longitudinal data into a series of conditionally independent observations. Hence, a great deal of variety and flexibility exists for modeling time-varying covariates but this is still a topic that needs more attention.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

National Institute of Occupational Safety and Health (OH008649 to M.-L.T.L. and M.L.P.); Natural Sciences and Engineering Research Council of Canada (4032-05 to G.A.W.).

REFERENCES

AALEN, O. O., BORGAN, Ø. AND GJESSING, H. K. (2008). *Survival and Event History Analysis*. New York: Springer.

AALEN, O. O. AND GJESSING, H. K. (2001). Understanding the shape of the hazard rate: a process point of view. *Statistical Science* **16**, 1–22.

AALEN, O. O. AND GJESSING, H. K. (2004). Survival models based on the Ornstein-Uhlenbeck process. *Lifetime Data Analysis* **10**, 407–423.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

BARR, D. R. AND SHERRILL, E. T. (1999). Mean and variance of truncated normal distributions. *American Statistician* **53**, 357–361.

CHHIKARA, R. S. AND FOLKS, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methods, and Applications*. New York: Marcel Dekker.

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

COX, D. R. AND MILLER, H. D. (1968). *The Theory of Stochastic Processes*. London: Chapman and Hall.

DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer.

DOKSUM, K. AND HOYLAND, A. (1992). Models for variable-stress accelerated testing experiments based on Wiener processes and the inverse Gaussian distribution. *Technometrics* **34**, 74–82.

DRZEWIECKI, K. T. AND ANDERSEN, P. K. (1982). Survival with malignant melanoma: a regression analysis of prognostic factors. *Cancer* **49**, 2414–2419.

FERGUSON, T. S. (1973). A Bayesian analysis of some non-parametric problems. *Annals of Statistics* **1**, 209–230.

GELFAND, A. AND SAHU, S. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**, 247–253.

GEORGE, E. I. AND MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

GRIFFITHS, W. (2002). A Gibbs sampler for the parameters of a truncated multivariate normal distribution. *Technical Report*. *Working Paper Series 856*. Victoria, Australia: Department of Economics, University of Melbourne.

HASTIE, T. AND TIBSHIRANI, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46**, 1005–1016.

HORROCKS, J. AND THOMPSON, M. (2004). Modelling event times with multiple outcomes using the Wiener process with drift. *Lifetime Data Analysis* **10**, 29–49.

HOUGAARD, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability* **28**, 695–701.

IBRAHIM, J. G., CHEN, M. AND SINHA, D. (2001). *Bayesian Survival Analysis*. New York: Springer.

KEIDING, N., ANDERSEN, P. K. AND KLEIN, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–224.

KLEIN, J. P. AND MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edition. New York: Springer.

LANCASTER, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society, Series A* **135**, 257–271.

LAWLESS, J. AND CROWDER, M. (2004). Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis* **10**, 213–227.

LEE, M. L. T., DEGRUTTOLA, V. AND SCHOENFELD, D. (2000). A model for markers and latent health status. *Journal of the Royal Statistical Society, Series B* **62**, 747–762.

LEE, M. L. T., GARSHICK, E., WHITMORE, G. A., LADEN, F. AND HART, J. (2004). Assessing lung cancer risk to rail workers using a first hitting time regression model. *Environmetrics* **15**, 1–12.

LEE, M. L. T. AND WHITMORE, G. A. (2006). Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Statistical Science* **21**, 501–513.

PADGETT, W. J. AND TOMLINSON, M. (2004). Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Analysis* **10**, 191–206.

PARK, C. AND PADGETT, W. J. (2005). Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. *Lifetime Data Analysis* **11**, 511–527.

PENNELL, M. L. AND DUNSON, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics* **62**, 1044–1052.

PETIT, L. I. AND YOUNG, K. D. S. (1999). Bayesian analysis for inverse Gaussian lifetime data with measures of degradation. *Journal of Statistical Computing and Simulation* **63**, 217–234.

SAEBO, S. AND ALMOY, T. (2004). Composite hazard functions reflecting competing latent processes. *Biometrical Journal* **46**, 103–113.

SAEBO, S., ALMOY, T. AND AASTVEIT, A. H. (2005). Disease resistance modelled as first-passage times of genetically dependent stochastic processes. *Applied Statistics* **54**, 273–285.

SHUBINA, M. (2005). Bayesian analysis for markers and degradation, [PhD. Thesis]. Boston, MA: Biostatistics Department, Harvard School of Public Health.

TIAN, L., ZUCKER, D. AND WEI, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association* **100**, 172–183.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701–1762.

VAUPEL, J. M., MANTON, K. G. AND STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.

WALKER, S. G. AND MALLICK, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society, Series B* **59**, 845–860.

WHITMORE, G. A. (1986). Normal-gamma mixtures of inverse Gaussian distributions. *Scandinavian Journal of Statistics* **13**, 211–220.

WHITMORE, G. A., CROWDER, M. J. AND LAWLESS, J. F. (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Analysis* **4**, 229–251.

[*Received September 23, 2008; revised September 11, 2009; accepted for publication September 14, 2009*]