Bayesian Analysis for Markers and Degradation

Mei-Ling Ting Lee[12], Maria Shubina[1], and Alan Zaslavsky[13]

[1] Biostatistics Department, Harvard School of Public Health, Boston, USA
[2] Channing Laboratory, Brigham and Women's Hospital, Boston, USA
(e-mail: meiling@channing.harvard.edu)
[3] Department of Health Care Policy, Harvard Medical School, Boston, USA

Abstract. Incorporating marker information into analysis of lifetime data is a topic treated in the current literature in many different ways. In this paper we apply a Bayesian approach to the model introduced by Whitmore, Crowder and Lawless in 1998. In their model they assumed that observable marker process and a latent "true" degradation process together follow a bivariate Wiener process with marker value available at the failure time with censoring. Using data augmentation method for the latent degradation for surviving subjects we construct a full Bayesian model with a closed form posterior distribution. As a sampling procedure we use Metropolis-Hastings within Gibbs algorithm. The model and estimating procedure are applied to a simulated data set from the original article by Whitmore, Crowder and Lawless in order to evaluate the performance of our algorithm. Our method appears to work well, while allowing also to incorporate prior information on the parameters of the model, which can be available from previous studies in similar populations.
Keywords: Marker, Degradation, Latent Models, Bayesian Inference.

1 Introduction

1.1 Markers, degradation and thresholds

Many articles in the literature have focused on incorporating auxiliary information, such as markers, in modelling lifetime data. For good reviews, see Fleming, Prentice, Pepe and Glidden [Fleming *et al.*, 1994], Lefkopoulou and Zelen [Lefkopoulou and Zelen, 1995], Jewell and Kalbfleisch [Jewell and J.D. Kalbfleisch, 1996], Shi, Taylor and Munoz [Shi *et al.*, 1996], among others. On the basis of both proportional and additive hazards models, Lin, Fleming and DeGruttola [Lin *et al.*, 1997] incorporated a time-varying covariate marker as a marker process and considered a variety of models for the marker process.

Another school of thought, represented by the work of Whitmore [Whitmore, 1979], [Whitmore, 1995], Doksum and Hoyland [Doksum and Hoyland, 1992], Doksum and Normand [Doksum and Normand, 1995], Lu [Lu, 1995], and Whitmore and Schenkelberg [Whitmore and Schenkelberg, 1997], considers several models that relate the occurrences of failure events directly to

an observable degradation process. An assumption in many of these models is that an event occurs when observable degradation reaches a threshold. Hence, these models were also referred to as "first-passage time", or "first-hitting time" models (Lee and Whitmore [Lee and Whitmore, 2003]).

Instead of a single observable degradation process to model event occurrences, Whitmore, Crowder and Lawless [Whitmore *et al.*, 1998] (abbreviated herein as WCL) introduced the joint distribution of an observable marker process and an unobservable degradation process. Specifically, they assume that the observable marker process and a latent but unobservable "true" degradation process together follow a bivariate Wiener process. This bivariate model deals with more realistic situations where failure is not deterministically related to an observable marker. The bivariate model also allows us to evaluate the reliability of the observed marker values in the assessment of the latent degradation of a subject. Although most of the earlier papers assume that degradation follows a Wiener process, other forms of degradation processes have recently been considered. Lawless and Crowder [Lawless and Crowder, 2004] used a gamma increment process; and Aalen and Gjessing [Aalen and Gjessing, 2004] modelled survival data using an Ornstein-Uhlenbeck process.

1.2 Motivation of the proposed Bayesian methods

Most of the papers listed above formulated their models and estimation procedures using a conventional frequentist paradigm. On the basis of the univariate degradation model discussed by Lu [Lu, 1995], Pettit and Young [Pettit and Young, 1999] adopt a conventional Bayesian approach. Using uniform priors for the threshold level and proper priors for parameters of degradation, Pettit and Young derived inferences for parameters of both the degradation process and the threshold level to make predictions regarding future events. They used a Gibbs sampler to sample from the posterior distributions of model parameters and estimated predictive distributions of failure times. For a newly enrolled subject, they estimated future degradation levels at different times using estimated parameters averaged over all samples. They applied their methods to a simulated dataset obtained from Lu [Lu, 1995] and compared their results to those obtained by using ML estimators. The paper by Pettit and Young, however, considered only the univariate degradation model.

In this article, we consider the use of Bayesian inference procedures for joint modelling of marker and degradation processes using the bivariate methodology introduced by Whitmore, Crowder, and Lawless (WCL) [Whitmore *et al.*, 1998]. Unlike the conventional Bayesian methods adopted by Pettit and Young, we needed to incorporate the data augmentation technique into our Bayesian models in order to derive the likelihood function in closed form.

We use a full Bayesian approach to make inferences on parameters of both the marker and degradation processes. Also, for surviving subjects, we can model the distribution of residual survival times. For newly enrolled subjects, we can predict their failure times. For subjects expected to survive until a given time with a given marker value, we can predict degradation levels. We applied our model to a simulated dataset from WCL.

2 Short Review of the Bivariate Marker and Degradation Model

The bivariate threshold model introduced by WCL assumes that every subject is represented by a path of a bivariate Wiener process $\mathrm{W}(\tau) = \{X(\tau), Y(\tau)\}$, $\tau > 0$, with initial values $\mathrm{W}(0) = \{X(0), Y(0)\} = \{0, 0\}$, drift $\mu = (\mu_X, \mu_Y)$ with nonnegative μ_X, and covariance matrix $\Sigma = \begin{vmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY} \end{vmatrix}$. The component $X(\tau)$ represents the latent process of an unobservable *degradation* (disease) state of a subject and component $Y(\tau)$ denotes a *marker* process that is correlated with the degradation process $X(\tau)$. The strength of the association between the two components of the bivariate Wiener process is described by the correlation coefficient ρ. The subject fails when the degradation process $X(\tau)$ reaches a failure threshold $a > 0$ for the first time. We denote this first-hitting time by the random variable S. It is well known that, when $X(\tau)$ follows a Wiener process, its first-hitting time S has an inverse Gaussian (IG) distribution with corresponding parameters (see, e.g., Chhikara and Folks[Chhikara and Folks, 1989]).

Each subject is observed during a fixed period of time $[0, t]$ with one of two possible outcomes:
(1) **failing subject** – subject fails at some time $s \in [0, t]$;
(2) **surviving subject** – subject is alive and censored at the time t.

For surviving subjects, the marker component $Y(\cdot)$ is measured at the end of the observation period t. For failing subjects, the marker component $Y(\cdot)$ is measured at the failure time.

As a result, the observed data consists of the following forms.
(1) For **failing** subjects:

i) failure time $S = s < t$,the first-passage time for the degradation X,
ii) value $y = Y(s)$, of the marker component Y at the failure time s ,
iii) $X(s) = a$, since failure is the first–passage to the threshold a;

(2) For **surviving** subjects:

i) time $t < S$, which implies that $X(\tau) < a$ for $\tau \in [0, t]$,
ii) value $y(t)$ of the marker component Y at the time t.

The model also assumes that the latent degradation component has a nonnegative drift $\mu_X \geq 0$ so as to ensure that all subjects will fail within a finite time period with probability 1.

The corresponding probability distributions for failing and surviving subjects were derived in WCL.

3 The Proposed Bivariate Model with Data Augmentation for the Degradation Process

Two approaches are possible to relate survival information to the latent degradation.

i) Consider only the observed data, and obtain a marginal p.d.f. of the marker component for a surviving subject (see [Whitmore *et al.*, 1998], 2.10). This strategy will result in a rather complicated combination of the p.d.f. and c.d.f. of normal distributions with different means and covariances;

ii) Alternatively, one can augment unobserved degradation values for surviving subjects and treat these values as additional parameters.

In this article, we will take the second approach and construct a full Bayesian model based on complete likelihood function.

3.1 Data augmentation

Assume that there are n independent subjects, and each subject could be observed during a fixed time period $[0, T_i]$, $i = 1, \ldots, n$. Let S_i denote the random failure time variable for the ith subject, $i = 1, \ldots, n$. If the ith subject failed at time $S_i = s_i \leq T_i$, the marker value $Y(s_i) = y_i$ is measured. If the ith subject did not fail during the observation period, then the survival time S_i is unobserved, and the marker value is measured at T_i with $Y(T_i) = y_i$. Let $\delta_i = I(S_i < T_i)$, $i = 1, \ldots, n$, be a *censoring* indicator.

We define a stopping time for the ith subject as $t_i = \begin{cases} s_i, \text{ if } \delta_i = 1, \\ T_i, \text{ if } \delta_i = 0. \end{cases}$

Thus, for n subjects, there are three vectors of length n of completely observed data, as follows.

- a vector of stopping times $t = (t_1, t_2, \ldots, t_n)$,
- a vector of censoring indicators $\delta = (\delta_1, \delta_2, \ldots, \delta_n)$,
- a vector $y = (y_1, y_2, \ldots, y_n)$ of values of the Y–component of the Wiener process $\mathrm{W}(\tau)$ at stopping times.

The latent degradation component X is observed only for the failed ($\delta_i = 1$) subjects and is equal to the failure threshold a. Thus the observed data are

$$D_{obs} = (t,\, y,\, \delta,\, (X(t_i) : \delta_i = 1)).$$

To get the "complete" data, we augment the observed data with latent degradation levels for surviving subjects as described below.

The augmented vector of the X–components for all n subjects is defined as $x = (x_1, x_2, \ldots, x_n)$, where $x_i = \begin{cases} a, & \text{if } \delta_i = 1, \\ \text{augmented value } x_i, & \text{if } \delta_i = 0. \end{cases}$

Thus, through data augmentation, we get an additional parameter vector $x_s = \{x_i : \delta_i = 0\}$ of length $n - k$, where $k = \sum_{i=1}^{n} \delta_i$.

3.2 Likelihood function for augmented data

Using the "complete" data consisting of $D = (t,\, x,\, y,\, \delta)$, we can easily derive from WCL (2.5), (2.7) and (2.12), and the condition $\mu_X \geq 0$, the likelihood function for the augmented set of parameters μ, Σ, a, x_s:

$$L\,(\mu, \Sigma, a,\, x_s \mid D_{obs}) = \tag{1}$$
$$\times \prod_{\delta_i=1} p_f(y_i, t_i \mid \mu, \Sigma, a) \prod_{\delta_i=0} p_s(x_i, y_i, t_i \mid \mu, \Sigma, a) I(\mu_X \geq 0).$$

Examination of densities $p_f(\cdot \mid \cdot)$ and $p_s(\cdot \mid \cdot)$ shows that they are overparameterized, and, without loss of generality, we can fix the failure threshold $a = 1$. To simplify notation, we denote the bivariate vector $\mathrm{w}_i = (x_i, y_i),\ i = 1, \ldots, n$. Thus

$$L\ \ (\mu, \Sigma, a,\, x_s \mid D_{obs}) = L(\mu, \Sigma,\, x_s \mid D_{obs}) =$$
$$= \frac{a^k}{(2\pi)^n |\Sigma|^{n/2}} \prod_{i=1}^{n} t_i^{-1-\delta_i} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(\mathrm{w}_i - t_i\mu)\Sigma^{-1}(\mathrm{w}_i - t_i\mu)'}{2t_i} \right) I(\mu_X \geq 0)$$
$$\times \prod_{\delta_i=0} \left(\left(1 - \exp\left(-\frac{2a(a - x_i)}{t_i \sigma_{XX}} \right) \right) I(x_i \leq a) \right). \tag{2}$$

The introduction of augmented latent variables resulted in a likelihood function with a closed form and three groups of parameters, namely a 2–dimensional vector of drift parameters μ, a 2×2 symmetric positive definite covariance matrix Σ, and a $(n - k)$–dimensional vector of augmented degradation values x_s.

4 The Prior and Posterior Distributions

To facilitate a Bayesian inference procedure, we need to specify prior distributions. Note, that conditionally on μ and Σ the prior distribution of augmented vector x_s is fully accounted for by the Wiener process model. We propose to use independent prior distributions for each group of parameters μ and Σ because they are related to different features of the trajectories of the Wiener process: drift describes the average path, whereas variance is responsible for the variability of each individual path. Therefore, the joint prior distribution has the following form:

$$\pi(\mu,\, \Sigma,\, x_s) = \pi(\mu)\pi(\Sigma)\pi(x_s \mid \mu,\, \Sigma). \tag{3}$$

Taking into account that the distribution of x_s is defined by the model, the joint posterior distribution has the form

$$p_{post}(\mu,\, \Sigma,\, x_s \mid D_{obs}) \propto L(\mu, \Sigma,\, x_s \mid D_{obs}) \cdot \pi(\mu)\pi(\Sigma). \tag{4}$$

It could be shown that, under some weak restrictions on observed data, and proper $\pi(\mu)$ and $\pi(\Sigma)$, the joint posterior distribution will be proper.

The proper prior distributions could be made by the choice of hyperparameters to be noninformative or informative, depending on availability of a priori information on marker behavior or/and patient population. Since the main part of the likelihood has a Gaussian form (though truncated for μ_X in the current formulation of the model), we suggest using traditional prior distributions for the class of Gaussian models:

$$\pi(\mu) \propto \exp\left(-\frac{1}{2}(\mu-\mu_0)\Sigma_0^{-1}(\mu-\mu_0)'\right) I(\mu_X \geq 0), \tag{5}$$

$$\pi(\Sigma)\text{simInverse Wish}_2(l, R). \tag{6}$$

5 Predictive distributions of survival times

For surviving subject i, the predictive distribution for residual time s_i^{res} is :

$$p(s_i^{res} \mid D_{obs}) = \iiint p(s_i^{res} \mid \mu, \Sigma, x_i) p_{post}(\mu, \Sigma, x_s \mid D_{obs}) d\mu\, d\Sigma\, dx_s. \tag{7}$$

where $p(s_i^{res} \mid \mu,\, \Sigma,\, x_i) = \dfrac{a - x_i}{\sqrt{2\pi\sigma_{XX}(s_i^{res})^3}} \exp\left(-\dfrac{(a - x_i - \mu_X s_i^{res})^2}{2\sigma_{XX} s_i^{res}}\right).$

The integral over the measure generated by the joint posterior distribution of the parameters can be estimated as a mean of the density $p(s_i^{res} \mid \mu,\, \Sigma,\, x_i)$ over a sample from the joint posterior distribution.

For a newly enrolled subject, the predictive distribution for survival time $p(s \mid D_{obs})$ have the same analytical forms as for residual survival times with initial degradation level $x(0) = 0$. The same estimation procedure applies.

6 Computational Implementation

We designed our computational model to be analytically convenient for the implementation of a Gibbs sampler to draw samples from the joint posterior distribution. We suggest using a form of Gibbs sampling, that allows sampling from conditional distributions for blocks of variables. There are three natural groups of parameters: μ, Σ, and x_s. In section 4 we specified the joint prior distribution for the parameters. From the analytical form of the joint posterior distribution we can see that the conditional posterior distributions for parameters μ and Σ are the products of respective conditional posterior distributions with flat priors and the respective prior distributions. The conditional posterior distribution for the vector of augmented values x_s is a product of conditional posterior distributions of its components.

6.1 Conditional posterior distributions for μ, Σ, and x_s

It can be easily shown that the conditional posterior distribution for μ with a flat improper prior is a truncated ($\mu_X \geq 0$) bivariate normal distribution with location parameter $\mu_{flat} = \bar{\mathrm{w}}/\bar{t}$, covariance parameter $\Sigma_\mu = \frac{1}{nt}\Sigma$, where $\bar{\mathrm{w}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{w}_i$, $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$.

For priors (5) and (6) the conditional posterior distribution for μ is also a truncated ($\mu_X \geq 0$) bivariate normal with location parameter $\mu'_{post} = \Sigma_{\mu_{post}}(\Sigma_\mu^{-1}\mu'_{flat} + \Sigma_0^{-1}\mu'_0)$, covariance parameter $\Sigma_{\mu_{post}} = (\Sigma_\mu^{-1} + \Sigma_0^{-1})^{-1}$.

The convenient way to specify the prior distributions for covariances is to specify them for the inverse matrices. Let denote $\Sigma^{-1} = I\Sigma = \begin{vmatrix} i\sigma_{XX} & i\sigma_{XY} \\ i\sigma_{XY} & i\sigma_{YY} \end{vmatrix}$, so that $\rho = -i\sigma_{XY}^2/\sqrt{i\sigma_{XX}i\sigma_{YY}}$. The kernel for the posterior conditional distribution of $I\Sigma$ with a flat prior can be written as

$$K(I\Sigma) = |I\Sigma|^{\frac{n}{2}} \prod_{\delta_i=0}\Big(1 - \exp\Big(-\frac{2a(a-x_i)i\sigma_{XX}(1-\rho^2)}{t_i}\Big)\Big) \tag{8}$$

$$\times \exp\Big(-\frac{1}{2}\mathrm{tr}\big(I\Sigma \cdot SE\big)\Big), \quad \text{where} \quad SE = \sum_{i=1}^{n}\frac{1}{t_i}(w_i - t_i\mu)'(w_i - t_i\mu).$$

For a Wishart prior for $I\Sigma$ corresponding to (6) with hyperparameters l,and R the kernel for the posterior conditional distribution of $I\Sigma$ is

$$K_{\text{post}}\,(\mathrm{I}\Sigma) = |I\Sigma|^{\frac{n+l-3}{2}} \prod_{\delta_i=0}\left(1 - \exp\Big(-\frac{2a(a-x_i)i\sigma_{XX}(1-\rho^2)}{t_i}\Big)\right) \tag{9}$$

$$\times \ \exp\left(-\frac{1}{2}\mathrm{tr}\left(I\Sigma \cdot R_{\text{post}}^{-1}\right)\right), \quad \text{where} \quad R_{\text{post}} = \left(SE + R^{-1}\right)^{-1}.$$

The kernel for a conditional posterior distribution of the component x_i of the augmented vector x_s for a surviving subject i ($\delta_i = 0$) is

$$K(x_i) = \exp\left(-\frac{(x_i - \mu_{X.Y}(t_i))^2}{2t_i\sigma_{XX.Y}}\right)\left(1 - \exp\left(-\frac{2a(a-x_i)}{t_i\sigma_{XX}}\right)\right) \times I(x_i \leq a),$$

$$\text{where} \ \ \mu_{X.Y} = \mu_X t_i + \sigma_{XY}/\sigma_{YY}(y_i - \mu_Y t_i), \ \sigma_{XX.Y} = \sigma_{XX}(1-\rho^2). \tag{10}$$

6.2 Sampling Schemes for the Conditional Posterior Distributions

As a sampling scheme we propose to use Metropolis-Hastings within Gibbs (MHwG) algorithm, described in Section 6.2 of Chib and Greenberg [Chib and Greenberg, 1995], for three "natural" groups of parameters: μ, Σ, and augmented values x_s of the process component $X(t)$ at the censoring time t. It can be applied to a joint distribution that has one of its conditional

distributions in an analytical form that makes it difficult to develop a direct sampling procedure. The MHwG algorithm is a Gibbs sampler that allows sampling the intractable conditional distribution using a Metropolis-Hastings algorithm, whereas all other conditional distributions are sampled directly. For the joint posterior distribution (4) with suggested priors the covariance matrix of the Wiener process Σ has an intractable conditional posterior distribution, drift parameter μ could be sampled directly. For independent components of vector x_s, we construct a convenient rejection sampling scheme.

7 Analysis of Simulated Data Set

In order to test the performance of our Bayesian scheme, we applied it to the simulated dataset obtained from WCL (see Table 5.1) and compared their ML estimates of parameters μ and Σ to the estimates based on the introduced Bayesian procedure. This dataset was generated by simulating a Wiener process $W = \{X, Y\}$ with parameters $(\mu_X, \mu_Y, \sigma_{XX}, \sigma_{YY}, \rho) = (.1, 1., .4^2, .1^2, .75)$. Fifty sample paths were generated by running steps with time increments $dt = .01$ until the cumulative sums exceeded the threshold level 1 or the number of steps reached 1000, which was equivalent to truncating the paths at time $T = 10$. The generated dataset contains 12 truncated observations.

Since we were interested in comparing our inference procedure to the maximum likelihood estimation of parameters from WCL, we needed to specify a noninformative set of priors. We chose the parameters for the priors (5) and (6) to make them noninformative comparing to the data. It can be shown that the location parameter $\mu_0 = (0, 0)$ and the covariance matrix $\Sigma_0 = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$ for (5), and $l = 3$ with the "covariance" matrix $R = 100.0 \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ with $\rho = 0$ for (6) will be sufficiently noninformative for the WCL dataset. The correlation 0 here corresponds to an a priori hypothesis of no association between the marker and the degradation.

The MHwG algorithm was implemented in S-Plus.

We ran one simulation chain of 12,000 iterations, starting with overdispersed initial values. We considered the first 2000 iterations a "warm-up run" and used the next $10,000$ iterations for inference . The plots of traces of simulated values for parameters μ, augmented latent degradation levels for survivors at the time of censoring, as well as variances and correlation looked rather homogeneous, and allowed us to conclude that the simulation chain has converged.

In Table 1 we present the true parameter values, the ML estimates from WCL, and the values of parameters μ_X, μ_Y, $\sigma_X = \sqrt{\sigma_{XX}}$, $\sigma_Y = \sqrt{\sigma_{YY}}$, and ρ estimated from the simulated Markov chain, We remind that the likelihood function in WCL is the likelihood for the observed data. Bayesian

estimates of parameters are sample means based on all simulated samples, even for parameters with large autocorrelation, because it was shown by S. N. MacEachern and L. M. Berliner [MacEachern and Berliner, 1994] that subsampling leads to less efficient estimates. Numbers in parenthesis represent the standard errors of the estimates based on an estimate of the inverse observed information matrix for ML estimates from WCL, and the estimates of the standard deviations of the posterior distributions of parameters. Based on the full sampled chain we calculated standard deviations as square roots from variance estimates. Median estimations are also based on the full chain. High density intervals are estimated by 250–th and 9750–th respective order statistics.

Table 1. True Values and Estimates for the Parameters of the Process

	True	Estimates				
	Values	from WCL	Bayesian			
Parameter		ML(SD)	Mean(SD)	Median	2.5%	97.5%
μ_X	0.1	0.120(0.023)	0.121(0.022)	0.122	0.077	0.164
μ_Y	1.0	1.012(0.005)	1.012(0.005)	1.012	1.002	1.022
σ_X	.4	0.364(0.039)	0.354(0.036)	0.352	0.289	0.424
σ_Y	.1	0.089(0.008)	0.088(0.008)	0.088	0.075	0.107
ρ	.75	0.737(0.063)	0.721(0.063)	0.729	0.600	0.828

Analysis of autocorrelation functions of parameter samples for μ, σ_{XX}, σ_{YY}, ρ and x_s indicates that samples of location parameters μ and augmented degradation levels x_s are relatively uncorrelated, whereas samples for σ_{XX}, σ_{YY}, and ρ have significant autocorrelations.

To check the stability of the behavior of the samples of σ_{XX}, σ_{YY} and ρ we analyzed subsamples of σ_X, σ_Y and ρ with lags 20 and 50, which are practically uncorrelated. The results are presented in Table 2. It can be

Table 2. Comparison of Estimates σ_X, σ_Y and ρ by subchains

	True	Estimates			
	Values	from WCL	Full Chain	Every 20th	Every 50th
Parameter		ML(SD)	Mean(SD)	Mean(SD)	Mean(SD)
σ_X	.4	0.364(0.039)	0.3539(0.0358)	0.3529(0.0355)	0.3515(0.0344)
σ_Y	.1	0.089(0.008)	0.0882(0.0084)	0.0878(0.0082)	0.0881(0.0088)
ρ	.75	0.737(0.063)	0.7208(0.0633)	0.7206(0.0616)	0.7227(0.0596)

seen that mean values and standard deviations are practically unchanged. Histograms for subsamples, which are not presented here, are also similar to those based on full simulated sample.

References

[Aalen and Gjessing, 2004]O.O. Aalen and H.K. Gjessing. Survival models based on the Ornstein-Uhlenbeck process. *LIDA*, pages 407–423, 2004.

[Chhikara and Folks, 1989]R. S. Chhikara and J. L. Folks. *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, 1989.

[Chib and Greenberg, 1995]S. Chib and E. Greenberg. Understanding Metropolis-Hastings algorithm. *The American Statistician*, pages 238–242, 1995.

[Doksum and Hoyland, 1992]K. Doksum and A. Hoyland. Models for variable-stress accelerated testing experiments based on Wiener processes and inverse Gaussian distribution. *Technometrics*, pages 74–82, 1992.

[Doksum and Normand, 1995]K.A. Doksum and S.-L. Normand. Gaussian models for degradation processes–Part I: Methods for the analysis of biomarker data. *Lifetime Data Analysis*, pages 131–144, 1995.

[Fleming *et al.*, 1994]T.R. Fleming, R.L. Prentice, M.S. Pepe, and D. Glidden. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine*, pages 167–178, 1994.

[Jewell and J.D. Kalbfleisch, 1996]N.P. Jewell and J.D. J.D. Kalbfleisch. Marker process in survival analysis. *Lifetime Data Analysis*, pages 15–29, 1996.

[Lawless and Crowder, 2004]J.F. Lawless and M.J. Crowder. Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, pages 213–227, 2004.

[Lee and Whitmore, 2003]M.-L. T. Lee and G. A. Whitmore. First hitting time models for lifetime data. In N. Balakrishnan and C.R. Rao, editors, *Handbook of Statistics*, pages 537–543, 2003.

[Lefkopoulou and Zelen, 1995]M. Lefkopoulou and M. Zelen. Intermediate clinical events, surrogate markers and survival. *LIDA*, pages 73–85, 1995.

[Lin *et al.*, 1997]D.Y. Lin, T.R. Fleming, and V. DeGruttola. Estimating the proportion of treatment effect explained by surrogate marker. *Statistics in Medicine*, pages 1515–1527, 1997.

[Lu, 1995]J. Lu. *A Reliability Model Based on Degradation and Lifetime Data*. McGill University, Montreal, Canada, 1995.

[MacEachern and Berliner, 1994]S. N. MacEachern and L. M. Berliner. Subsampling the Gibbs sampler. *The American Statistician*, pages 188–190, 1994.

[Pettit and Young, 1999]L.I. Pettit and K.D.S. Young. Bayesian analysis for inverse Gaussian lifetime data with measures of degradation. *J. Statist. Comput. Simul.*, pages 217–234, 1999.

[Shi *et al.*, 1996]M. Shi, J.M.G. Taylor, and A. Munoz. Models for residual time to AIDS. *Lifetime Data Analysis*, pages 31–49, 1996.

[Whitmore and Schenkelberg, 1997]G.A. Whitmore and F. Schenkelberg. Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis*, pages 27–45, 1997.

[Whitmore *et al.*, 1998]G.A. Whitmore, M.J. Crowder, and J.F. J.F. Lawless. Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Analysis*, pages 229–251, 1998.

[Whitmore, 1979]G.A. Whitmore. An inverse Gaussian model for labour turnover. *Journal of the Royal Statistical Society, Series A*, pages 468–478, 1979.

[Whitmore, 1995]G.A. Whitmore. Estimating degradation by a Wiener diffusion process subject to measurement error. *LIDA*, pages 307–319, 1995.