Flexible Modeling of Frailty Effects in Clustered Survival Data

Samuel Manda

Abstract Survival data that have a multivariate structure occur in many health sciences including biomedical research, epidemiology studies and, clinical trials. In most cases, an analysis of multivariate survival data deals with association structures among survival times within same subjects or clusters. Under the conditional (frailty) model approach, the baseline survival functions are modified using mixed effects that incorporates cluster-specific random effects. This approach can be routinely applied and implemented in several statistical software packages with tools to handle analyses of clustered survival data. The random cluster terms are typically assumed to be independently and identically distributed from a known parametric distribution. However, in most practical application, the random effects may change over time, and the assumed parametric random effect distribution could be incorrect. In such cases, nonparametric forms could be used instead. In this chapter, we develop and apply two approaches that assume (a) time dependent random effects and (b) nonparametric random effect distribution. For both approaches, full Bayesian inference using the Gibbs sampler algorithm is used for computing posterior parameter for the mixing distribution and regression coefficients. The proposed methodological approaches are demonstrated using real data sets.

1 Introduction

Modeling and analysis of clustered survival data complicate the estimation procedures since independence between the survival times can no longer be assumed. It has been proven that ignoring clustering effects when an analysis of such data

S. Manda (✉)
Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa

Department of Statistics, University of Pretoria, Pretoria, South Africa

School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
e-mail: samuel.manda@mrc.ac.za

is undertaken could lead to biased estimates of both fixed effect and variance parameters (Heckman and Singer 1984; Pickles and Crouchley 1995; Ha et al. 2001). Random effects models and marginal models are the two most common approaches to modeling clustered survival data. Marginal models focus on the population average of the marginal distributions of multivariate failure times, and the correlation is treated as a nuisance parameter to reduce the dependence of the marginal models on the specification of the unobservable correlation structure of clustered survival data. Parameters are often estimated using generalized estimating equations and the corresponding variance-covariance estimators are corrected properly to account for the possible dependence structure. An excellent overview, with examples, of the marginal approach is discussed in Lin (1994).

On the other hand, random effects models explicitly formulate the underlying dependence in the survival data using random effects. They provide insights into the relationship among related failure times. Conditional on the cluster-specific random effects, the failure times are assumed to be independent. A more general framework for incorporating random effects within a proportional hazards model is given in Sargent (1998) and Vaida and Xu (2000). Random effect survival models are commonly referred to as *shared frailty models* because observations within a cluster share the same cluster-specific effects (Clayton 1991). The nomenclature of frailty effect survival data has its roots from seminal work on univariate survival analyses developed by Vaupel et al. (1979). The frailty effect is taken to collectively represent unknown and unmeasured factors, which affect the cluster-specific baseline risk. The frailty effects can be nested at several clustering levels (Sastry 1997; Bolstad and Manda 2001).

Most of the methodological and analytical developments in the context of frailty models have been based on assuming that the random effects are independent and time-invariant. Within parametric Bayesian hierarchical models (see, for example, Gilks et al. 1996), the usual set-up uses first stage modeling in the observed outcomes. The second stage involves an exchangeable prior distribution on the unobservables (the random frailties), which parameterize the distribution for the observables. However, for some diseases, there may be an increase in the patient's frailty, especially after the first failure event. In such cases, there is need to accommodate the effect of past infection patterns as well as the possibility of time-dependent frailty, for example in recurrent event data. This first part is usually accomplished by a *monitoring risk* variable, which is introduced as a fixed effect covariate. This measures the effect of the deterministic time-dependent component of frailty which can be modeled by the number of prior episodes as in Lin (1994) and Lindsey (1995), or by the total time from study entry as in McGilchrist and Yau (1996) and Yau and McGilchrist (1998). A positive coefficient corresponding to a monitoring risk variable would imply that the rate of infection increases once a first infection has occurred and might indicate serial dependence in the patient's frailty. Alternatively, time-dependent frailty could be modelled using an autoregressive process prior. A simple model uses a stochastic AR(1) prior for the subject (cluster)-specific frailty and residual maximum likelihood estimation procedures have been used (Yau and McGilchrist 1998).

Perhaps a more restrictive assumption is on the parametric model for the frailty effects. For computational convenience, in many applications, the random frailty effects are usually assumed to be independent and identically distributed from a distribution of some known parametric family; in particular, under multiplicative frailty effects, the frailty distribution is from gamma or lognormal families. In practice, information about the distribution of the random effects is often unavailable, and this might lead to poor parameter estimates when the distribution is mis-specified (Walker and Mallick 1997). Furthermore, estimates of covariate effects may show changes in both sign and magnitude depending on the form of frailty distribution (Laird 1978; Heckman and Singer 1984). To reduce the impact of distributional assumptions on the parameter estimates, finite mixture models for frailty effects have been studied (Laird 1978; Heckman and Singer 1984; Guo and Rodriguez 1992; Congdon 1994). A more flexible approach uses nonparametric modelling of frailty effect terms, though this is not used widely in practice (Walker and Mallick 1997; Zhang and Steele 2004; Naskar 2008). Furthermore, asymptotic unbiasedness for estimates of frailty variance depend on the form of the frailty distribution (Ferreira and Garcia 2001). Thus, it is not correct to always assume that the random frailty effects are constant over the study period nor that they arise from a known parametric distribution with its restrictive unimodality and shape. It is important to choose the distribution of the frailty effects to be more flexible in order to account for arbitrary multimodality and unpredictable skewness types (Walker and Mallick 1997). These possibilities can be addressed when the frailty effects distribution is drawn from a large class of distributions. Such a large class could be formed by using nonparametric approaches to model the frailty effect distribution.

In this chapter, we develop methodologies for the analysis of clustered survival where frailty random terms are modelled as time-dependent and nonparametrically. For time-dependent random frailty effect, both the deterministic monitoring risk and stochastic AR(1) model are developed. For the nonparametric frailty model, a Dirichlet process prior is employed (Ferguson 1973). In both frailty effects constructions, parameters are estimated within full Bayesian framework by implementing the Gibbs sampling algorithm in WinBUGS, a statistical software package for Bayesian inference (Spiegelhalter et al. 2004). Both methodologies are applied to example data, and comparisons are made to the constant frailty model using the deviance information criteria (DIC) (Spiegelhalter et al. 2002). Further details on the methods described here could be found in Manda and Meyer (2005) and Manda (2011). We develop the methodology for the nonparametric frailty model. In Sect. 2, we describe the standard conditional survival model and possible extensions. The Dirichlet process prior for nonparametric modelling of the unknown distribution of the frailty effects is presented in Sect. 3, which also describes two constructions of the process: the *Polya urn scheme* and the *stick-breaking* construction. We also describe how the model can be computed using the Gibbs sampler. The child survival data and the resulting parameter estimates are given in Sect. 4. Section 5 presents the proposed time-dependent frailty model for recurrent event analysis, and in Sect. 5.1, we apply the proposed methodology to a data set from Fleming and Harrington (1991) on patients suffering from chronic granulomatous disease. The proposed methodologies are discussed in Sect. 6.

2   Conditional Frailty Survival Model

2.1   Basic Model and Notation

Following Andersen and Gill (1982) (see also Clayton (1991) and Nielsen et al.
(1992)), the basic proportional hazards model is formulated using counting process
methodology. A counting process $N(t)$ $t \geq 0$, is a stochastic process with $N(0) = 0$.
The sample path of $N(t)$ is increasing, piecewise constant, with jumps of size 1 and
right-continuous. Additionally, a stochastic process $Y(t)$ is defined to indicate *alive*
and under observation at time t.

   We extend the counting process methodology to account for clustered events
data. The methodology could easily be adopted to work with recurrent-times
data. Let J be the number of clusters and each has K_j subjects. For subject
jk ($j = 1, \ldots, J$; $k = 1, \ldots, K_j$), a process $N_{jk}(t)$ is observed, which is
a cumulative count of observed events experienced by the subject by time t. In
addition, a process $Y_{jk}(t)$, which indicates whether the subject was at risk for
the event at time t is also observed. We also measure a possibly time-varying p-
dimensional vector of covariates, $x_{jk}(t)$. Thus, for the $(jk)^{th}$ subject the observed
data are $D = \{N_{jk}(t), Y_{jk}(t), x_{jk}(t); t \geq 0\}$, and are assumed independent. Let
$dN_{jk}(t)$ be the increment of $N_{jk}(t)$ over an infinitesimal interval $[t, t + dt)$; i.e.
$dN_{jk}(t) = dN_{jk}[(t + dt)^-] - dN_{jk}(t^-)$; ($t^-$ is time just before t). For right-
censored survival data, the change dN_{jk} takes a value 1 if an event occurred at time
t or 0, otherwise. Suppose F_{t-} is the available data just before time t. Then

$$E[dN_{jk}(t)|F_{t-}] = P(dN_{jk}(t) = 1|F_{t-}) = Y_{jk}(t)h_{jk}(t)dt$$

is the mean increase in $N_{jk}(t)$ over the short interval $[t, t + dt)$, where $h_{jk}(t)$
is hazard function for subject jk. The process $I_{jk}(t) = Y_{jk}(t)h_{jk}(t)$ is called the
intensity process of the counting process. The effect of the covariates on the intensity
function for subject jk at time t is given by the Cox proportional covariate effects
function (Cox 1972)

$$I_{jk}(t|\lambda_0, \beta, x_{jk}(t), w_j) = Y_{jk}(t)\lambda_0(t)w_j e^{\beta^T x_{jk}(t)} \tag{1}$$

where β is a p-dimensional parameter vector of regression coefficients; w_j is the
cluster-specific unobserved frailty, which captures the risk of the unobserved or
unmeasured risk variables; and $\lambda_0(t)$ is the baseline intensity, which is unspecified
and to be modelled nonparametrically. In the present study, the frailty effect w_j is
assumed time-invariant, but this can be relaxed in certain situations (Manda and
Meyer 2005). Under non-informative censoring, the (conditional) likelihood of the
observed data D is proportional to

$$\prod_{j=1}^{J} \prod_{k=1}^{K_j} \prod_{t \geq 0} (I_{jk}(t|\lambda_0(t), \beta, x_{jk}(t), w_j))^{dN_{jk}(t)} e^{-I_{jk}(t|\lambda_0(t), \beta, x_{jk}(t), w_j)dt}. \quad (2)$$

This is just a Poisson likelihood taking the increments $dN_{jk}(t)$ as independent Poisson random variables with means $I_{jk}(t|\lambda_0, \beta, x_{jk}(t), w_j)dt = Y_{jk}(t)w_j \exp(\beta^T x_{jk}(t))d\Lambda_0(t)$, where $d\Lambda_0(t)$ is the increment in the integrated baseline hazard function in interval $[t, t+dt)$. We conveniently model the increment $d\Lambda_0(t)$ by a gamma $d\Lambda_0(t) \sim \text{gamma}(cdH_0(t), c)$ prior, where $H_0(t)$ is a known non-decreasing positive function representing the prior mean of the integrated baseline hazard function and c is the confidence attached to this mean (Kalbfleisch 1978; Clayton 1991). Mixtures of beta or triangular distributions could also be used to nonparametrically model the baseline hazard function (Perron and Mengersen 2001), but this is not done here.

2.2 Prior on the Frailty Distribution

In a nonparametric Bayesian hierarchical structure, prior uncertainty is at the level of the frailty distribution function F (Green and Richardson 2001). One such prior is the Dirichlet process, which models nonparametrically the distribution function F as a random variable. The use of the Dirichlet process prior to model a general distribution function arises from the work of Ferguson (1973). However, the resulting flexibility comes with higher cost due to increased computational complexity of the analysis. A number of algorithms have been proposed recently for fitting nonparametric hierarchical Bayesian models using Dirichlet process mixtures; namely Gibbs sampling, sequential imputations and predictive recursions. In a comparative analysis of these three algorithms using an example from multiple binary sequences, Quintana and Newton (2000) found the Gibbs sampler, though computational intensive, was more reliable. It is the algorithm of choice for studies involving the Dirichlet process mixture to model random effects distributions in linear models (Kleinman and Ibrahim 1998) and in multiple binary sequences (Quintana and Newton 2000).

The Dirichlet process has previously been successfully used by Escobar (1994) and Maceachern (1994) to estimate a vector of normal means. Recently, Dubson (2009) introduced many interesting applications of Bayesian nonparametric priors for inference in biomedical problems. A number of examples provide motivation for non-parametric Bayes methods in bio-statistical applications. These ideas have been expanded upon by (Muller and Quintana 2009) focussing more on inference for clustering. Naskar et al. (2005) and Naskar (2008) suggest Monte Carlo Conditional Expected Maximisation (EM), a hybrid algorithm to analyse HIV infection times in a cohort of females and recurrent infections in kidney patients, respectively, using a Dirichlet process mixing of frailty effects.

3   Nonparametric Dirichlet Frailty Process Frailty Model

3.1   Dirichlet Process Prior

Parametric Bayesian hierarchical modelling uses first stage modelling in the observed outcomes $dN_{11}(t), dN_{12}(t), \ldots, dN_{JK_J}(t)$; $t \geq 0$, and the second stage uses an assumed exchangeable prior, F, usually a gamma or lognormal distribution, on the unobservable $w_1, w_2, \ldots, w_J$. Thus, we have

$$\text{Stage 1}: I_{jk}(t|\lambda_0, \beta, x_{jk}(t), w_j) = Y_{jk}(t)\lambda_0 w_j e^{\beta^T x_{jk}(t)}$$
$$\text{Stage 2}: w_j \sim F$$

A stage where F is allowed to be an unknown random variable is added. Thus, we have

$$\text{Stage 3}: F \sim \text{DP}(M_0, F_0(\gamma))$$

where $\text{DP}(M_0, F_0(.))$ is a Dirichlet process (DP) prior on the distribution function F. The DP prior has two parameters: the function $F_0(.)$, which is one's prior best guess for F, and the scalar M_0, which measures the strength of our prior belief on how well F_0 approximates F. The Dirichlet process prior for distribution function F stems from the work of Ferguson (1973). The property of the process is that for any finite partition $(A_1, \ldots, A_q)$ on the real line R^+, the random vector of prior probabilities $(F(A_1), \ldots, F(A_q))$ has a Dirichlet distribution with parameter vector $(M_0 F_0(A_1), \ldots, M_0 F_0(A_q))$. Using the moments of the Dirichlet distribution, the DP has prior mean $M_0 F_0(0, .]/M_0 F_0(0, \infty) = F_0(.)$ and variance $[M_0^2(1 - F_0(.))F_0(.)]/(M_0 + 1)$.

   A number of useful constructive characterisations of the Dirichlet process (DP) have been proposed in the literature. For instance, Sethuraman (1994) (see also Ishwaran and James (2001)), proposed that the unknown distribution F could be represented as

$$F \sim \sum_{j=1}^{N} \pi_j \delta_{w_j} \tag{3}$$

a random probability measure with N components; $(w_1, w_2, \ldots, w_N)$ are independent and identically distributed random variables with a distribution F_0; $(\pi_1, \pi_2, \ldots, \pi_N)$ are random weights independent of $(w_1, w_2, \ldots, w_N)$. The random weights are chosen using a stick-breaking construction

$$\pi_1 = V_1 \text{ and } \pi_j = (1 - V_1)(1 - V_2) \cdots (1 - V_{j-1})V_j; \; j = 2, \ldots, N \tag{4}$$

where V_j's are independent Beta $(1, M_0)$ random variables. It is necessary to set $V_N = 1$ to ensure that $\sum_{j=1}^{N} \pi_j = 1$. In this paper, we use another important characterisatation of the DP mixture model: the Polya urn and Gibbs sampling inference.

3.2 Polya Urn Sampling Scheme

The Dirichlet process is not very useful for sampling purposes. Blackwell and Macqueen (1973) presented the process as being generated by a Polya urn scheme. In this scheme, w_1 is drawn from F_0 and then the jth subsequent cluster effect w_j is drawn from this mixture distribution:

$$p(w_j | w_1, \ldots, w_{j-1}) = \sum_{k=1}^{m_j} \frac{m_{k,j}^*}{M_0 + j - 1} \delta_{w_{k,j}^*} + \frac{M_0}{M_0 + j - 1} F_0(.)$$

where $\delta_{w_{k,j}^*}$ is a degenerate distribution giving mass 1 to the point $w_{k,j}^*$; $(w_{1,j}^* \ldots, w_{m_j,j}^*)$ are unique set of values in $(w_1, \ldots, w_{j-1})$ with frequency $(m_{1,j}^*, \ldots, m_{m_j,j}^*)$; i.e. $m_{1,j}^* + \ldots + m_{m_j,j}^* = j - 1$. Thus, if M_0 is very large compared to J, little weight is given to previous samples of w_j, implying that the Dirichlet process leads to a full parametric modelling of the random frailty effects by F_0. On the other hand, if M_0 is small, then the process leads to draws from the previous sampled frailty effect.

Polya urn scheme essentially generates samples of w_j from a finite mixture distribution where the components of the mixture are all prior draws $\{w_1, \ldots, w_{j-1}\}$ and $w_1 \sim F_0$ with probabilities described in the Polya urn scheme above. Polya urn sampling is a special case of what is known as a Chinese Restaurant Process (CRP), a distribution on partitions obtained from a process where J customers sit down in a Chinese restaurant with an infinite number of tables (Teh et al. 2005). The basic process specifies that the first customer sits at the first table, then the jth subsequent customer w_j sits at table $w_{k,j}^*$, with probability proportional to the number $m_{k,j}^*$ of customers already seated at the table, otherwise the customer sits at a new table with probability proportional to M_0

Using this representation of the Dirichlet model, the joint prior distribution of $w = (w_1, \ldots, w_J)$ is given by

$$p(w) = \prod_{j=1}^{J} \frac{\sum_{l=1}^{j-1} \delta_{w_l^*} + M_0 f_0(w_j)}{M_0 + j - 1} \tag{5}$$

where δ_b is a degenerate distribution giving mass 1 to the point b, and $f_0(b) = dF_0(b)$.

The Polya urn sampling results in a discrete distribution for a continuous frailty effect by partitioning the J clusters into latent sub-clusters whose members have identical values of w distinct from members in other sub-clusters. The number of latent sub-clusters depends on the parameter value M_0 and the number of the frailty terms (Escobar 1994). This inherent property of partitioning a continuous sample space into discrete clusters, has often led to the Dirichlet process being criticized as selecting a discrete distribution for an otherwise continuous random variable.

3.3   Posterior Distribution Using Polya Urn Scheme

The marginal and joint posterior distributions of the model parameters given the data are obtained from the product of the data likelihood (2); the priors for Λ_0, β, w and the hyperprior for γ. To overcome this computational difficulty, the Gibbs sampler (Tierney 1994), a Markov Chain Monte Carlo (MCMC) method, is used to obtain a sample from the required posterior distribution. The Gibbs sampler algorithm generates samples from the joint posterior distribution by iteratively sampling from the conditional posterior distribution of each parameter, given the most recent values of the other parameters. Let $\pi(.|.)$ denote the conditional posterior distribution of interest, $\Lambda_{0(-t)}$ and w_{-j} be the vectors Λ_0 and w, excluding the element t and j respectively.

The conditional of w_j: From Theorem 1 in Escobar (1994), conditional on the other w and the data, w_j has the following mixture distribution:

$$\pi(w_j|\beta, w_{-j}, N(t), Y(t)) \tag{6}$$

$$\propto \frac{\sum_{l \neq j} L(N_j(t), Y_j(t)|\Lambda_0, \beta, w_l)\delta_{w_l} + M_0 g_0(w_j|\gamma)L(N_j(t), Y_j(t)|\Lambda_0, \beta, w_j)}{M_0 \int L(N_j(t), Y_j(t)|\Lambda_0, \beta, w_j)f_0(w_j|\gamma)dw_j + \sum_{l \neq j} L(N_j(t), Y_j(t)|\Lambda_0, \beta, w_l)}$$

where $L(N_j(t), Y_j(t)|\Lambda_0, \beta, w_j)$ is the sampling distribution of the data in the jth cluster. We note that

$$\pi(w_j|\Lambda_0, \beta, \gamma, N_j(t), Y_j(t)) = f_0(w_j|\gamma)L(N_j(t), N_j(t)|\Lambda_0, \beta, w_j)/\pi(N_j(t), Y_j(t))$$

where $\pi(N_j(t), Y_j(t))$ is the marginal density of $(N_j(t), Y_j(t))$ which, typically, is evaluated by numerical integration. This can be costly when the number J of clusters is large. In order to ease the computational burden, the base measure F_0 is selected to be conjugate to the likelihood of the data. We chose F_0 to be a gamma distribution with mean 1 and variance $\sigma^2 = 1/\gamma$. The marginal distribution of the cluster-specific observed data $(N_j(t), Y_j(t))$ when w_j has the gamma f_0 prior is

$$\pi(N_j(t), Y_j(t)) \propto \int \prod_{k=1}^{K_j} \prod_{t \geq 0} \left[(w_j d\Lambda_0(t)e^{\beta^T x_j(t)})^{dN_j(t)} e^{-Y_j(t)w_i e^{\beta^T x_j(t)} d\Lambda_0(t)} \right]$$

$$w_j^{\gamma-1} e^{-\gamma w_j} dw_j$$

$$= \frac{\Gamma(\gamma + D_j)}{(\gamma + H_j)^{\gamma + D_j}} \prod_{t \geq 0} (d\Lambda_0(t) e^{\beta^T x_j(t)})^{dN_j(t)} \tag{7}$$

where D_j and $H_j = \sum_t Y_j(t) \exp(\beta^T x_j(t)) d\Lambda_0(t)$ are the total number of events and integrated hazard function for cluster j, respectively. Substituting (5) into (4), we have a two-component mixture for the posterior distribution of w_j, which is drawn in the following manner. Using probability proportional to

$$\prod_{k=1}^{K_i} \prod_{t \geq 0} (w_i d\Lambda_0(t) e^{\beta^T x_{ik}(t)})^{dN_{ik}(t)} \exp(-Y_{ik}(t) w_i e^{\beta^T x_{ik}(t)} d\Lambda_0(t))$$

the selection is made from δ_{w_i}, which means that $w_j = w_i$; and with probability proportional to $M_0 \pi(N_j(t), Y_j(t))$ we select w_j from its full conditional posterior distribution $\pi(w_j | \Lambda_0, \beta, \gamma, N_j(t), Y_j(t))$, which is given by

$$\pi(w_j | \Lambda_0, \beta, \gamma, N_j(t), Y_j(t)) \propto F_0(w_j|\gamma) L(N_j(t), Y_j(t) | \Lambda_0, \beta, w_j)$$

$$= \frac{(\gamma + H_j)^{\gamma + D_j}}{\Gamma(\gamma + D_j)} w_j^{\gamma + D_j - 1} e^{-(\gamma + H_j) w_j}$$

a gamma$(\gamma + D_j, \gamma + H_j)$ distribution. This is a mixture of point masses and a gamma distribution. This process draws a new frailty term for cluster j more often from the sampled effects w_l, $l \neq j$ if the observed likelihood for the cluster conditional on the w_l's are relatively large; otherwise the draw is made from its conditional gamma distribution.

The conditional distribution of $d\Lambda_0(t)$ is

$$\pi(d\Lambda_0(t) | \Lambda_{0-t}, \beta, w, N(t), Y(t)) \propto d\Lambda_0(t)^{cdH_0(t)+dN_+(t)-1} e^{-(c+R_+(t|\beta, w_i))d\Lambda_0(t)}$$

where $dN_+(t) = \sum_j^J \sum_k^{K_j} dN_{jk}(t)$ and $R_+(t|\beta, w_i) = \sum_j^J \sum_k^{K_j} Y_{jk}(t) w_j$ $\exp(\beta^T x_{jk}(t))$. This conditional is the gamma$(cdH_0(t) + dN_+(t), c + R_+(t|\beta, w_i))$ distribution, and will be sampled directly.

The conditional posterior distribution of β is

$$\pi(\beta | \Lambda_0, w, \gamma, N(t), Y(t)) \propto \pi(\beta) \prod_{j=1}^{J} \prod_{k=1}^{K_j} \prod_{t \geq 0} w_j (e^{\beta^T x_{jk}(t)})^{dN_{jk}(t)}$$

$$\exp(-Y_{jk}(t) w_j e^{\beta^T x_{jk}(t)} d\Lambda_0(t)) \tag{8}$$

where $\pi(\beta)$ is a prior density for β, commonly assumed to be the multivariate normal distribution with mean zero and the covariance matrix having zero off-diagonal terms. This conditional does not simplify to any known standard density. The vector β can readily be sampled using a Metropolis-Hastings step with a

multivariate Gaussian density centred at the last draw (Tierney 1994). That is, at iteration m, a candidate β^* is generated from $N(\beta^{m-1}, D_\beta)$, and with probability

$$\min\left\{\frac{\pi(\beta^*|\Lambda_0, w, \gamma, N(t), Y(t))}{\pi(\beta^{(m-1)}|\Lambda_0, \beta, w, \gamma, N(t), Y(t))}, 1\right\}$$

the draw $\beta^{(m)}$ is set to β^*, otherwise $\beta^{(m)} = \beta^{(m-1)}$. The covariance matrix D_β may be tuned to allow for an acceptance rate of around 50%.

We model the hyperparameter γ by a hyperprior gamma distribution with known shape κ_1 and scale κ_2. Its conditional distribution uses the $J' \leq J$ distinct values $w^* = (w_1^*, \ldots, w_{J'}^*)$, which are regarded as a random sample from F_0 (Quintana and Newton 2000). Subjects sharing a common parameter value in w^* have a common distribution and thus are in the same group. The conditional posterior density of γ is

$$\pi(\gamma|\beta, w^*, N(t), Y(t)) \propto \gamma^{\kappa_1-1} e^{-\kappa_2\gamma} \frac{\gamma^{J'\gamma}}{(\Gamma(\gamma))^{J'}} \prod_{j=1}^{J'} (w_j^*)^{\gamma-1} e^{-\gamma \sum_{j=1}^{J'} w_j^*} \quad (9)$$

$$\propto \frac{\gamma^{J'\gamma+\kappa_1-1}}{(\Gamma(\gamma))^{J'}} e^{-\gamma(sum_j - \log prod_j + \kappa_2)}$$

where $sum_j = \sum_{j=1}^{J'} w_j^*$ is the sum of the distinct frailty values, and $prod_j = \prod_{j=1}^{J'} w_j^*$ is their product. This conditional does not simplify to any standard distribution, thus requiring a non-standard method of sampling from it.

4 Application of the Model

4.1 Data

The example data concerns child survival collected in the 2000 Malawi Demographic and Health Survey. The data are hierarchically clustered in 559 enumeration areas (EAs); these form our community clustering units. We concentrate on all of the 11,926 births in the 5 years preceding the survey. The distribution of the number of births per community had a mean of 21 and median of 20 with an interquartile range of 16–26. The minimum and maximum number of births per community was 5 and 68, respectively. Infant and under-five mortality rates per 1000 live births were estimated to be 104 and 189, respectively (NSO 2001). These rates are still very high when compared to those around the world (56 and 82, respectively) (Bolstad and Manda 2001). We considered some standard explanatory variables used in the analysis of child mortality in sub-Saharan countries (Manda 1999). The distributions of the predictor variables are presented in Table 1.

Table 1 Descriptive statistics of explanatory variables used in the analysis

Variable	Frequency	Percent
Gender of child		
Male	5951	49.9
Female	5975	51.1
Preceding birth interval		
First birth	2924	24.5
<24 months	1514	12.7
24–36 months	3569	29.9
>36 months	3919	32.9
Maternal education		
No education	3547	29.7
Primary	7213	63.0
Secondary and higher	866	7.3
Residence		
Urban	2084	17.5
Rural	9842	82.5
Region		
Northern	1936	16.2
Central	4394	36.8
Southern	5596	46.9
Total	11,926	100.0
	Mean	*Median*
Birth order	3.4	3.0
Maternal age	25.8	24.3

4.2   *Implementation and Results*

In a previous analysis of childhood mortality using the 1992 Malawi Demographic Health Survey data, Manda (1999) found that the rate of death for children under five years was 0.0055 deaths per birth per month. Thus, we set the mean cumulative increment in the baseline child death rate as $dH_0(t) = 0.0055\,dt$, where dt denoted a one-month interval. Our prior confidence in the mean hazard function is reflected by assigning $c = 0.1$, which is weakly informative. We adopt a vague proper gamma $(1, 0.1)$ prior for the precision of the community frailty effect γ. This implies that *apriori* the community-specific frailty w_j has variance $1/10 = 0.1$, which we thought to be reasonable as the differences in risk are likely to be minimal because we have adjusted for some important community-level factors: region and type of residence. We do not have sufficient prior information on the prior precision and concentration parameter M_0, so we assigned it a gamma $(0.001, 0.001)$ prior implying a mean of 1, but with reasonable coverage across its space.

The computation of the parameter estimates was run in *WinBUGS* software (Spiegelhalter et al. 2004). For each model considered, three parallel Gibbs sampler chains from independent starting positions were ran for 50,000 iterations. All fixed

effects and covariance parameters were monitored for convergence. Trace plots of sample values of each of these parameters showed that they were converging to the same distribution. We formally assessed convergence of the three chains using the German-Rubin reduction factor, and it stabilised to 1.0 by 2000 iterations. However, for posterior inference, we used a combined sample of the last 30,000 iterations. Using the posterior samples of the parameters, we calculated 50% (median) and (2.5% 97.5%) (95% CI) percentiles for posterior summaries. We also performed limited model selection using the Bayesian Information Criterion (BIC). BIC is defined as BIC $= -2 * \log$ likelihood $+ P \log g$, where P is the number of unknown parameters in the model and g is the sample size. A better fitting model has a smaller BIC value. The parametric gamma and the nonparametric model had BIC values of 13,657.68 and 11,797.96, respectively, showing that the nonparametric approach to frailty was a better fitting model for the data.

Posterior summaries are presented in Table 2 for both the gamma frailty model and the Dirichlet process frailty model. For the fixed effects, the results are presented on the logarithm scale where no risk is represented by 0. The estimates of the fixed

Table 2 Posterior median and 95% credibility interval (CI) of the parameters for the Malawi child survival data

Parameter	The parametric frailty		The nonparametric frailty	
	Median	95% CI	Median	95% CI
Male child (0: No, 1: Yes)	0.095	−0.008, 0.200	0.091	−0.008, 0.186
Birth order	−0.025	−0.112, 0.066	−0.027	−0.133, 0.073
Birth order^2	0.003	−0.006, 0.010	0.002	−0.006, 0.010
Preceding birth interval				
<2 years	0	−, −		−, −
2–3 years	−0.383	−0.528, −0.243	−0.381	−0.522, −0.252
>2 years	−0.404	−0.555, −0.254	−0.406	−0.570, −0.254
Maternal age				
(Age-26)/10	0.009	−0.154, 0.174	0.019	−0.154, 0.189
((Age-26)/10)^2	0.123	0.018, 0.230	0.117	0.006, 0.227
Maternal education				
Secondary+ (0: No, 1: Yes)	−0.197	−0.472, 0.060	−0.206	−0.465, 0.052
Region of residence				
Northern region	0	−, −		
Central region	−0.101	−0.275, 0.077	−0.106	−0.278, 0.079
Southern region	0.014	−0.155, 0.184	0.013	−0.152, 0.190
Rural residence (0: No, 1: Yes)	0.305	0.142, 0.479	0.295	0.127, 0.457
Frailty effects				
No of risk classes			21	3, 79
Variance of sub-risk frailties			0.073	0.026, 0.379
DP precision (M_0)			4.172	0.454, 33.83
Variance of frailty effects	0.056	0.027, 0.101		

effects are largely the same between a fully parametric and a nonparametric model. Based on the 95% CI, not all of the fixed effects are significant; however, the median estimated effects of the modelled covariates support the findings in previous studies on child mortality in sub-Saharan Africa and other less developed countries (Sastry 1997; Manda 1999). A male child has slightly higher risk of death. Birth order has a decreasing risk effect. A long preceding birth interval greatly reduces the risk of death for the index child. The coefficient of the quadratic part of the age of the mother indicates a child born to a younger or older mother has higher risk. Maternal education is clearly an important factor: higher educational level is a surrogate for many positive economic and social factors. Living in rural areas has an increased risk. Region of residence is a risk factor, though not greatly pronounced.

The nonparametric Dirichlet mixing process has identified 21 classes into which the 559 communities can be separated according to their childhood mortality risk. The distinct latent risk values have a posterior median variance of 0.073, slightly larger than the variance of the frailty effects under the full parametric gamma frailty. The precision of the DP process has a median of 4, well below the total number of communities, indicating that distribution of the community effects is more likely multimodal and nonparametric.

5 Time Dependent Frailty

We have seen that in Sect. 3, stage 2 could be modelled conveniently, using a gamma distribution with mean one and unknown variance. However, apart from being restrictive, the gamma distribution has some undesirable properties in that it is not symmetric or scale-invariant; a property which ensures that the inference does not depend on the measuring units (Vaida and Xu 2000). We have shown how this frailty effect could be modeled nonparametrically. In here, we still use stage 2 modeling as in Sect. 3, where some flexibility is imposed on the random frailty effects by assuming it is time-dependent. Thus, the notation for subject frailty effect is slightly changed to depend on t. Thus, in the following, $w_{i(t)}$ is assumed to follow a first-order autocorrelated AR(1) process prior:

$$w_{i(t)}|w_{i(t-1)} = \phi w_{i(t-1)} + e_{i(t)}; \quad i = 1, \ldots, I, \ t \geq 0 \qquad (10)$$

where $w_{i(0)} \sim \text{Normal}(0, \sigma_w^2)$ and $e_{i(t)}$ are *i.i.d* random variables having a Normal$(0, \sigma_w^2)$ distribution. The parameter ϕ is constrained to lie between -1 and 1 and it measures the degree of serial correlation in the subject-specific frailty. The prior density of the frailty vector $w = (w_{1(t)}, \ldots, w_{I(t)})$ is given by

$$p(w|\phi, \sigma_w^2) \propto \prod_{i=1}^{I} \prod_{t \geq 0}^{c} (\sigma_W^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_w^2}(w_{i(t)} - \phi w_{i(t-1)})^2\right)$$

5.1 Example Data

The example data is taken from Fleming and Harrington (1991) and it describes a double-blinded placebo controlled randomised trial of gamma interferon (γ-IFN) in chronic granulomatous disease (CGD). The disease is a group of rare inherited disorders of the immune function, characterized by recurrent pyogenic infections. These infections result from the failure to generate microbicidal oxygen metabolites, within the phagocytes, to kill ingested micro-organisms. The disease usually presents early in life and may lead to death in childhood. There is evidence that gamma interferon is an important macrophage activating factor which could restore superoxide anion production and bacterial killing by phagocytes in CGD patients. A total of 128 patients were followed for about a year and 203 infections/censorings were observed, with the number of infections per patient ranging from 1 to 8. Of the 65 patients on placebo, 30 had at least one infection, but only 14 of the 63 patients on gamma interferon treatment had more than one infection. This resulted in 56 and 20 infections amongst the placebo and gamma interferon groups respectively.

The original data set is shown by treatment group in Table 3. The use of corticosteroid on entry is not sufficiently varied in the data (a very small proportion of the subjects were using it), hence it is not used in the present analysis. These data have previously been analysed using gap-times between infections and total times using the number of preceding events and a shared frailty model. In our analysis, we also use the logarithm of the number of previous infections plus 1 (labelled as *PEvents (t)*) as the deterministic time-dependent component of frailty. Its parameter ω measures the effect of past infections on the risk of the current infection. We used 5-day intervals for each patient, resulting in 88 such intervals.

Table 3 Baseline characteristics according to treatment group

Baseline characteristic	Placebo	Gamma interferon
No. of patients	65	63
Pattern of inheritance		
(0: autosomal recessive, 1: X-linked)	41 (63.1%)	45 (71.4%)
Age (in years)[a]	14.98 (9.64)	14.29 (10.12)
Height (in cm)[a]	140.55 (34.10)	139.57 (27.24)
Weight (in kg)[a]	42.30 (24.32)	38.76 (19.92)
Corticosteroid use (0: No, 1: Yes)	2 (3.1%)	1 (1.6%)
Prophylactic antibiotic use (0: No, 1: Yes)	55 (84.6%)	56 (88.9%)
Gender (0: male, 1: female)	12 (18.5%)	12 (19.1%)
Hospital region (0: USA, 1: Europe)	22 (33.8%)	17 (27.0%)

[a]Mean with standard deviation (SD) in parenthesis

5.2 Prior Specification and Model Comparison

In running the Gibbs sampling algorithm, the prior specifications were as follows: For both the baseline and fixed effect parameters, σ_0^2 was set to 1000, resulting in a normal distribution which is very uninformative. In specifying noninformativeness for the frailty variance σ_W^2, the hyperparameters κ_0 and ω_0 of the inverse-gamma prior are usually set to very low values, typically 0.001 is used in the WinBUGS software. However, κ_0 and ω_0 values near 0 can result in an improper posterior density. Moreover, when the true frailty variance is near 0, inferences become sensitive to choices of κ_0 and ω_0 (Gelman 2004). For numerical stability, we chose $\sigma_W^2 \sim$ inverse-gamma$(1, 1)$. In specifying a Beta(ξ_0, φ_0) prior for ξ, the values of ξ_0 and φ_0 were set at 3 and 2. This centres ξ at 0.6 (and ϕ at 0.2), but is flat far away from this value.

We considered a number of competing models for these data and compared them using the Deviance Information Criterion (DIC) developed by Spiegelhalter et al. (2002). The DIC is defined as DIC $= \bar{D} + p_D$ where $\bar{D}$ is the posterior mean of the deviance (measures model *fit*) and p_D is the effective number of parameters in the model (measures model *complexity*). The parameter p_D is calculated using $p_D = \bar{D} - D(\bar{\psi})$, where $D(\bar{\psi})$ is the deviance evaluated at the posterior mean of the unknown parameters. The DIC is particularly useful in situations involving complex hierarchical models in which the number of parameters used is not known. It is a generalisation of the Akaike Information Criterion (AIC) and works on similar principles. The models we compared were:

- Model 1: $\eta_{i(t)} = \beta^T X_i(t) + \omega P Events(t)$. This model allows for differences in the infection times that only depend on the measured risk variables including a deterministic time-dependent component of frailty; it does not allow for the random frailty component.
- Model 2: $\eta_{i(t)} = \beta^T X_i(t) + \omega P Events(t) + W_i$. This model allows for differences depending on the covariates as well as the random frailty component for each patient. However, it assumes that the patient random frailty is constant over time. Thus, W_i can be considered as the frailty on entry for the ith patient and it does not change with time.
- Model 3: $\eta_{i(t)} = \beta^T X_i(t) + \omega P Events(t) + W_{i(t)}$. Rather than assuming a constant model for the random subject frailty effect, Model 3 uses a time-dependent random frailty. This became our main model, which we suggest is better at explaining all the variation in the data.

For each model, the Gibbs sampler was run for 100,000 iterations and using trace plots of sample values, we found very rapid convergence for the baseline and fixed effect parameters, but a longer *burn-in* was required for the convergence of variance and correlation parameters. The first 50,000 iterations were discarded and the remaining 50,000 samples were used for posterior inference. The estimates of DIC and p_D for the three competing models are presented in Table 4. It should be noted that the use of DIC is not meant to show a true model, but rather to compare

Table 4 DIC parameters for model selection

Model I	$\bar{D}$	$D(\bar{\psi})$	p_D	DIC
Model I	812.00	803.04	8.96	821.33
Model II	783.24	750.14	33.10	816.74
Model III	724.74	651.11	73.62	798.36

different sets of models. Model choice is based on a combination of model fit, complexity and the substantive importance of the model. Starting with the fixed effect Model 1, the effective number of parameters p_D is 8.96, nearly equal to the correct number of parameters which is 10 (8 fixed effects, a longitudinal parameter and a constant baseline hazard). This model does not offer any meaningful insights into the variation in the data above that given by the observed covariates. Thus, we are motivated to consider random effects models, which are Models 2 and 3. For Models 2 and 3, the estimates of p_D are roughly 33 and 74 respectively. The number of unknown is 139 and 11,266 for Models 2 and 3, respectively. Thus, even though the number of unknown parameters in Model 3 is about 80 times larger than that in Model 2, its effective number of parameters is just a little over twice as large. This suggest that Model 3 is still quite sparse. Furthermore, it has the lowest goodness-of-fit as measured by $\bar{D}$. The effectiveness of Model 3 is further supported by its lowest DIC value, indicating that this model is best in terms of overall model choice criterion that combines goodness-of-fit with a penalty for complexity. In these circumstances, we prefer Model 3, as it is more effective and explains the variation in the data better.

5.3   Results

The posterior estimates from Model 1 are given in columns two and three in Table 5. The posterior distribution of the treatment effect has mean -0.875 and standard deviation (SD 0.276), showing that the (γ-IFN) treatment has the effect to reduce substantially the rate of infection in CGD patients. Other covariates that have an effect on the rate of infection include pattern of inheritance, age, the use of prophylactic at study entry, gender and hospital region. Height and weight were not significantly related to rate of infection. The results also show that the risk of recurrent infection significantly increases as the number of previous infections increases (mean 0.712 and SD 0.225). For comparison, the treatment and the longitudinal effects in Lindsey (1995), without controlling for the other covariates, were estimated at -0.830 (SD 0.270) and 1.109 (SD 0.227), respectively, using maximum likelihood estimation. When we left out the number of previous infections, the effect of treatment increases to -1.088; the same was also observed in Lindsey (1995) where it changed to -1.052. Thus, the effect of treatment is partially diminished by the inclusion of the number of previous infections.

Table 5 Posterior estimates of parameters using the Gibbs sampling

Parameter	Model 1		Model 2		Model 3	
	Mean	SD	Mean	SD	Mean	SD
γ-IFN	−0.875	0.276	−1.065	0.353	−0.889	0.283
Inheritance	−0.458	0.297	−0.542	0.393	−0.466	0.308
Age	−0.063	0.034	−0.073	0.045	−0.063	0.035
Height	0.004	0.010	0.008	0.015	0.005	0.011
Weight	0.001	0.016	0.005	0.022	0.009	0.017
Prophylactic	−0.541	0.323	−0.726	0.466	−0.561	0.336
Sex	−0.507	0.400	−0.536	0.520	−0.507	0.412
Hospital region	−0.553	0.312	−0.658	0.400	−0.556	0.319
ω	0.712	0.225	0.146	0.330	0.699	0.247
σ_W^2	–	–	0.788	0.459	0.879	0.507
ϕ	–	–	–	–	0.313	0.278

Next, we consider Model 2 which compares to analyses in McGilchrist and Yau (1996), Yau and McGilchrist (1998) and Vaida and Xu (2000) using (restricted) maximum likelihood estimation on inter-event times and the EM algorithm on the Cox model, respectively. The results are presented in columns four and five of Table 5. The posterior estimates of the risk variables are fairly close to those obtained under Model 1, except that the effect of the longitudinal parameter ω is now much reduced and no longer significant. Similar results were found in Vaida and Xu (2000), where the effect of the number of previous infections was reduced to a non-significant result when the random frailty term was included in the analysis. The variance of patient frailty effect is estimated with a posterior mean 0.788 and SD 0.459 and is significant. In comparison, the estimates of the frailty variance were 0.237 (SD 0.211) and 0.593 (SD 0.316) under ML and REML methods, respectively, in McGilchrist and Yau (1996) when the number of the previous infections was adjusted for, and 0.743 (SD 0.340) in Yau and McGilchrist (1998) without controlling for the number of previous infections. It appears that the number of previous infections explains most of the variation in the patient's frailty effects. However, we argue for the inclusion of both deterministic time-dependent and random components of frailty as they control for order and common dependence in recurrent-event times. Moreover, using a random frailty effect will also account for the effects of the missing covariates.

Finally, we examine the results from Model 3 and these are given in the last two columns of Table 5. Posterior means and SDs of the risk variables are in general agreement with those obtained in Models 1 and 2. The frailty correlation parameter ϕ is estimated with a posterior mean 0.313 and SD 0.278, indicating the presence of a positive serial correlation between the recurrent event times, but not sufficiently significant. The variance of frailty has posterior mean 0.879 and SD 0.507, showing that, apart from the serial dependence, the recurrent event times within a subject share a common frailty effect that partially summarizes the dependence within the subject.

The general criticism of using a normal model for the subject random effects is its susceptibility to outliers. Thus, we considered three other comparable models for the subject random effects: double exponential, logistic and t_d distributions for sensitivity analyses on the parameter estimates. Using a small d, say 4, the t_4 distribution can also be used as a robust alternative model for the subject random effects. We fitted each alternative model as in Model 3 with the same prior specification for β and the variance parameters. The fixed effects were essentially the same as those in Table 5. For instance the effect of treatment was estimated with posterior mean -0.909, -0.894 and -0.883 with SD 0.296, 0.284 and 0.281, respectively, for the double exponential, logistic and t_4 models. There was a slight sensitivity to the frailty variance estimate with mean and SD values 0.255 (0.093), 0.418 (0.112) and 0.349 (0.122) respectively. However, the overall conclusions that the frailty variance is significant is the same for the models and in general this application is insensitive and robust.

Our analyses yield posterior means and SDs of fixed effect parameters that are lower than those obtained by Yau and McGilchrist (1998) using either ML or REML estimation. However, we obtain an estimate of the frailty variance, which is larger than the ML estimate (0.200, SD 0.480), but similar to the REML estimate (0.735, SD 0.526). The Bayesian estimate of the correlation parameter is also lower than both the ML (0.729, SD 0.671) and REML (0.605, SD 0.325) estimates. It should be noted that they used total time since the first failure to define a deterministic time-dependent frailty, in addition to using inter-event times in their models, so these differences might be due to the modelling strategy of recurrent event times and the deterministic time-varying component of frailty.

6 Discussion

This chapter has introduced two flexible approaches to dealing with associations between survival times in a conditional model for the analysis of multivariate survival data. The two approaches were demonstrated using real data sets. Firstly, we assumed that frailty effect could be modelled nonparametrically using a Dirichlet process prior, which specifies prior uncertainty for random frailty effects at the level of the distribution function; this offers infinite alternatives. This could have practical benefits in many applications, where, often, the concern is grouping units into strata of various degrees of risk. In our analysis of the child survival data, the nonparametric approach allowed us to categorize the sample of the communities into 21 classes of risk of childhood mortality. Understanding that communities can be classified according to their risk of childhood mortality provides useful guidance on the effective use of resources for childhood survival and preventive interventions. The identified 21 sub-classes of risk could be administratively convenient and manageable for child health intervention programs.

We note that the model could be extended to have the covariate link function unspecified and modelled nonparametrically, for instance, by a mixture of beta or

triangular distributions (Perron and Mengersen 2001). Such an extension would make the proportional hazards model fully nonparametric in all parts of the model. A similar model involving minimal distributional assumptions on fixed and frailty effects was presented in Zhang and Steele (2004). However, implementation of such models would be computationally more complex and intensive, especially when there are many short intervals for the counting process.

For the data example, the estimates of the fixed effects and 95% confidence intervals were similar under both frailty model assumptions; thus substantive conclusions are not affected by whether or not we use a nonparametric form for the frailty effects. The fact that a seemingly more complex model produces similar results to those under a simpler model is irrelevant when we consider that there is no reason *aprior* to believe that a gamma shape is adequate for the distribution of the random frailties (Dubson 2009). A parametric gamma model would have been unable to model frailty adequately if the distribution of the frailties had arbitrary shapes and if there were interactions between the observed and the unobserved predictors. Thus, a nonparametric frailty model validates the parametric gamma model in this application.

In the second approach, we have shown that a model encompassing dependence through a time-varying longitudinal parameter with complex structure, that accounts for both intra-subject and order correlation, provides a better fit than other traditional models of dependence. It has also been shown for these data that only the deterministic time component of frailty is important. In the more complex models, using an AR(1) frailty model, the correlation and the random component show high positive dependence and variation, respectively, though with minimal significance since these estimates are marginally larger than their standard deviations. Despite this, we are satisfied that a model has been presented for consideration in the analysis of similar data structures, whether binary or count data.

References

Andersen, P. K., & Gill, R. D. (1982). Cox's regression models for counting processes. *The Annals of Statistics, 10*, 1100–1120.

Blackwell, D., & Macqueen, J. B. (1973). Ferguson distribution via Polya Urn schemes. *Annals of Statistics, 1*, 352–355.

Bolstad, W. M., & Manda, S. O. M. (2001). Investigating child mortality in Malawi using family and community random effects: a Bayesian analysis. *Journal of the American Statistical Association, 96*, 12–19.

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

Congdon, P. (1994). Analysing mortality in London: Life-Tables with frailty. *Statistician, 43*, 277–308.

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

Dubson, D. P. (2009). Nonparametric Bayesian application to biostatistics. In N. Njort, C. Holmes, P. Muller, & S. Walker (Eds.), *Bayesian nonparametrics in practice* Cambridge: Cambridge University Press.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association, 89*, 268–277.

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

Ferreira, A., & Garcia, N. L. (2001). Simulation study for mis-specifications on frailty model. *Brazilian Journal of Probability and Statistics, 15*, 121–134.

Fleming, T. R, Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Gelman, A. (2004). Parameterization and Bayesian Modeling. *Journal of the American Statistical Association, 99*(466), 537–545. https://doi.org/10.1198/016214504000000458.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Green, P. J., & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandanavian Journal of Statistics, 28*, 355–375.

Guo, G., & Rodriguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with application to child survival in Guatemala. *Journal of the American Statistical Association, 87*, 969–979.

Ha, I. D., Lee, Y., & Song, J.-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika, 88*, 233–243.

Heckman, J., & Singer, B. (1984). A Method of minimising the impact of distributional assumptions in econometric models for duration data. *Econometrical, 52*, 231–241.

Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association, 96*, 161–173.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B, 40*, 214–221.

Kleinman, K. P., & Ibrahim, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics, 54*, 921–938.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association, 73*, 805–811.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine, 13*, 2233–2247.

Lindsey, J. K. 1995. Fitting parametric counting processes by using log-linear models. *Applied Statistics, 44*, 201–212.

Maceachern, S. N. (1994). Estimating normal means by a conjugate style Dirichlet process prior. *Communications in Statistics, 23*, 727–741.

Manda, S. O. M. (1999). Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. *Social Science and Medicine, 48*, 301–312.

Manda, S. O. M. (2011). A nonparametric frailty model for clustered survival data. *Communications in Statistics: Methods and Theory, 40*(5), 863–875.

Manda, S. O. M., & Meyer, R. (2005). Bayesian inference for recurrent events data using time-dependent frailty. *Statistics in Medicine, 24*, 1263–1274.

McGilchrist, C. A, & Yau, K. K. W. (1996). Survival analysis with time dependent frailty using a longitudinal model. *Australian Journal of Statistics, 38*, 53–60.

Muller, P., & Quintana,F. (2009). *More nonparametric Bayesian for biostatistics*. Available at www.mat.puc.cl/~quintana/mnbmfb.pdf. Accessed 21 May 2009.

Naskar, N. (2008). Semiparametric analysis of clustered survival data under nonparametric frailty. *Statistica Neerlandica, 62*, 155–172.

Naskar, N., Das, K., & Ibrahim, I. G. (2005). A semiparametric mixture model for analysing clustered competing risk data. *Biometrics, 61*, 729–737.

National Statistical Office (NSO). (2001). *Malawi demographic and health survey 2000*. Calverton, MD: National Statistical Office and Macro International Inc.

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

Perron, F., & Mengersen, K. (2001). Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics, 57*, 518–528.

Pickles, A., & Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine, 14*, 1447–1461.

Quintana, F. A., & Newton, M. A. (2000). Computational aspects of nonparametric Bayesian analysis with applications to the modelling of multiple binary sequences. *Journal of Computational and Graphical Statistics, 9*, 711–737.

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

Sastry, N. (1997). A nested frailty model for survival data, with application to study of child survival in Northeast Brazil. *Journal of American Statistical Association, 92*, 426–435.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica, 4*, 639–650.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, 64*, 583–639.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2004). *BUGS: Bayesian inference using Gibbs sampling, version 1.4.1*. Medical Research Council Biostatistics Unit, Cambridge University.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2005). *Hierarchical Dirichlet Processes*. Technical Report 653. Department of Statistics, University of California at Berkeley.

Tierney, L. (1994). Markov chain for exploring posterior distributions. *Annals of Statistics, 22*, 1701–1762.

Vaida, F., & Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine, 19*, 1309–3324.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*, 439–454.

Walker, S. G., & Mallick, B. K. (1997). Hierarchical generalised linear models and frailty models with Bayesian nonparametric mixing. *Journal of Royal Statistical Society B, 59*, 845–860.

Yau, K. K. W., & McGilchrist, C. A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine, 17*, 1201–1213.

Zhang, W., & Steele, F. (2004). A semiparametric multilevel survival model. *Journal of the Royal Statistical Society, Series C, 53*, 387–404.