Published in final edited form as:

Biometrics. 2015 September ; 71(3): 760–771. doi:10.1111/biom.12298.

Bayesian Path Specific Frailty Models for Multi-state Survival Data with Applications

Mário de Castro¹, Ming-Hui Chen², and Yuanye Zhang³

Mário de Castro: mcastro@icmc.usp.br; Ming-Hui Chen: ming-hui.chen@uconn.edu; Yuanye Zhang: yuanyevickiezhang@gmail.com

¹Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos, SP, Brazil

²Department of Statistics, University of Connecticut, Storrs, CT, USA

³Novartis Institutes for BioMedical Research, Inc., Cambridge, MA, USA

Summary

Multi-state models can be viewed as generalizations of both the standard and competing risks models for survival data. Models for multi-state data have been the theme of many recent published works. Motivated by bone marrow transplant data, we propose a Bayesian model using the gap times between two successive events in a path of events experienced by a subject. Path specific frailties are introduced to capture the dependence structure of the gap times in the paths with two or more states. Under improper prior distributions for the parameters, we establish propriety of the posterior distribution. An efficient Gibbs sampling algorithm is developed for drawing samples from the posterior distribution. An extensive simulation study is carried out to examine the empirical performance of the proposed approach. A bone marrow transplant data set is analyzed in detail to further demonstrate the proposed methodology.

Keywords

Gamma frailty; Gap time; Gibbs sampler; Piecewise exponential model; Survival analysis

1. Introduction

Markov models and Markov extension models are widely used to model multi-state data, in which the transition probabilities and transition intensities are the major study focus. There are two major time scales used for studying transition intensities: the study time since the study origin and the duration time in the current state. Based on the time scale to use, Markov model and Markov extension models are classified into several categories. The non-

9. Supplementary Materials

The detailed computation of the survival probabilities for terminating events in Section 4.3 and the required steps for implementing the Gibbs sampling algorithm for posterior inference in Section 5.2 are presented in Web Appendix A and Web Appendix C, respectively. The proofs of propositions and theorems are given in Web Appendix B. Online supplementary materials include also figures for the three and four states models in Sections 4.1 and 6. Figures S3–S7 display plots of the posterior distribution for some elements. Tables S2 and S3 pertain to Sections 2 and 6, respectively, whereas Table S4 is discussed in Section 7. Web appendices, tables, and figures referenced in the paper are available with this paper at the Biometrics website on Wiley Online Library

homogeneous Markov model assumes that the upcoming transition intensity depends on the history only via the current state and the study time. The homogeneous Markov model (classical Markov model) further assumes that all transition hazards are constant. The homogeneous semi-Markov model assumes that the transition intensity, given the current state and the duration time in the current state, is independent of the history. The non-homogeneous semi-Markov model allows the transition intensities to depend on both of the study and duration time scales. In our paper, we are going to deal with Markov and semi-Markov models.

Models for multi-state data have been the subject of many published works, most of which are based on frequentist methods. Comprehensive reviews about the development and applications are given by, for example, Commenges (1999), Hougaard (1999), Hougaard (2000), Andersen and Keiding (2002), Andersen and Perme (2008), Meira-Machado et al. (2009), Zhao (2009), Andersen and Perme (2013), and references therein. In the paper by Fiocco et al. (2008), a Markovian model is adopted, semi-parametric Cox models with transition-specific covariates are fitted, and the transition probability matrix is estimated using the Aalen-Johansen estimator (Aalen and Johansen, 1978). Fiocco et al. (2008) also consider a model in which some of the baseline hazards are assumed to be proportional (see also Keiding et al., 2001). Titman and Sharples (2010a) propose several tools for model diagnostics. Titman and Sharples (2010b) formulate a semi-Markov model for discrete time survival data with an extension to situations in which observed states are subject to classification error. Xu et al. (2010) present an interesting connection between semi-competing risks models and illness-death processes (see Section 4.1). Ferguson et al. (2012) develop an R package for summary calculations (e.g., state occupation probabilities) in a general, possibly non-Markov, multi-state model without covariates. The semi-competing risks models proposed by Zeng et al. (2012) can also be interpreted as multi-state models.

Under a Bayesian viewpoint, Kneib and Hennerfeind (2008) devise a model with transition intensities specified in a multiplicative way, enabling the inclusion of flexible nonparametric and time-varying effects. Subject specific variation not accounted for by covariates is captured by transition specific frailties. The baseline transition intensities are represented by penalized splines. Armero et al. (2012) present a Weibull semi-Markov model for the three states disability model (see Section 4.1). Kim et al. (2012) develop a semi-parametric regression model based on a Markov process and a beta-Dirichlet process for the cumulative intensity functions. Conlon et al. (2014) present a multi-state Markov model with a cured fraction that jointly models recurrence and death.

As an anonymous referee pointed out, the flowgraph models discussed in Huzurbazar (2005) and Warr and Huzurbazar (2010) represent an alternative for modeling multi-state data. These models are based on the specification of the moment generating function or the complex Laplace transform for the waiting times between two states. Up to the best of our knowledge, flowgraph models including frailties have not yet been investigated.

Several important issues remain unaddressed in the existing literature. First, the dependence structure of the gap times between two states is not always explicitly accounted for. In this paper, we take a different approach to model multi-state data by developing Bayesian gap

time models with path specific frailties. An anonymous referee pointed out that the frailty represents the vulnerability of a subject. From a clinical point of view, patients going to different paths shall have different vulnerabilities. As discussed in Zhang et al. (2014), the marginal distributions of the gap times after integrating out the frailty belong to the class of generalized odds-rate hazards (GORH) models (Banerjee et al., 2007). As the GORH model is a non-proportional hazards model, the proposed path specific frailty model is more robust to the proportional hazards assumption than the classical Markov and semi-Markov models in which the proportional hazards models are assumed for the gap times. Second, within the Bayesian framework, we characterize the conditions for posterior propriety under improper noninformative priors specified for regression coefficients as well as other model parameters. Third, the path probabilities are difficult to compute under most of the existing models in the literature. However, under the proposed path specific frailty model, the path probabilities are easily estimated as a by-product of the Gibbs sampler. Fourth, as stressed by Zhang et al. (2014), the computation is difficult to carry out when fitting a frailty model with unknown parameters in the frailty distribution since the estimates of these unknown parameters are generally unstable. The computation becomes even more challenging for fitting the complex multi-state survival data. In this regard, we develop an efficient Gibbs sampling algorithm via the introduction of latent variables and reparameterizations. In addition, in the proposed Bayesian approach, the derivation of the closed form expression of the likelihood after integrating out latent frailties facilitates the development of model comparison criteria such as the deviance information criterion (DIC) and the logarithm of the pseudo marginal likelihood (LPML). Our simulation study in Section 6 shows that these versions of DIC and LPML yield a good statistical power to select the true model.

The rest of the paper is organized as follows. Section 2 presents a detailed description of the bone marrow transplant (BMT) data. The building blocks of the model are described in Section 3 and some model properties are examined in Section 4. The prior distribution is specified and the propriety of the posterior distribution is established in Section 5. Moreover, model comparison criteria are also presented. The performance of the proposed methodology is assessed through a simulation study conducted in Section 6. In Section 7, the methodology is applied to the analysis of the bone marrow transplant data set presented in Section 2. Some general concluding remarks are given in Section 8.

2. Bone Marrow Transplant Data

According to Fiocco et al. (2008), bone marrow transplantation is an effective and standard treatment for acute leukemia, but the procedure is associated with considerable morbidity and mortality. The BMT data used in this paper are available in the *mstate* package in R (de Wreede et al., 2011). Six states are considered. As shown in Figure 1, after transplant, the patients may experience platelet recovery (PR), acute graft-versus-host disease (AGvHD) or both PR and AGvHD, all of which are considered as nonterminating events. The patients may also experience either relapse or death in remission, both of which are considered as terminating events. Graft-versus-host disease (GvHD), either acute or chronic, is the most common non-relapse complication. Patients who develop acute GvHD are more likely to develop chronic GvHD than others. Overall, there are 12 transitions and the number of patients in these transitions are shown in Figure 1. These data motivate our proposed

methodology. The data are comprised of 2279 patients treated between 1985 and 1998. Table S2 (in the online supplementary materials) presents the baseline prognostic factors.

We can also view Figure 1 as a path diagram, with boxes representing states and arrows representing transitions from a parent state to a child state. In Figure 1, there are $K = 10$ possible complete paths and their corresponding states and transitions are given as follows: P1: $1 \rightarrow 5$, P2: $1 \rightarrow 2 \rightarrow 5$, P3: $1 \rightarrow 2 \rightarrow 6$, P4: $1 \rightarrow 2 \rightarrow 4 \rightarrow 5$, P5: $1 \rightarrow 2 \rightarrow 4 \rightarrow 6$, P6: $1 \rightarrow 3 \rightarrow 4 \rightarrow 5$, P7: $1 \rightarrow 3 \rightarrow 4 \rightarrow 6$, P8: $1 \rightarrow 3 \rightarrow 5$, P9: $1 \rightarrow 3 \rightarrow 6$, and P10: $1 \rightarrow 6$. There are five possible incomplete paths as well, including IP1: 1, IP2: $1 \rightarrow 2$, IP3: $1 \rightarrow 3$, IP4: $1 \rightarrow 2 \rightarrow 4$, and IP5: $1 \rightarrow 3 \rightarrow 4$.

Among the 2279 patients who underwent transplant, the numbers of patients belonging to these paths are 95 patients for P1, 112 patients for P2, 39 patients for P3, 33 patients for P4, 60 patients for P5, 74 patients for P6, 77 patients for P7, 56 patients for P8, 197 patients for P9, 160 patients for P10, 332 patients for IP1, 407 patients for IP2, 221 patients for IP3, 134 patients for IP4, and 282 patients for IP5. The median gap times in years of the transitions in Figure 1 are 0.066 for $1 \rightarrow 2$, 0.045 for $1 \rightarrow 3$, 0.627 for $1 \rightarrow 5$, 0.144 for $1 \rightarrow 6$, 0.027 for $2 \rightarrow 4$, 0.586 for $2 \rightarrow 5$, 0.479 for $2 \rightarrow 6$, 0.041 for $3 \rightarrow 4$, 0.675 for $3 \rightarrow 5$, 0.205 for $3 \rightarrow 6$, 0.578 for $4 \rightarrow 5$, and 0.444 for $4 \rightarrow 6$.

3. Proposed Models

In this section, we introduce the two main components of our models. We begin with some definitions. If a state does not have any child states, then it is an absorbing state, whereas if a state does not have a parent state, then it is a starting state. In the studies that we consider, there is one starting state and one or more absorbing states. We only consider the situation in which all the transitions are in one direction from an early state to a late state but not vice versa. If a state is neither a starting state nor an absorbing state, then it is a transient state. A transient state must have both parent and child states. As in Section 2, a sequence of connected states from the starting state to an absorbing state is called a path. There are $J - 2$ states and $K - 1$ paths. We assume that each subject goes through just one path.

3.1 Model for Immediate Child States

For each parent state j , let $\mathcal{P}_j = \{l : \text{state } l \text{ is an immediate child state of state } j\}$ denote the collection of all possible immediate child states of state j , $j = 1, \dots, J$. Under our setting, the states are numbered so that we always have $l > j$ when $l \in \mathcal{P}_j$. Let δ_j denote a possible value of a child state for the parent state j , independent of the gap times. We assume a multinomial logistic regression model for δ_j as follows:

$$P(\delta_j = l | \mathbf{z}_1, \boldsymbol{\alpha}^{(j)}) = \exp(\mathbf{z}_1' \boldsymbol{\alpha}_l^{(j)}) / \sum_{l^* \in \mathcal{P}_j} \exp(\mathbf{z}_1' \boldsymbol{\alpha}_{l^*}^{(j)}), \quad (1)$$

where $\mathbf{z}_1$ is a $p \times 1$ vector of covariates, $\boldsymbol{\alpha}_{l^*}^{(j)}$ is a $p \times 1$ vector of parent and child specific regression coefficients for $l^* \in \mathcal{P}_j$, and $\boldsymbol{\alpha}^{(j)} = ((\boldsymbol{\alpha}_l^{(j)})' : l \in \mathcal{P}_j)'$. To ensure identifiability, we assume $\boldsymbol{\alpha}_{l_j}^{(j)} = \mathbf{0}$, where $l_j = \max\{l : l \in \mathcal{P}_j\}$.

3.2 Models for the Gap Times with Path Specific Frailty

Let T_{jl} denote the gap time between two connected states j and l . In Figure 1, we have a total of 12 gap times. Let $S_k = \{l : \text{state } l \text{ is in path } k\}$ denote the collection of states along path P_k for $k = 1, \dots, K$, and $|S_k|$ denote the cardinality of the set S_k . For each path P_k , in order to model variability in the hazard not accounted for observable covariates, we introduce a frailty term w , which has a path specific distribution with parameter τ_k . Each subject eventually ends up with a certain path. Each path is composed by a number of states representing intermediate points before reaching a terminating state. The vulnerability, represented by the frailty, depends on the sequence of events experienced by the subject.

For complete paths, we assume that

$$w|\tau_k \sim \text{gamma}(1/\tau_k, 1/\tau_k) \quad (2)$$

independently, with density function $f(w|\tau_k) = (1/\tau_k)^{1/\tau_k} w^{1/\tau_k - 1} \exp(-w/\tau_k) / \Gamma(1/\tau_k)$ for $w > 0$. The distribution of the frailty can be determined only if we know which path the transition $j \rightarrow l$ belongs to. We emphasize that the frailty is subject specific and the subjects undergoing a given path have frailties that follow the same distribution.

For complete paths, it is observed that if $|S_k| = 2$ for the path P_k , then the transition in S_k between two connected states j and l cannot be shared by other paths. If $|S_k| > 2$, then the transition $j \rightarrow l$ can be shared by other paths. In this case, there will be more than one path specific conditional distributions for the corresponding gap time T_{jl} according to (2).

Next, given P_k , for $j, l \in S_k$, the path specific hazard function for T_{jl} when states j and l are connected is assumed as

$$h_{jl}(t|z_2, \beta_{jl}, P_k) = wh_{jl0}(t) \exp(z_2' \beta_{jl}), \quad (3)$$

where the distribution of $w|\tau_k$ is given in (2), $h_{jl0}(\cdot)$ is the transition specific baseline hazard function, z_2 is a $q \times 1$ vector of covariates, and β_{jl} is a vector of transition specific regression coefficients. In the case that $\delta_j = l$, we assume that $T_{jl} = \infty$. Notice that in the marginal model, the proportionality of the hazards is relaxed, even when $|S_k| = 2$. For notational simplicity, we assume that the covariates z_1 and z_2 in (1) and (3) are the same for all transitions. For example, for path $P_4: 1 \rightarrow 2 \rightarrow 4 \rightarrow 5$, (3) implies that

$$h_{12}(t|z_2, \beta_{12}, P_4) = wh_{120}(t) \exp(z_2' \beta_{12}), \quad h_{24}(t|z_2, \beta_{24}, P_4) = wh_{240}(t) \exp(z_2' \beta_{24}), \quad \text{and} \\ h_{45}(t|z_2, \beta_{45}, P_4) = wh_{450}(t) \exp(z_2' \beta_{45}), \quad \text{where } w|\tau_4 \sim \text{gamma}(1/\tau_4, 1/\tau_4).$$

The baseline hazard function $h_{jl0}(\cdot)$ is represented by a piecewise constant function. First we create a partition of the gap time axis with M_{jl} intervals and cutpoints $0 = c_{jl0} < c_{jl1} < \dots < c_{jlM_{jl}}$, where $c_{jlM_{jl}} > t_{ijl}$ for all subjects i sharing the transition $j \rightarrow l$. In this way, the intervals are $(0, c_{jl1}]$, $(c_{jl1}, c_{jl2}]$, $\dots$, $(c_{jlM_{jl}-1}, c_{jlM_{jl}}]$. We also define an interval indicator I_{ijlm} such that $I_{ijlm} = 1$ if a subject i sharing the transition $j \rightarrow l$ failed or was right-censored in the

m -th interval; $I_{ijlm} = 0$ otherwise. In the m -th interval, we assume a constant hazard λ_{jlm} , $m = 1, \dots, M_{jl}$, so that

$$h_{j|0}(t) = \lambda_{jlm} \quad \text{and} \quad H_{j|0}(t) = \lambda_{jlm}(t - c_{jl,m-1}) + \sum_{g=1}^{m-1} \lambda_{jlg}(c_{jlg} - c_{jl,g-1}), \quad (4)$$

when $c_{jl,m-1} < t < c_{jlm}$. In the results reported in Sections 6 and 7, for $m = 1, \dots, M_{jl}$, we chose intervals $(c_{jl,m-1}, c_{jlm}]$ based on the percentiles of the gap times for subjects with a complete path.

3.3 Likelihood Function

Let n denote the number of subjects. For the i -th subject, let y_{ij} denote the observed event time or right-censored time at state j . When $j = 1$, which is the starting state, let $y_{i1} = 0$. Let δ_{ij} denote a possible value of child states for the parent state j . If $\delta_{ij} = l$, then the gap time between the parent state j and its child state $l \in \mathcal{P}_j$ can be expressed as $t_{ijl} = y_{il} - y_{ij}$. This gap time can be the gap time between two events l and j or the gap time between a censoring time and an event time. For a given observation i , let v_i denote the indicator of an absorbing state, with $v_i = 1$ if an absorbing state is reached and $v_i = 0$ otherwise. Let $\mathcal{S}_i = \{s_{i1}, s_{i2}, \dots, s_{iJ_i}\}$ denote the set of states visited by the i -th subject, comprising a complete or incomplete path with $s_{i1} = 1$, where $J_i - 1$ is the number of states. We let $D = (n, \mathbf{t}, \mathbf{v}, \mathcal{S}_1, \dots, \mathcal{S}_n, \mathbf{Z}_1, \mathbf{Z}_2)$ denote the observed data, where $\mathbf{t}$ is a vector with elements $t_{i,1,s_{i2}}, \dots, t_{i,s_{iJ_i-1},s_{iJ_i}}$ for $i = 1, \dots, n$, $\mathbf{v} = (v_1, \dots, v_n)'$, $\mathbf{Z}_1$ is the $n \times p$ matrix of covariates with the i -th row $\mathbf{z}_{i1}'$, and $\mathbf{Z}_2$ is the $n \times q$ matrix of covariates with the i -th row $\mathbf{z}_{i2}'$.

If s_{iJ_i} is an absorbing state, $\ell(\mathcal{S}_i)$ denotes the number of the path corresponding to $\mathcal{S}_i$. If $v_i = 0$, then s_{iJ_i} is a transient state. In this case, let $\mathcal{U}(s_{iJ_i})$ stand for the set of subpaths from state s_{iJ_i} to every absorbing state through all the possible transient states. Each subpath $g \in \mathcal{U}(s_{iJ_i})$ has states $\{s_{i,g1}, s_{i,g2}, \dots, s_{i,gJ_{ig}}\}$, with $s_{i,g1} = s_{iJ_i}$ and such that $\mathcal{S}_i \cup g$ represents a complete path. For example, in Figure 1 we see that $\mathcal{U}(2) = \{\{2, 5\}, \{2, 6\}, \{2, 4, 5\}, \{2, 4, 6\}\}$. For a set $\mathcal{S}_i^*$ representing a complete path ($\mathcal{S}_i^*$ can be either $\mathcal{S}_i$ or $\mathcal{S}_i \cup g$), let

$$a(\tau_{\ell(\mathcal{S}_i^*)}) = \prod_{j=1}^{J_i-2} (1 + j\tau_{\ell(\mathcal{S}_i^*)}), \text{ if } J_i > 2; \quad a(\tau_{\ell(\mathcal{S}_i^*)}) = 1, \text{ if } J_i \leq 2, \quad (5)$$

noticing that $a(\cdot)$ is obtained when we integrate out w using the Laplace transformation of a $\text{gamma}(1/\tau_{\ell(\mathcal{S}_i^*)} + J_i - 1, 1/\tau_{\ell(\mathcal{S}_i^*)})$ random variable. If $\mathcal{S}_i$ is a complete path ($v_i = 1$), let

$$H^*(\mathcal{S}_i) = \left\{ 1 + \tau_{\ell(\mathcal{S}_i)} \sum_{j=1}^{J_i-1} H_{s_{ij}, s_{i,j+1}, 0}(t_{i,s_{ij}, s_{i,j+1}}) \exp(\mathbf{z}_{i2}' \boldsymbol{\beta}_{s_{ij}, s_{i,j+1}}) \right\}^{-1/\tau_{\ell(\mathcal{S}_i)} - J_i + 1}. \quad \text{If } \mathcal{S}_i \text{ is an incomplete path } (v_i = 0), \text{ the gap times are right-censored at } t_{i,s_{iJ_i}, *}. \text{ In this case, let}$$

$$H^*(\mathcal{S}_i \cup g) = [1 + \tau_{\ell(\mathcal{S}_i \cup g)} \{ \sum_{j=1}^{J_i-1} H_{s_{ij}, s_{i,j+1}, 0}(t_{i, s_{ij}, s_{i,j+1}}) \exp(z'_{i2} \beta_{s_{ij}, s_{i,j+1}}) + H_{s_{iJ_i}, s_{i,g2}, 0}(t_{i, s_{iJ_i}, *}) \times \exp(z'_{i2} \beta_{s_{iJ_i}, s_{i,g2}}) \}]^{-1/\tau_{\ell(\mathcal{S}_i \cup g)} - J_i + 1} \quad (6)$$

The vector $\theta = (\alpha', \beta', \lambda', \tau^*, \eta)'$ encapsulates all the parameters in our model, where τ^* refers to the two states paths ($|S| = 2$). The latent vector τ comprising all paths with $|S| > 2$ is included in the likelihood function to ease the computations in Section 5.2 and Web Appendix C. Using these notations and definitions, the likelihood function can be written as

$$L(\theta | \tau, D) = \prod_{i=1}^n \left\{ a(\tau_{\ell(\mathcal{S}_i)}) H^*(\mathcal{S}_i) \prod_{j=1}^{J_i-1} P(\delta_{i, s_{ij}} = s_{i,j+1} | z_{i1}, \alpha^{(s_{ij})}) h_{s_{ij}, s_{i,j+1}, 0}(t_{i, s_{ij}, s_{i,j+1}}) \right. \\ \times \exp(z'_{i2} \beta_{s_{ij}, s_{i,j+1}}) \left. \right\}^{I(\nu_i=1)} \left\{ \prod_{j=1}^{J_i-1} P(\delta_{i, s_{ij}} = s_{i,j+1} | z_{i1}, \alpha^{(s_{ij})}) h_{s_{ij}, s_{i,j+1}, 0}(t_{i, s_{ij}, s_{i,j+1}}) \exp(z'_{i2} \beta_{s_{ij}, s_{i,j+1}}) \right. \\ \times \sum_{g \in \mathcal{U}(\mathcal{S}_i)} a(\tau_{\ell(\mathcal{S}_i \cup g)}) H^*(\mathcal{S}_i \cup g) \prod_{j=1}^{J_{ig}-1} P(\delta_{i, s_{i,gj}} = s_{i,g(j+1)} | z_{i1}, \alpha^{(s_{i,gj})}) \left. \right\}^{I(\nu_i=0)} \quad (7)$$

For example, for a subject i in path P2, the contribution to the likelihood function is

$$P(\delta_{i1}=2 | z_{i1}, \alpha^{(1)}) P(\delta_{i2} = 5 | z_{i1}, \alpha^{(2)}) h_{120}(t_{i12}) h_{250}(t_{i25}) \exp\{z'_{i2}(\beta_{12} + \beta_{25})\} (1 + \tau_2) [1 + \tau_2 \{H_{120}(t_{i12}) \exp(z'_{i2} \beta_{12}) + H_{250}(t_{i25}) \exp(z'_{i2} \beta_{25})\}]^{-1/\tau_2 - 2} \quad (\text{for the remaining paths, see Web Appendix C}).$$

4. Model Properties

4.1 A Special Case: A Three States Model

The three states model is the most simple case of our general models, with one starting state (healthy), one transient state (diseased), and one absorbing state (dead). The data structure is shown in Figure S1 (in the online supplementary materials) and sometimes it is also referred as a disability model (Hougaard, 2000, Chapter 5). Under this special case, the Markov property is examined.

Proposition 1—The three states model corresponds to a homogeneous Markov renewal process when the baseline hazard function is constant.

4.2 Path Probability

Let $\mathcal{P}_j^k$ denote the subset of $\mathcal{P}_j$ such that $\mathcal{P}_j^k = \{l: \text{state } l \text{ is an immediate child state of state } j \text{ in path } P_k\}$. If j is an absorbing state, then $\mathcal{P}_j^k = \emptyset$. Based on the above models, the path probability can be computed as

$$p_k = p_k(z_1^c) = \prod_{j \in S_k, l \in \mathcal{P}_j^k} P(\delta_j = l | z_1^c, \alpha^{(j)}), \quad (8)$$

where z_1^c denotes the fixed values of some covariates upon which we condition and $\mathcal{P}_j^k = \emptyset$. This is the probability that a subject having characteristics z_1^c will eventually end up with path P_k . It is worthy to mention that under our model, the path probabilities are easy to compute. These probabilities can be useful for classifying subjects with a given set of characteristics. However, these path probabilities are computationally intensive under the nonparametric Markov, Cox semi-parametric Markov, and non-Markov models. Furthermore, these clinically important probabilities have not been examined in the literature, including, for example, Keiding et al. (2001) and Fiocco et al. (2008), for analyzing the bone marrow transplant data discussed in Section 2.

4.3 Survival Probability for Terminating Events

In Figure 1, we see that relapse and death in remission are terminating events. The relapse free probability is denoted by $P(T_5 > t | z_1^c, z_2^c)$, with z_1^c and z_2^c as in (8). Following the paths in Figure 1, we obtain

$$\begin{aligned} P(T_5 > t | z_1^c, z_2^c) = & \int_{\mathcal{X}_1^0} \int_{\mathcal{X}_2^0} \left[P(\delta_1 = 5 | z_1^0, z_1^c, \alpha^{(1)}) \{1 + \tau_1 H_{150}(t) e^{z_2' \beta_{15}}\}^{-1/\tau_1} \right. \\ & + P(\delta_1 = 2 | z_1^0, z_1^c, \alpha^{(1)}) P(\delta_2 = 5 | z_1^0, z_1^c, \alpha^{(2)}) \int_0^\infty P(T_{12} + T_{25} > t | z_2^0, z_2^c, \beta_{12}, \beta_{25}, w) \\ & \times f(w | \tau_2) dw + P(\delta_1 = 2 | z_1^0, z_1^c, \alpha^{(1)}) P(\delta_2 = 4 | z_1^0, z_1^c, \alpha^{(2)}) P(\delta_4 = 5 | z_1^0, z_1^c, \alpha^{(4)}) \\ & \times \int_0^\infty P(T_{12} + T_{24} + T_{45} > t | z_2^0, z_2^c, \beta_{12}, \beta_{24}, \beta_{45}, w) f(w | \tau_4) + P(\delta_1 = 3 | z_1^0, z_1^c, \alpha^{(1)}) \\ & \times P(\delta_3 = 4 | z_1^0, z_1^c, \alpha^{(3)}) P(\delta_4 = 5 | z_1^0, z_1^c, \alpha^{(4)}) \int_0^\infty P(T_{13} + T_{34} + T_{45} > t \\ & | z_2^0, z_2^c, \beta_{13}, \beta_{34}, \beta_{45}, w) f(w | \tau_6) dw + P(\delta_1 = 3 | z_1^0, z_1^c, \alpha^{(1)}) P(\delta_3 = 5 | z_1^0, z_1^c, \alpha^{(3)}) \\ & \times \int_0^\infty P(T_{13} + T_{35} > t | z_2^0, z_2^c, \beta_{13}, \beta_{35}, w) f(w | \tau_8) dw \Big] \int_{z_1^0, z_2^0} (z_1^0, z_2^0) dz_1^0 dz_2^0, \end{aligned} \quad (9)$$

where $\mathcal{X}_1^0$ and $\mathcal{X}_2^0$ denote the sample spaces of the free covariates z_1^0 and z_2^0 , respectively. The detailed computations of this probability and the survival probability $P(T_6 > t | z_1^c, z_2^c)$ are given in Web Appendix A. Since the covariates in the BMT data are categorical (see Table S2), integration with respect to the covariates is performed using their joint empirical distribution. A more general nonparametric scheme is presented in Zhang et al. (2014).

The piecewise constant hazard function in (4) facilitates the computation of the survival probability for terminating events. In particular, for the three and four state models in Figures S1 and S2 (in the online supplementary materials), the survival probability has a closed form expression, which can be derived using the results given in Web Appendix A.

Summary calculations, including state occupation and transition probabilities, as well as state entry and exit distributions (Ferguson et al., 2012), in models such as nonparametric Markov, Cox semi-parametric Markov, and non-Markov have been the subject of many studies, as we can see in the papers by Kalbfleisch and Lawless (1985), Pepe (1991), Keiding et al. (2001), Putter et al. (2007), Andersen and Perme (2008), Meira-Machado et al. (2009), and Ferguson et al. (2012), to name just a few. State occupation and transition

probabilities can be computed through equations like (9) and the results in Web Appendix A. Once we have the occupation probabilities, the state entry and exit distributions can be determined. For the sake of space, the derivations of these probabilities are omitted here for brevity.

5. Bayesian Inference

5.1 Prior and Posterior Distributions

The Bayesian framework requires the specification of a prior distribution for θ . We assume an improper joint prior distribution in which the components of θ are *a priori* independent with $\pi(\alpha) \propto 1$ and $\pi(\beta) \propto 1$. The prior for λ is specified as $\lambda_{jlm} \sim \text{gamma}(\gamma_{jlm1}, \gamma_{jlm2})$, with $\gamma_{jlm1} = 0$, $\gamma_{jlm2} = 0$, and $\pi(\lambda_{jlm} | \gamma_{jlm1}, \gamma_{jlm2}) \propto \lambda_{jlm}^{\gamma_{jlm1}-1} e^{-\gamma_{jlm2}\lambda_{jlm}}$, for $m=1, \dots, M_{jl}$ and all transitions $j \rightarrow l$. If $\gamma_{jlm1} = \gamma_{jlm2} = 0$, we get a Jeffreys-type prior for λ_{jlm} . In order to maintain a unique marginal distribution for the gap times, we further assume that $\tau_k | \eta \sim \text{exponential}(\eta)$ independently for the paths P_k such that $|S_k| > 2$, with density function $f(\tau_k | \eta) = \exp(-\tau_k/\eta)/\eta$, for $\tau_k > 0$. Then, after integrating out the frailty w and τ_k , there will be only one marginal distribution for the time T_{jl} even when the transition $j \rightarrow l$ may belong to more than one paths. The prior for η is taken as an inverse gamma distribution with $\pi(\eta) \propto (1/\eta)^{\gamma_{01}+1} \exp(-\gamma_{02}/\eta)$, where $\gamma_{01} > 0$ and $\gamma_{02} > 0$ are hyperparameters chosen to express vague prior distributions. For the components τ_k in τ^* , we assume independent inverse gamma distributions with hyperparameters $\gamma_{03k} > 0$ and $\gamma_{04k} > 0$. Hence, the prior has expression

$$\pi(\theta) \propto \left\{ \prod_{j \rightarrow l} \prod_{m=1}^{M_{jl}} \pi(\lambda_{jlm}) \right\} \pi(\tau^*) \pi(\eta), \quad (10)$$

so that the posterior distribution is given by

$$\pi(\theta | D) \propto \pi(\theta) \int L(\theta | \tau, D) \prod_{k: |S_k| > 2} f(\tau_k | \eta) d\tau_k. \quad (11)$$

In the sequel, we study some conditions in order to establish the propriety of the posterior distribution in (11).

Proposition 2—The likelihood function in (7) satisfies $L(\theta | \tau, D) = A_0 L_1(\alpha | D) L_2(\beta, \lambda | \tau, D)$, where A_0 is a positive constant,

$$L_1(\alpha | D) = \prod_{i=1}^n \left[\left\{ \prod_{j=1}^{J_i-1} P(\delta_{i,s_{ij}} = s_{i,j+1} | z_{i1}, \alpha^{(s_{ij})}) \right\}^{I(\nu_i=1)+I(\nu_i=0, J_i>1)} \right], \text{ and}$$

$$L_2(\beta, \lambda | \tau, D) = \prod_{i=1}^n \left[\left\{ a(\tau_{\ell(\mathcal{S}_i)}) H^*(\mathcal{S}_i) \prod_{j=1}^{J_i-1} h_{s_{ij}, s_{i,j+1}, 0}(t_{i,s_{ij}, s_{i,j+1}}) \exp(z'_{i2} \beta_{s_{ij}, s_{i,j+1}}) \right\}^{I(\nu_i=1)} \right].$$

For each parent state j , let $P^{(j)}$ be the set of $n_{P(j)}$ subjects visiting the paths, either complete or incomplete, such that the state j has at least one child, that is, $s_{ij_i} = j$. Define

$\mathbf{z}_{i1,1}^{(j)} = (\mathbf{z}_{i1,1}', \mathbf{0}', \dots, \mathbf{0}')', \dots, \mathbf{z}_{i1,|\mathcal{P}_j|-1}^{(j)} = (\mathbf{0}', \dots, \mathbf{0}', \mathbf{z}_{i1}')'$, and $\mathbf{z}_{i1,|\mathcal{P}_j|}^{(j)} = (\mathbf{0}', \dots, \mathbf{0}')'$, with $\mathcal{P}_j$ as in Section 3.1. Further define the $n_{\mathcal{P}(j)} \times p(|\mathcal{P}_j| - 1)$ matrix

$$\mathbf{Z}_1^{(j)} = \left[(\mathbf{z}_{i1,\delta_{ij}}^{(j)} - \mathbf{z}_{i1,l}^{(j)})' : l \neq \delta_{ij}, l \in \mathcal{P}_j, \text{ and } i \in \mathcal{P}^{(j)} \right]$$

Theorem 1—Assume the prior distribution given in (10). For each parent state j , assume further that the following conditions hold: (i) The matrix $\mathbf{Z}_1^{(j)}$ is of full rank and (ii) There exists a positive vector $\mathbf{e} \in \mathbb{R}^{n_{\mathcal{P}(j)}}$ such that $\mathbf{Z}_1^{(j)'} \mathbf{e} = \mathbf{0}$. Then, $\int L_1(\boldsymbol{\alpha}|D) d\boldsymbol{\alpha} < \infty$ if and only if conditions (i) and (ii) hold.

The binary case ($|\mathcal{P}_j| = 2$) studied by Chen and Shao (2001) is a special case of Theorem 1.

Theorem 2—Assume the prior distribution in (10) with $\gamma_{jlm1} = \gamma_{jlm2} = 0$. Assume further that for each transition $j \rightarrow l$, there exists at least one complete path $\mathcal{P}$ with $n_{\mathcal{P}}$ subjects such that $\mathbf{Z}_{2,jl}$ is of full rank, where $\mathbf{Z}_{2,jl}$ is an $n_{\mathcal{P}} \times (M_{jl} + q)$ matrix with rows $(I_{ijl1}, \dots, I_{ijlM_{jl}}, \mathbf{z}_{i2}')$ for $i \in \mathcal{P}$. Then, $\int L_2(\boldsymbol{\beta}, \boldsymbol{\lambda} | \boldsymbol{\tau}, D) f(\boldsymbol{\tau} | \boldsymbol{\eta}) \pi(\boldsymbol{\eta}) \pi(\boldsymbol{\lambda}) d\boldsymbol{\beta} d\boldsymbol{\lambda} d\boldsymbol{\tau} d\boldsymbol{\eta} < \infty$.

By combining the results in Proposition 2, Theorem 1, and Theorem 2, we conclude that the posterior distribution in (11) is proper, that is, $\int \pi(\boldsymbol{\theta}|D) d\boldsymbol{\theta} < \infty$.

Roy and Hobert (2007, Appendix A) present a linear programming problem to check condition (ii) in Theorem 1. In Theorem 2, we essentially require that at least one event occurs in each interval $(c_{jl,m-1}, c_{jlm}]$ and the corresponding covariate matrix is of full rank. The rank conditions in Theorems 1 and 2 can be checked numerically.

5.2 Bayesian Computations and Model Comparison

The analytical form of the posterior distribution in (11) is not available. Therefore, we develop an efficient Gibbs sampling scheme (Robert and Casella, 2004) to draw samples from the posterior distribution. To this end, we introduce many latent variables and perform reparameterizations. The details of our computational development are given in Web Appendix C. Bayesian computations using the Gibbs sampler were implemented in the FORTRAN language using IMSL subroutines with double precision arithmetic. The convergence of the Gibbs sampler was checked using several diagnostic tools discussed in Robert and Casella (2004).

To carry out Bayesian model comparison, we consider the deviance information criterion (DIC) and the logarithm of the pseudo marginal likelihood (LPML). We define the deviance $\text{Dev}(\boldsymbol{\vartheta}) = -2 \log L(\boldsymbol{\theta} | \boldsymbol{\tau}, D)$, where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\tau})'$ and $L(\boldsymbol{\theta} | \boldsymbol{\tau}, D)$ is given in (7). Let $\bar{\boldsymbol{\vartheta}}$ and $\overline{\text{Dev}} = E\{\text{Dev}(\boldsymbol{\vartheta}|D)\}$ denote the posterior means of $\boldsymbol{\vartheta}$ and $\text{Dev}(\boldsymbol{\vartheta})$, respectively. According to Spiegelhalter et al. (2002), the DIC measure is defined as $\text{DIC} = \text{Dev}(\bar{\boldsymbol{\vartheta}}) + 2p_D^\theta$, where $p_D^\theta = \overline{\text{Dev}} - \text{Dev}(\bar{\boldsymbol{\vartheta}})$ is the effective number of model parameters. The smaller the DIC value, the better the model fits the data. The posterior means $\bar{\boldsymbol{\vartheta}}$ and $\overline{\text{Dev}}$ can be estimated by

$$\bar{\boldsymbol{\vartheta}} = \sum_{j=1}^B \boldsymbol{\vartheta}_j / B \text{ and } \overline{\text{Dev}} = \sum_{j=1}^B \text{Dev}(\boldsymbol{\vartheta}_j) / B, \text{ where } \boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_B \text{ are samples from the}$$

posterior distribution. LPML is another useful Bayesian measure of goodness-of-fit, which is defined based on the conditional predictive ordinate (CPO). For the i -th observation, we define CPO as $\text{CPO}_i = \int L(\boldsymbol{\theta} | \boldsymbol{\tau}, D_i) \pi(\boldsymbol{\theta} | D^{(-i)}) d\boldsymbol{\theta}$, where D_i is the observed data for the i -th subject, $L(\boldsymbol{\theta} | \boldsymbol{\tau}, D_i)$ is the likelihood for the i -th subject, which is the term inside the product in (7), $D^{(-i)}$ is the data with D_i deleted, and $\pi(\boldsymbol{\theta} | D^{(-i)})$ is the posterior density of $\boldsymbol{\theta}$ based on the data $D^{(-i)}$. According to Geisser and Eddy (1979) and Gelfand and Dey (1994), an

approximation is given by $\text{LPML} = \sum_{i=1}^n \log(\hat{\text{CPO}}_i)$, where

$\hat{\text{CPO}}_i = \left[\sum_{j=1}^B 1/L(\boldsymbol{\theta}_j | \boldsymbol{\tau}, D_i) \right]^{-1} / B$. The larger the LPML value, the better the model fits the data.

6. Simulation Studies

In this section, we conduct a simulation study to assess some properties of the Bayesian model. For the sake of simplicity and taking into account that the simulations are time consuming, our study was carried out with the four states model in Figure S2 (in the online supplementary materials). In Figure S2, there are $K = 3$ possible complete paths and their corresponding states and transitions are given as follows: P1: $1 \rightarrow 4$, P2: $1 \rightarrow 2 \rightarrow 3$, and P3: $1 \rightarrow 2 \rightarrow 4$. There are two possible incomplete paths as well, namely, IP1: 1 and IP2: $1 \rightarrow 2$. In the data generation with $p = 3$, $q = 2$, and $z_{i11} = 1$, we first generate n independent $z_{i13} = z_{i22} \sim N(0, 1)$ and given z_{i13} , we sample $z_{i12} = z_{i21} \sim \text{Bernoulli}\{1/\exp(-0.5 - 0.3z_{i13})\}$, independent, $i = 1, \dots, n$, with $n = 1000$. These values remain fixed throughout the 500 repetitions of the simulations. Sample sizes of about 1000 and even bigger are not uncommon in multi-state data sets.

For the transition $1 \rightarrow 4$ we adopt the proportional hazards model. The likelihood function is derived from (7). The prior distribution in (10) is as in Theorem 2 with $\gamma_{jlm1} = \gamma_{jlm2} = 0$, for all j , l , and m , whereas γ_{01} and γ_{02} are such that the prior mean and variance of η are $(\tau_2 + \tau_3)/2$ and twice the mean, respectively. For τ_1 , the prior is inverse gamma(10, 1). In the Gibbs sampling algorithm, after discarding the first 2000 iterations, the next 3000 iterations were used for posterior inference. In our simulations, the gap times are generated from an exponential model with cumulative baseline $H_0(t) = t$, $\tau_2 = 2.0$, and $\tau_3 = 1.5$. The multinomial probabilities and the gap times are generated from (1) and (3), whereas the censoring time has a uniform distribution on $(0.3t_{\max}, 1.5t_{\max})$, with $t_{\max} = 80$ chosen to control the censoring rate, leading to 21.6%, 30.4%, 32.0%, 12.3%, and 3.7% of observations on average in the paths P1, P2, P3, IP1, and IP2, respectively. The true values of the parameters are given in Table 1. Note that the fitted model is not the true model, since the data are generated under a proportional hazards assumption for the transition $1 \rightarrow 4$.

In the first part of our simulation study, we compare the fits of the Markov, semi-Markov, and frailty models. In the semi-Markov model, the gap time of the previous transition is included as a covariate in the model for the hazard function, as in (3). The hazard functions are given by (4). The likelihood functions for the Markov and semi-Markov models follow from the expressions in Hougaard (1999, Sections 3.2 and 4). These models were fitted using specific Gibbs sampling algorithms with the same prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ as in the frailty model. Samples from the posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are drawn applying the

same techniques described in Web Appendix C. For the sake of space, the details are omitted.

For each model used in the generation of the simulated data sets (true model) in Table S3 (in the online supplementary materials), we fit the frailty, Markov, and semi-Markov models. The averages of LPML and DIC over the 500 repetitions of the simulations, as well as the percentage of data sets in which the true model was selected (correctness), are shown in Table S3. When the model with frailties in paths P2 and P3 is the true one, our frailty model (even not the true model), by far, outperforms the Markov and semi-Markov models. On the other hand, when the data are generated from the Markov model, both the Markov and semi-Markov models yield similar results, so that the correctness of the Markov model, less than 100%, is not unexpected.

Moreover, we also assess some frequentist properties of the Bayesian model. The results are shown in Table 1. In Table 1(a), the biases are almost negligible. There is a good agreement between the averages of the posterior standard deviation (SD) and the root mean squared error of the posterior means (RMSE), suggesting that the uncertainty in the estimates is adequately accounted for. The coverage probabilities of the 95% highest posterior distribution (HPD) intervals differ from the nominal value by at most 2.8%.

Table 1(b) shows the survival probabilities for terminating events, which are computed using the results established in Web Appendix A. The true probabilities are obtained by integrating out z_1 and z_2 with respect to their true joint distribution. Since our main goal here is to estimate the survival probabilities, we take just one piece for each transition with $\lambda_{121} = 0.5$, $\lambda_{141} = 1.2$, $\lambda_{231} = 0.8$, and $\lambda_{241} = 0.5$, whereas the true values of α , β , and η are those in Table 1(a). When the events are more extreme, the estimators are more biased. However, the biases are reduced when the sample size increases to 2000. Furthermore, SD and RMSE decrease at a rate of about $n^{-1/2}$. The coverage probabilities of the 95% HPD intervals are in general close to the nominal value for all events. On the whole, despite the small number of samples for posterior inference when compared to Section 7, the results in Table 1 demonstrate a good performance of the point and interval Bayesian estimators under the scenarios considered in our study and using the piecewise constant baseline hazard described in Section 3.2.

7. Analysis of the BMT Data

We carry out a detailed analysis of the BMT data described in Section 2. The prognostic factors in Table S2 are the covariates both for the probabilities in (1) and for the hazard functions in (3). In (10) the hyperparameters were set at $\gamma_{jlm1} = \gamma_{jlm2} = 0$, for all j , l , and m , $\gamma_{01} = \gamma_{02} = 0.01$, and $\gamma_{03} = \gamma_{04} = 5$. When running the Gibbs sampling algorithm, the first 2000 iterations were discarded. Then, we performed 200,000 additional iterations with thinning equal to 20, leading to 10,000 samples for each parameter. Table S4 (in the online supplementary materials) shows the values of the comparison criteria for frailty models having different numbers of intervals in the transitions. According to the DIC and LPML values, we select model 6 as our working model, noticing that model 6i denotes the independent gap time (i.e., no-frailty) model. For model 6, Figures S3–S6 (in the online

supplementary materials) show the trace plots for α_{34} , β_{15} , λ_{24} , and $(\tau_1, \dots, \tau_{10})'$, respectively, whereas Figure S7 shows the trace, histogram, ergodic mean, and autocorrelation plots for η . These plots, as well as the omitted ones, indicate that the chains have a good mixing and converge. Regarding $\pi(\tau_1)$ and $\pi(\tau_{10})$, a sensitivity analysis for model 6 with values of γ_{03} and γ_{04} in $\{1, 2, 5, 10, 50\}$ shows that the best fit is achieved when $\gamma_{03} = \gamma_{04} = 5$.

Next we present some results obtained from the samples of the posterior distribution. Table 2 shows the posterior means and standard deviations for the regression coefficients under model 6. A coefficient is significant when its 95% HPD interval does not contain 0. We see, for example, that at a 5% level, donor recipient gender mismatch is not significant for the probabilities in (1). For the transition $4 \rightarrow 5$, there are no significant effects of the prognostic factors. With respect to the hazards in (3), for the transitions $1 \rightarrow 5$, $2 \rightarrow 5$, $3 \rightarrow 4$, and $3 \rightarrow 5$, the effects of the prognostic factors are not significant.

The point estimate and the HPD interval for the frailty standard deviation η are 0.151 and (0.048, 0.291), respectively. The posterior means of $\tau_2, \dots, \tau_9$ in (2) range from 0.016 (P5) to 0.323 (P3), so that there is some heterogeneity in the frailty distribution for the paths with more than two states. We also see that the largest means correspond to the shortest paths (P3, P8, P2, and P9). For paths P1 and P10, the posterior means of τ_1 and τ_{10} are 3.92 and 1.61, giving an indication that the proportional hazards assumption is not tenable.

Figure 2(a) and (b) display the differences in the path probabilities taking the age class ≤ 20 years at transplant as reference class and fixing the levels of the remaining prognostic factors. In some cases, there is a significant difference between the path probabilities for the upper age classes with respect to the reference age class. In Table 2, the posterior means of the components in α_{12} , α_{13} , and α_{15} corresponding to the age classes are negative and the coefficients are significant, so that the results for paths P1 and P10 would be expected. In Figure 2(c), we consider only the age class after integrating out with respect to the remaining factors, as discussed in Section 4.3. Paths P1, P3, and P7 stand out as the most probable paths, with the posterior means of path probabilities (0.23, 0.18, 0.14), (0.20, 0.18, 0.21), and (0.18, 0.14, 0.16), respectively, for the age classes ≤ 20 , (20, 40], and > 40 .

In Figure 3(a), we see that donor recipient gender mismatch has a short time decreasing significant effect on the relapse free probability. With respect to the survival probability in Figure 3(b), the behavior is not monotonic and the effect of gender mismatch is not significant.

The effect of age at transplant on the relapse free probability is not significant, as shown in Figure 4(a). In Figure 4(b), we see that the survival probability of the youngest patients at transplant becomes greater than the probability of older patients for after a few months. The HPD intervals for the differences in survival probabilities do not contain 0 after about 3 years. For example, when the time is set at 6 years, the posterior means of the differences for the age classes (20, 40] and > 40 are -0.0498 and -0.0533 , with HPD intervals $(-0.0850, -0.0125)$ and $(-0.0947, -0.0128)$, respectively. When the time is set at 10 years, the

posterior means are -0.0535 and -0.0613 with HPD intervals $(-0.0895, -0.0176)$ and $(-0.102, -0.0206)$, respectively.

8. Discussion and Future Directions

When the number of internal states is large and the sample size is not sufficiently large, posterior propriety may be an issue due to the small number of events in certain paths. In this case, the path diagram is sparse and the conditions in Theorem 2 may not be satisfied. However, under the Bayesian approach, with a moderately informative prior we can handle the sparsity of the data. Certainly, a sensitivity analysis on the specification of priors needs to be carried out. On the other hand, if there are too many paths, the model may not be clinically useful. By collapsing some transient states, we can obtain a simpler path diagram.

Although lots of work have been done, there are various areas for future research. Our proposed model covers the situations in which the path diagram is unidirectional. The methodology can be extended to a more general setting such as the disability model in Figure S1 (in the online supplementary materials) allowing for recovery from the disease. We emphasize that under our model, the path probabilities are easily estimated as a by-product of the Gibbs sampler. Paths with outstanding estimates might be useful to clinicians and can be used to classify the patients into different risk groups. The inclusion of time dependent covariates imposes additional challenges. In our formulation, the conditions in Section 5.1 require that all complete paths must be observed in the data. Otherwise, more informative prior distributions should be elicited from experts or historical data, if available, can be used to construct a power prior distribution (see, e.g., Chen et al., 1999).

Our model assumes equal frailties for the gap times of a given subject. This can be extended to a correlated frailty model (Wienke, 2011, Chapter 5) with transition specific frailties in a given path. A multivariate log-normal distribution for the vector of frailties $\mathbf{w} = (w_{jl} : j, l \in S_k)'$ can replace (2), allowing a more flexible correlation structure for the gap times in a path. Moreover, the gamma frailty imposes a positive correlation between two gap times in the same path. This can be relaxed by using a copula model (Wienke, 2011, Chapter 6) with path specific parameters.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the Editor, an Associate Editor, and two referees for their constructive comments and suggestions, which have led to an improved version of the paper. Dr. de Castro's research was partially supported by CNPq, Brazil. Dr. Chen's research was partially supported by US NIH grants #GM 70335 and #CA 74015.

References

- Aalen OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*. 1978; 5:141–150.
- Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research*. 2002; 11:91–115. [PubMed: 12040698]

- Andersen PK, Perme MP. Inference for outcome probabilities in multi-state models. *Lifetime Data Analysis*. 2008; 14:405–431. [PubMed: 18791824]
- Andersen, PK.; Perme, MP. Multistate models. In: Klein, JP.; van Houwelingen, HC.; Ibrahim, JG.; Scheike, TH., editors. *Handbook of Survival Analysis*. Boca Raton: Chapman & Hall/CRC; 2013. p. 417–439.
- Armero C, Cabras S, Castellanos ME, Perra S, Quirós A, Oruezábal MJ, Sánchez-Rubio J. Bayesian analysis of a disability model for lung cancer survival. *Statistical Methods in Medical Research*. 2012;10.1177/0962280212452803
- Banerjee T, Chen MH, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Analysis*. 2007; 13:241–260. [PubMed: 17401683]
- Chen MH, Shao QM. Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*. 2001; 129:293–302.
- Chen MH, Ibrahim JG, Yiannoutsos C. Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society B*. 1999; 61:223–242.
- Commenges D. Multi-state models in epidemiology. *Lifetime Data Analysis*. 1999; 5:315–327. [PubMed: 10650740]
- Conlon ASC, Taylor JMG, Sargent DJ. Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in Medicine*. 2014; 33:1750–1766. [PubMed: 24307330]
- de Wreede LC, Fiocco M, Putter H. mstate: an R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*. 2011; 38:1–30.
- Ferguson N, Datta S, Brock G. msSurv: an R package for nonparametric estimation of multistate models. *Journal of Statistical Software*. 2012; 50:1–24. [PubMed: 25317082]
- Fiocco M, Putter H, van Houwelingen HC. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine*. 2008; 27:4340–4358. [PubMed: 18425994]
- Geisser S, Eddy WF. A predictive approach to model selection. *Journal of the American Statistical Association*. 1979; 74:153–160.
- Gelfand AE, Dey DK. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*. 1994; 56:501–514.
- Hougaard P. Multi-state models: a review. *Lifetime Data Analysis*. 1999; 5:239–264. [PubMed: 10518372]
- Hougaard, P. *Analysis of Multivariate Survival Data*. New York: Springer; 2000.
- Huzurbazar, AV. *Flowgraph Models for Multistate Time-to-Event Data*. Hoboken: Wiley; 2005.
- Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*. 1985; 80:863–871.
- Keiding N, Klein JP, Horowitz MM. Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*. 2001; 20:1871–1885. [PubMed: 11406847]
- Kim Y, James L, Weissbach R. Bayesian analysis of multistate event history data: beta-Dirichlet process prior. *Biometrika*. 2012; 99:127–140.
- Kneib T, Hennerfeind A. Bayesian semi parametric multi-state models. *Statistical Modelling*. 2008; 8:169–198.
- Meira-Machado L, de Uña Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*. 2009; 18:195–222. [PubMed: 18562394]
- Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*. 1991; 86:770–778.
- Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*. 2007; 26:2389–2430. [PubMed: 17031868]
- Robert, CP.; Casella, G. *Monte Carlo Statistical Methods*. 2. New York: Springer; 2004.
- Roy V, Hobert JP. Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society B*. 2007; 69:607–623.

- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*. 2002; 64:583–639.
- Titman AC, Sharples LD. Model diagnostics for multi-state models. *Statistical Methods in Medical Research*. 2010a; 19:621–651. [PubMed: 19654169]
- Titman AC, Sharples LD. Semi-Markov models with phase-type sojourn distributions. *Biometrics*. 2010b; 66:742–752. [PubMed: 19912172]
- Warr RL, Huzurbazar AV. Expanding the statistical flowgraph model framework to use any transition distribution. *Journal of Statistical Theory and Practice*. 2010; 4:529–539.
- Wienke, A. *Frailty Models in Survival Analysis*. Boca Raton: Chapman & Hall/CRC; 2011.
- Xu J, Kalbfleisch JD, Tai B. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*. 2010; 66:716–725. [PubMed: 19912171]
- Zeng D, Chen Q, Chen MH, Ibrahim JG. Amgen Research Group. Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. *Biometrika*. 2012; 99:167–184. [PubMed: 23049136]
- Zhang Y, Chen MH, Ibrahim JG, Zeng D, Chen Q, Pan Z, Xue X. Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime Data Analysis*. 2014; 20:76–105. [PubMed: 23543121]
- Zhao, L. Doctoral dissertation. Department of Statistics and Actuarial Science. Simon Fraser University; Canada: 2009. Multi-state Processes with Duration-dependent Transition Intensities: Statistical Methods and Applications.

Figure 1.
Path diagram of the BMT data.

Figure 2.

Differences in the path probabilities and 95% HPD intervals according to age at transplant class in years under model 6 (reference class: 20, left: (20, 40], and right: > 40). (a) Donor-recipient gender mismatch: no, prophylaxis: no, and year of transplant: 1990–1994, (b) donor-recipient gender mismatch: yes, prophylaxis: yes, and year of transplant: 1985–1989, and (c) path probabilities and 95% HPD intervals according to age at transplant class in years under model 6 (left: 20, middle: (20, 40], and right: > 40).

Figure 3.

Differences and 95% HPD intervals (dashed lines) according to donor recipient gender mismatch under model 6 (reference class: no gender mismatch). (a) Relapse free probability and (b) survival probability.

Figure 4. Differences and 95% HPD intervals (dashed lines) according to age at transplant class in years under model 6 (reference class: 20, “●”: (20, 40] and “○”: > 40). (a) Relapse free probability and (b) survival probability.

Table 1

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

(a)					
Par	True	Est	SD	RMSE	CP
α_{121}	1.0	1.00	0.12	0.12	0.964
α_{122}	-0.5	-0.51	0.09	0.09	0.932
α_{123}	0.5	0.51	0.16	0.15	0.954
α_{231}	0.7	0.71	0.16	0.16	0.938
α_{232}	1.0	1.02	0.11	0.12	0.938
α_{233}	-1.0	-1.02	0.19	0.19	0.952
β_{121}	1.0	1.01	0.08	0.08	0.944
β_{122}	-1.5	-1.50	0.15	0.16	0.940
β_{121}	-1.5	-1.51	0.08	0.08	0.960
β_{142}	-0.5	-0.50	0.15	0.14	0.958
β_{231}	0.3	0.30	0.11	0.11	0.962
β_{232}	-0.5	-0.50	0.21	0.21	0.942
β_{241}	-1.5	-1.50	0.11	0.11	0.948
β_{242}	-1.0	-1.01	0.20	0.21	0.942
τ_2	2.0	1.98	0.23	0.22	0.948
τ_3	1.5	1.52	0.21	0.20	0.950
λ_{121}	1.0	1.02	0.13	0.13	0.952
λ_{141}	1.0	1.09	0.13	0.15	0.936
λ_{231}	1.0	1.02	0.16	0.16	0.960
λ_{241}	1.0	1.05	0.20	0.21	0.938
(b)					
t_0	True $P(T_3 > t_0)$	Est	SD	RMSE	CP
1	0.338	0.339	0.020	0.019	0.952

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(b)						
t_0	True $P(T_3 > t_0)$	Est	SD	RMSE	CP	
5	0.240	0.231	0.038	0.039	0.958	
10	0.190	0.177	0.050	0.053	0.960	
20	0.145	0.128	0.063	0.066	0.956	
<hr/>						
	$n = 2000$	0.136	0.043	0.043	0.946	
<hr/>						
t_0	True $P(T_4 > t_0)$	Est	SD	RMSE	CP	
1	0.484	0.480	0.019	0.018	0.948	
5	0.332	0.323	0.031	0.032	0.974	
10	0.258	0.244	0.044	0.046	0.962	
20	0.189	0.172	0.059	0.062	0.954	
<hr/>						
	$n = 2000$	0.181	0.043	0.042	0.962	
<hr/>						

Table 2

Posterior means (posterior standard deviations) of the regression coefficients under model 6 for the BMT data.

Multinomial probabilities (α)													
Intercept and prognostic factor	Categories	Transition											
		1→2	1→3	1→5	2→4	2→5	3→4	3→5	4→5				
Intercept		1.61 (0.10)	1.77 (0.10)	0.94 (0.11)	-0.70 (0.09)	-1.29 (0.12)	0.84 (0.09)	0.32 (0.11)	-1.60 (0.12)				
Donor recipient gender mismatch	Yes	-0.18 (0.20)	-0.16 (0.20)	0.03 (0.22)	0.37 (0.20)	0.43 (0.26)	0.03 (0.21)	-0.27 (0.24)	-0.49 (0.27)				
Prophylaxis	Yes	-0.41* (0.20)	-0.39* (0.20)	0.12 (0.22)	-0.53* (0.23)	0.31 (0.27)	-0.14 (0.21)	-0.56* (0.24)	-0.05 (0.28)				
Year of transplant	1990–1994	0.95 (0.22)	0.64 (0.21)	0.39 (0.23)	-0.22 (0.24)	-0.23 (0.30)	1.24* (0.23)	0.83* (0.23)	0.11 (0.29)				
	1995–1998	1.14 (0.24)	0.57 (0.23)	0.49 (0.26)	-0.13 (0.24)	0.09 (0.31)	1.44* (0.24)	0.15 (0.27)	-0.32 (0.32)				
Age at transplant (years)	(20,40]	-0.66* (0.25)	-0.51* (0.25)	-0.92* (0.26)	0.56* (0.22)	-0.27 (0.27)	-0.85* (0.24)	-0.31 (0.26)	0.34 (0.28)				
	> 40	-0.60* (0.29)	-0.64* (0.29)	-1.21* (0.32)	0.72* (0.25)	-0.21 (0.31)	-1.03* (0.29)	-0.67* (0.33)	0.00 (0.35)				
Hazards (β)													
Prognostic factor	Categories	Transition											
		1→2	1→3	1→5	1→6	2→4	2→5	2→6	3→4	3→5	3→6	4→5	4→6
Donor recipient gender mismatch	Yes	-0.01 (0.10)	-0.23* (0.09)	0.29 (0.37)	-0.25 (0.39)	-0.43* (0.16)	0.25 (0.30)	0.28 (0.43)	-0.08 (0.12)	-0.22 (0.37)	-0.75* (0.24)	0.70* (0.31)	0.49* (0.18)
Prophylaxis	Yes	-0.41* (0.11)	-0.04 (0.09)	0.62 (0.36)	-0.79* (0.37)	-0.12 (0.21)	0.05 (0.34)	-0.04 (0.40)	-0.20 (0.13)	0.53 (0.34)	-0.27 (0.21)	0.13 (0.31)	-0.16 (0.22)
Year of transplant	1990–1994	-0.06 (0.12)	-0.14 (0.10)	0.49 (0.39)	0.79* (0.38)	-0.17 (0.20)	0.13 (0.37)	-0.75 (0.41)	0.05 (0.14)	-0.27 (0.31)	-0.12 (0.22)	-1.03* (0.30)	-0.33 (0.23)
	1995–1998	0.11 (0.12)	-0.13 (0.10)	0.34 (0.45)	0.49 (0.44)	-0.05 (0.21)	0.30 (0.37)	-0.83 (0.46)	0.02 (0.14)	-0.38 (0.46)	0.07 (0.26)	-0.68* (0.33)	-0.32 (0.23)
Age at transplant (years)	(20,40]	0.36* (0.11)	-0.18 (0.09)	-0.09 (0.37)	-0.10 (0.50)	-0.40 (0.20)	-0.25 (0.38)	0.17 (0.52)	-0.02 (0.12)	0.24 (0.38)	-0.37 (0.27)	0.19 (0.34)	0.82* (0.28)
	> 40	0.30* (0.12)	-0.23* (0.11)	-0.02 (0.52)	-0.17 (0.57)	-0.27 (0.22)	-0.10 (0.36)	1.55* (0.51)	0.20 (0.15)	0.60 (0.46)	-0.48 (0.31)	0.86* (0.40)	1.38* (0.29)

For the estimates marked with *, the 95% HPD intervals do not contain 0.