Biometrics, 2024, 80(3), ujae075
https://doi.org/10.1093/biomtc/ujae075
Biometric Methodology

The multivariate Bernoulli detector: change point estimation in discrete survival analysis

Willem van den Boom [1,*], Maria De Iorio [1,2], Fang Qian[1], Alessandra Guglielmi [3]

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore, [2]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research , Singapore 117609, Singapore, [3]Department of Mathematics, Politecnico di Milano, Milan 20133, Italy

*Corresponding author: Willem van den Boom, Yong Loo Lin School of Medicine, National University of Singapore, 1E Kent Ridge Road, NUHS Tower Block, Level 12, Singapore 119228, Singapore (vandenboom@nus.edu.sg).

ABSTRACT

Time-to-event data are often recorded on a discrete scale with multiple, competing risks as potential causes for the event. In this context, application of continuous survival analysis methods with a single risk suffers from biased estimation. Therefore, we propose the multivariate Bernoulli detector for competing risks with discrete times involving a multivariate change point model on the cause-specific baseline hazards. Through the prior on the number of change points and their location, we impose dependence between change points across risks, as well as allowing for data-driven learning of their number. Then, conditionally on these change points, a multivariate Bernoulli prior is used to infer which risks are involved. Focus of posterior inference is cause-specific hazard rates and dependence across risks. Such dependence is often present due to subject-specific changes across time that affect all risks. Full posterior inference is performed through a tailored local-global Markov chain Monte Carlo (MCMC) algorithm, which exploits a data augmentation trick and MCMC updates from nonconjugate Bayesian nonparametric methods. We illustrate our model in simulations and on ICU data, comparing its performance with existing approaches.

KEYWORDS: Bayesian statistics; competing risks; discrete failure time models; discrete time-to-event data; grouped survival data; local-global Markov chain Monte Carlo.

1 INTRODUCTION

Time-to-event data are common in many applications such as finance, medicine, and engineering with examples including time to payment and survival and failure times. Most approaches to survival data consider continuous event times. Nevertheless, it is more and more common to record time-to-event data on a discrete scale (eg, patient-reported outcomes or time to pregnancy measured in number of menstrual cycles it takes a couple to conceive). See Schmid and Berger (2020) for further examples. Discrete recording of the timings of events (Allison, 1982) may occur when time is truly discrete, or when continuous time is partitioned into nonoverlapping intervals (corresponding, for instance, to days, weeks, or months) and only the interval in which an event occurs is recorded (King and Weiss, 2021). This special case of interval censoring is usually referred to as *grouped time* and, in this work, we consider discrete survival models that arise as probabilistic grouped versions of continuous-time frailty models from survival analysis (see, eg, Hougaard, 1986, 1995; Andersen et al., 1993, Chapter IX; Hougaard et al., 1994; Kalbfleisch and Prentice, 2002). Indeed, direct application of continuous-time methods to discretely recorded data may result in biased estimation (Lee et al., 2018).

Moreover, our focus is on discrete survival data in the presence of multiple, competing risks that can cause an event. Traditional analyses often consider a single risk with events due to other risks, for example, other causes of death, treated as censoring. However, this generally violates the common assumption of independent censoring (Schmid and Berger, 2020). Moreover, such analyses can lead to misestimation of hazards and covariate effects (Andersen et al., 2012). Here, we consider data from the Medical Information Mart for Intensive Care IV (MIMIC-IV, Johnson et al., 2023) database on length of stay in an intensive care unit (ICU), typically reported discretely in days. While the MIMIC-IV database documents admission and discharge times down to the minute, it is recommended to perform survival analysis using discrete day units. This approach is preferred because admission and discharge times within a single day are significantly determined by hospital protocols and staff decisions, rather than purely reflecting the patients' health conditions. The study considers 3 competing events that can terminate an ICU stay: discharge to home (69.0%), transfer to another medical facility (21.4%), and in-hospital death (6.1%).

The 2 main approaches for competing risks with discrete times are cause-specific hazard functions and the subdistribution hazard model (Schmid and Berger, 2020). The latter is more suitable when interest is in one out of many risks. The first approach usually exploits methods from generalized linear models (GLMs), enabling maximum likelihood estimation, variable selection, and other methods from GLMs, such as in Tutz (1995) and Möst et al. (2016). Our work places itself within this approach, focusing on scenarios with few risks and thus considering cause-specific hazards, that is, a hazard

Received: August 21, 2023; Revised: May 23, 2024; Accepted: July 24, 2024

function for each risk, introducing dependence across risks by building on recent advances in change point analysis.

A characteristic of traditional discrete survival models as compared to GLMs is that they present unconstrained baseline hazards. This leads to a large number of parameters to estimate and, as a consequence, to unstable estimation, especially if for certain time points the number of events is small. To improve stability, regularization of hazard functions has been proposed. See, for example, Luo et al. (2016), Heyard et al. (2019), and Möst et al. (2016). Fahrmeir and Wagenpfeil (1996) and King and Weiss (2021) employ random walks to smooth the hazard function. In all these works, cause-specific baseline hazards are treated independently. Vallejos and Steel (2017) focus on risk-specific covariate selection, still assuming independence across risks. On the other hand, dependence across risks is plausible since multiple hazards can be affected by changes to the individual across time, and as such should be incorporated in the model.

The main methodological contribution of this work lies in modeling explicit and interpretable dependencies across risks. We introduce a multivariate change point model for baseline hazards, which offers 2 key advantages (Kozumi, 2000). Firstly, it reduces the number of parameters, promoting parsimony. Secondly, it accommodates the variability in hazard rates across different survival times. We adopt a Bayesian approach, assigning priors on the number and location of overall change points, thereby inducing marginal dependence among them. For each overall change point, we use a multivariate Bernoulli prior to determine which risks are involved, a method previously applied in time series analysis (Dobigeon et al., 2007; Harlé et al., 2016). We term our approach the multivariate Bernoulli detector, building on prior work by Harlé et al. (2016).

Change points have been widely studied in continuous survival analysis (eg, Matthews and Farewell, 1982; Goodman et al., 2011) but less so in the discrete case: Kozumi (2000) considers a single risk modeled via a Markov chain with a prespecified number of change points, while Wang and Ghosh (2007) use posterior predictive checks for change point detection, without allowing for covariates in the model. On the other hand, we allow for covariates in the model and perform cause-specific variable selection. Moreover, we estimate the number of change points, testing for the presence and location of change points using Bayes factors or posterior probabilities.

The paper is structured as follows. Section 2 introduces the model and describes a tailored Markov chain Monte Carlo (MCMC) sampler. Section 3 discusses an application to the ICU data. Section 4 compares our approach with existing ones. We conclude the paper in Section 5.

2 MODEL

2.1 Setup and notation

We follow the notation in Tutz and Schmid (2016). Let T_i denote the time to event for individual $i \in \{1, \ldots, n\}$, where n is the number of individuals. In the discrete-time setting, $T_i \in \{1, \ldots, t_{\max} + 1\}$ for some maximum time $t_{\max}$. The random variable T_i can, for instance, arise as the discretization, also known as grouping (Kalbfleisch and Prentice, 2002), of a latent continuous time T_i^c into $t_{\max} + 1$ intervals $[\omega_0, \omega_1), [\omega_1, \omega_2), \ldots, [\omega_{t_{\max}}, \infty)$ with $\omega_0 = 0$. In this case, there is a one-to-one correspondence between the set of integers $\{1, \ldots, t_{\max} + 1\}$ and the intervals of the real line where the continuous-time random variables T_i^c are defined. As such, the interpretation of $T_i = t_{\max} + 1$ is similar to censoring, in the sense that the event will occur at time $t > t_{\max}$ (ie, $T_i^c \geq \omega_{t_{\max}}$) and $t_{\max} + 1$ is simply a "latent time" that groups together individuals for which it is known that the even has not occurred up to time $t_{\max}$. The time-to-event distribution is usually characterized by the overall hazard function $\lambda(t \mid \theta_i) = P(T_i = t \mid T_i \geq t, \theta_i) = P\{T_i^c \in [\omega_{t-1}, \omega_t) \mid T_i^c \geq \omega_{t-1}, \theta_i\}$ for some vector of parameters θ_i.

Additionally, we assume that observations are subject to censoring. That means that only a portion of the observed times can be considered as exact survival times. Let C_i be the censoring time of individual i. C_i assumes values in $\{1, \ldots, t_{\max}\}$, with T_i and C_i independent (random censoring). Moreover, we assume that the censoring mechanism is noninformative, that is, it does not depend on any parameters used to model the event times (Schmid and Berger, 2020). Let $\delta_i = \mathbb{1}[T_i \leq C_i]$ be a censoring indicator, where $\mathbb{1}[\cdot]$ denotes the indicator function, and $t_i = \min(T_i, C_i)$ the observed time.

We consider competing risks with m different types of events. For instance, the events can correspond to death due to m different causes. Then, $R_i \in \{1, \ldots, m\}$ denotes the event type experienced by individual i at time T_i, for which we observe a value r_i only in the absence of censoring, that is, $\delta_i = 1$. Finally, the cause-specific hazard function is $\lambda_r(t \mid \theta_i) = P(T_i = t, R_i = r \mid T_i \geq t, \theta_i)$ such that $\lambda(t \mid \theta_i) = \sum_{r=1}^{m} \lambda_r(t \mid \theta_i)$.

2.2 Likelihood

We assume independence across individuals such that the likelihood is a product over individual-specific terms. For ease of explanation, we consider the likelihood contribution of 1 individual and drop subscripts i in the remainder of this section unless otherwise specified. Under the assumption that C is independent of T and θ, the likelihood for θ is given by

$$P(T = t,\ R = r \mid \theta)^{\delta} P(T > t \mid \theta)^{1-\delta} = \lambda_r(t \mid \theta)^{\delta} \prod_{l=1}^{t-\delta} \{1 - \lambda(l \mid \theta)\} \quad (1)$$

and we specify $\lambda_r(t \mid \theta)$ and thus $\lambda(t \mid \theta)$ according to the multinomial logit model, which is the most popular for categorical responses (Tutz and Schmid, 2016). Specifically, $\lambda_r(t \mid \theta) = \exp(\eta_{rt}) / \{1 + \sum_{\rho=1}^{m} \exp(\eta_{\rho t})\}$, where $\eta_{rt} = \alpha_{rt} + \boldsymbol{x}^\top \boldsymbol{\beta}_r$ is a cause- and time-specific linear predictor, and $\theta = \{\eta_{rt}\}$. The intercept α_{rt} represents the cause-specific baseline hazard. The p-dimensional vector $\boldsymbol{\beta}_r = (\beta_{r1}, \ldots, \beta_{rp})$ consists of the cause-specific regression coefficients of the covariates in the p-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_p)$. Such likelihood specification is also known as proportional continuation ratio model (Tutz and Schmid, 2016) as increasing x_j by one unit increases the cause-specific odds by the factor $\exp(\beta_{rj})$. Note that, when

$r = 1$ and $\delta = 1$, $\lambda_r(t \mid \theta)$ corresponds to ϕ_t defined in Section 2.5, and the considerations on prior specification there discussed will be relevant when building a change point model on α_{rt}.

2.3 The multivariate Bernoulli detector

We propose a model on the baseline hazards that is flexible, yet has interpretable structure. Specifically, the sequence $\alpha_{r1}, \ldots, \alpha_{r,t_{\max}}$ is set to follow a piecewise constant function. We do so through a change point model with dependence across risks r. In our setup, a change point corresponds to a time point where the hazard of at least 1 risk r changes. We specify a hierarchical prior on the change points, which has 3 main components: (i) a prior on the number of change points; (ii) a prior on the location of change points; and (iii) a prior on which risks (at least 1) have a change point at that particular time location, given that a change point at time t occurs.

2.3.1 *Prior specification on overall change points*

In this section, we describe prior specification on the number and location of change points. Let $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{mt})$. Then, $\gamma_t = \mathbb{1}[\boldsymbol{\alpha}_t \neq \boldsymbol{\alpha}_{t-1}]$ indicates whether there is an overall change point at $t \in \{1, \ldots, t_{\max}\}$, that is, if the hazard of at least a risk changes at time t. Furthermore, $K = \sum_{t \in \mathcal{T}} \gamma_t$ denotes the number of change points. Here, $\mathcal{T}$ defines the set of possible change point locations. We specify the prior on $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{t_{\max}})$ hierarchically, by specifying a prior $p(K)$ on the number of change points and then $p(\boldsymbol{\gamma} \mid K)$ as this provides explicit regularization on K: i.i.d. γ_t would imply a binomial distribution on K.

To motivate the next model development, consider the bottom plots in Figure 2, where 2 observations with $t_i = 7$ out of 500 are not used when inferring change points. Then, the lack of observations at time $t = 7$ results in spurious change points at that time location and the next. We restrict our inference to avoid such sensitivity to a few observations: To aid identifiability, considering the flexibility of the underlying time-varying geometric distribution, which is discussed in Section 2.5, we only allow change points for a subset of times $\mathcal{T} \subset \{1, \ldots, t_{\max}\}$ such that $\gamma_t = 0$ if $t \notin \mathcal{T}$. Firstly, as it is typical in change point applications, we do not allow change points at the support boundary, in our case $t = 1$ and $t = t_{\max}$. Moreover, we do not allow a change point at time t if both t and $t - 1$ have no observed events as the data lack information on which of the 2 time points would be a change point. Also, we do not allow change points at a time t with no observed events if both neighboring times $t - 1$ and $t + 1$ have observed events, because this would lead to spurious change points due to the flexibility of the underlying time-varying geometric, as seen in Figure 2 (bottom row). On the other hand, we prefer to introduce parsimony in the estimation of change points to improve interpretability. We explore the effect of the restriction on change point locations in a simulation study in Web Section E.4. There, the restriction (i) does not deteriorate inference, even if the true change point is not in $\mathcal{T}$; (ii) avoids spurious change points at time locations without observed events.

We assume a geometric distribution with success probability π_K truncated to $K \leq |\mathcal{T}|$ as prior on the number of change points. We denote such prior as $\text{Geo}_{|\mathcal{T}|}(\pi_K)$. For the locations of overall change points given K, we use the uniform distribution on possible configurations $p(\boldsymbol{\gamma} \mid K) = 1/\binom{|\mathcal{T}|}{K}$. In summary, the joint prior on the number and location of change points has a hierarchical specification: $p(K, \boldsymbol{\gamma}) = p(K)\, p(\boldsymbol{\gamma} \mid K)$.

2.3.2 *Cause-specific change point configuration*

In this section, we discuss the prior on which risks present a jump in the hazard, given the vector $\boldsymbol{\gamma}$. Let $z_{rt} = \mathbb{1}[\alpha_{rt} \neq \alpha_{r(t-1)}]$ be an indicator variable denoting if a change point occurs at time t for risk r. If there is no change point at t for any r, then $\gamma_t = 0$ and $z_{rt} = 0$. If $\gamma_t = 1$, then $z_{rt} = 1$ for at least one r.

Let $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{mt})$. Conditionally on $\gamma_t = 1$, we assume that $\boldsymbol{z}_t$ follows a multivariate Bernoulli distribution (eg, Teugels, 1990). An m-dimensional binary vector $\boldsymbol{z}_t$ can assume 2^m possible values corresponding to a combination of $z_{rt} \in \{0, 1\}$. The multivariate Bernoulli distribution is then parameterized by a 2^m-dimensional vector, whose elements correspond to the probability of each particular outcome (ie, configuration). In our case, when modeling $\boldsymbol{z}_t$ given $\gamma_t = 1$, we exclude the configuration of all zeros, that is, $z_{rt} = 0$ for every r. Thus, $\boldsymbol{z}_t$ can assume only $2^m - 1$ possible values. We denote such distribution as $\text{Ber}_0(\boldsymbol{\psi})$, where $\boldsymbol{\psi}$ denotes the $(2^m - 1)$-dimensional vector of configuration probabilities. In summary, the prior specification for $\boldsymbol{z}_t$ is

$$\boldsymbol{z}_t \mid \gamma_t \sim \begin{cases} \text{Ber}_0(\boldsymbol{\psi}) & \text{if } \gamma_t = 1 \\ \delta_{\mathbf{0}} & \text{if } \gamma_t = 0 \end{cases}$$

where $\delta_{\mathbf{0}}$ is a point mass at the zero vector. We refer to the joint prior on $(K, \boldsymbol{\gamma}, \boldsymbol{z})$ as multivariate Bernoulli detector, where $\boldsymbol{z} = \{\boldsymbol{z}_t\}_{t=1}^{t_{\max}}$.

2.4 Further prior specification

Model specification is completed by specifying independent prior distributions on α_{rt} and $\boldsymbol{\beta}_r$. We specify a prior on α_{rt} conditionally on the number and location of change points. Since $\boldsymbol{\alpha}_r = (\alpha_{r1}, \ldots, \alpha_{rt_{\max}})$ is a piecewise constant function for each risk r, assuming constant values between change points, let $\alpha^\star_{r\ell}$ denote the unique value of α_{rt} over each time interval for risk r. Note that for each risk, a change point can be activated or not, with the only constraint that a change point needs to be activated for at least 1 risk. We assume $\alpha^\star_{r\ell} \sim \mathcal{N}(\mu_\alpha, \sigma^2_\alpha)$ independently across ℓ and r.

Furthermore, to identify important effects, we assume a variable selection prior for the regression coefficients, $\boldsymbol{\beta}_r = (\beta_{r1}, \ldots, \beta_{rp})$, which allows for risk-specific variable selection. Here, we consider a spike-and-slab prior (Mitchell and Beauchamp, 1988): $\beta_{rj} \sim \pi_\beta\, \mathcal{N}(0, \sigma^2_\beta) + (1 - \pi_\beta)\, \delta_0$, where π_β is the prior inclusion probability. We use the hyperprior $\pi_\beta \sim \mathcal{U}(0, 1)$. In the application in Section 3, some variables are grouped as they are dummy variables associated with a categorical covariate. We modify the prior accordingly to perform groupwise variable selection as detailed in Web Appendix B. We note that other possible prior choices are available in the literature to perform variable selection, such as shrinkage priors (Bhadra et al., 2019), which offer computational advantages at the cost of depending on arbitrary thresholds to identify relevant effects.

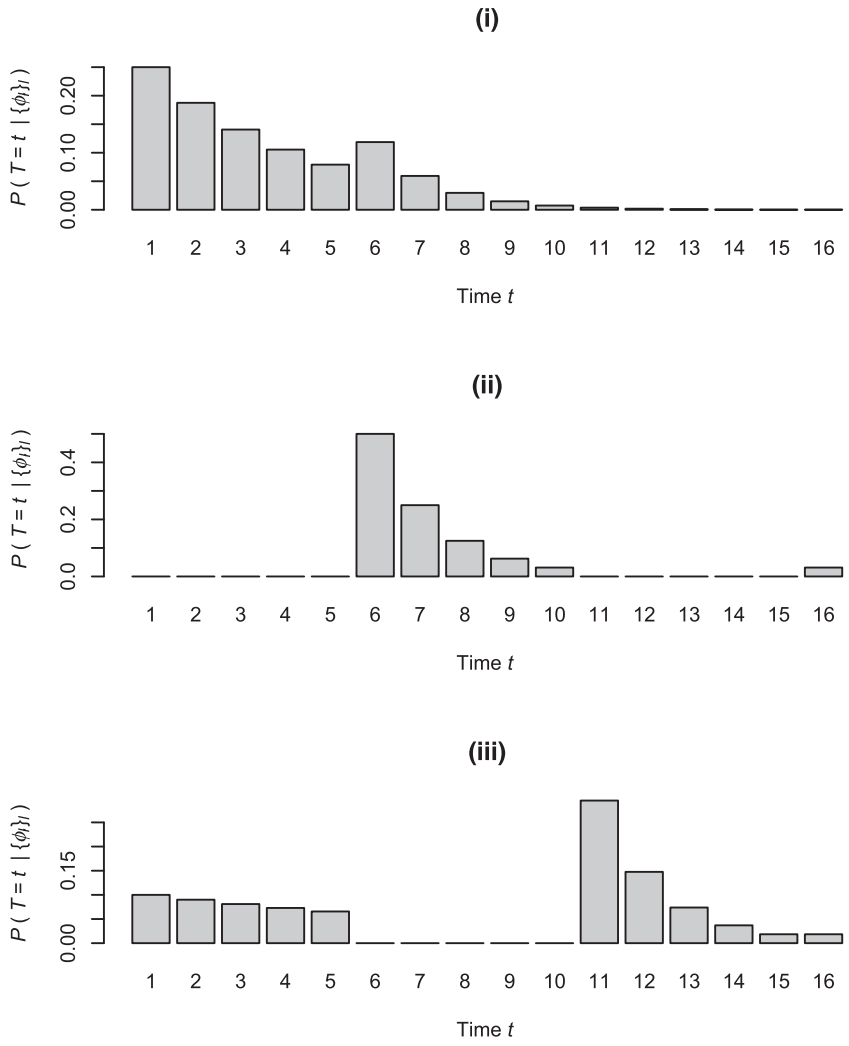

FIGURE 1 Probability mass function of the time-varying geometric distribution: visualizations of Equation 2 with $t_{\max} = 15$ and success probabilities (i) $\phi_t = 0.25$ for $t \leq 5$ and $\phi_t = 0.5$ otherwise; (ii) $\phi_t = 0$ for $t \leq 5$ or $t \geq 11$, and $\phi_t = 0.5$ otherwise; and (iii) $\phi = 0.1$ for $t \leq 5$, $\phi_t = 0.5$ for $t \geq 11$, and $\phi_t = 0$ otherwise.

In the simulation studies and the application on ICU data, we set the parameters as follows: $\sigma_\beta^2 = 1$, $\pi_K = 0.5$, all elements of $\boldsymbol{\psi}$ equal to $1/(2^m - 1)$, $\mu_\alpha = -9$, and $\sigma_\alpha^2 = 3$. The particular prior choice for $\alpha^\star_{r\ell}$ derives from the interpretation of the model in terms of the time-varying geometric. In Section 2.5, we highlight the importance of shrinking the probabilities ϕ_t toward zero. This is equivalent, in absence of covariates, to shrinking $\exp(\alpha^\star_{r\ell})/\{1 + \exp(\alpha^\star_{r\ell})\}$ toward zero and, consequently, $\alpha^\star_{r\ell}$ toward -9. Roughly speaking, a $\mathcal{N}(-9, 3)$ on $\alpha^\star_{r\ell}$ is equivalent to a Beta$(0.01, 1)$ prior on ϕ_t, which is shown to have a good performance in Section 2.5. Finally, we note that we could specify a prior on $\boldsymbol{\psi}$ to favor sparsity or a large number of change points.

2.5 Rationale behind modeling strategy

Prior specification for the parameters governing the multivariate Bernoulli detector (see Section 2.4) is derived from the following considerations. In the uncensored ($\delta = 1$) single-risk ($m = 1$) case, the distribution of the time to event in the discrete case is a time-varying geometric distribution (Landau and Zachmann, 2019) with time-varying success probability $\phi_t = \lambda(t \mid \theta)$, that is,

$$P(T = t \mid \{\phi_l\}_l) = \begin{cases} \phi_t \prod_{l=1}^{t-1}(1 - \phi_l), & t = 1, \ldots, t_{\max} \\ \prod_{l=1}^{t-1}(1 - \phi_l), & t = t_{\max} + 1 \end{cases}. \quad (2)$$

The time-varying geometric is fully flexible in that it can represent any distribution on $\{1, \ldots, t_{\max} + 1\}$ by appropriately choosing ϕ_t (Mandelbaum et al., 2007). It is analogous to the piecewise exponential distribution in continuous survival analysis (Gamerman, 1991), if we assume a change point model for the ϕ_t. See Figure 1 for widely varying realizations of the distribution for certain $\{\phi_t\}_t$. This flexibility should be taken into account when inferring ϕ_t. It relates to the potential lack of stability of unconstrained estimation of baseline hazards mentioned in Section 1. For instance, there is a separate parameter ϕ_t for each time point, but we might not have observed an event at each time point. To avoid such overpameterization, some subsequent ϕ_t can be assumed to be equal to each other, as in Figure 1, resulting in a change point model: A change point is a time t for which $\phi_t \neq \phi_{t-1}$. Moreover, given

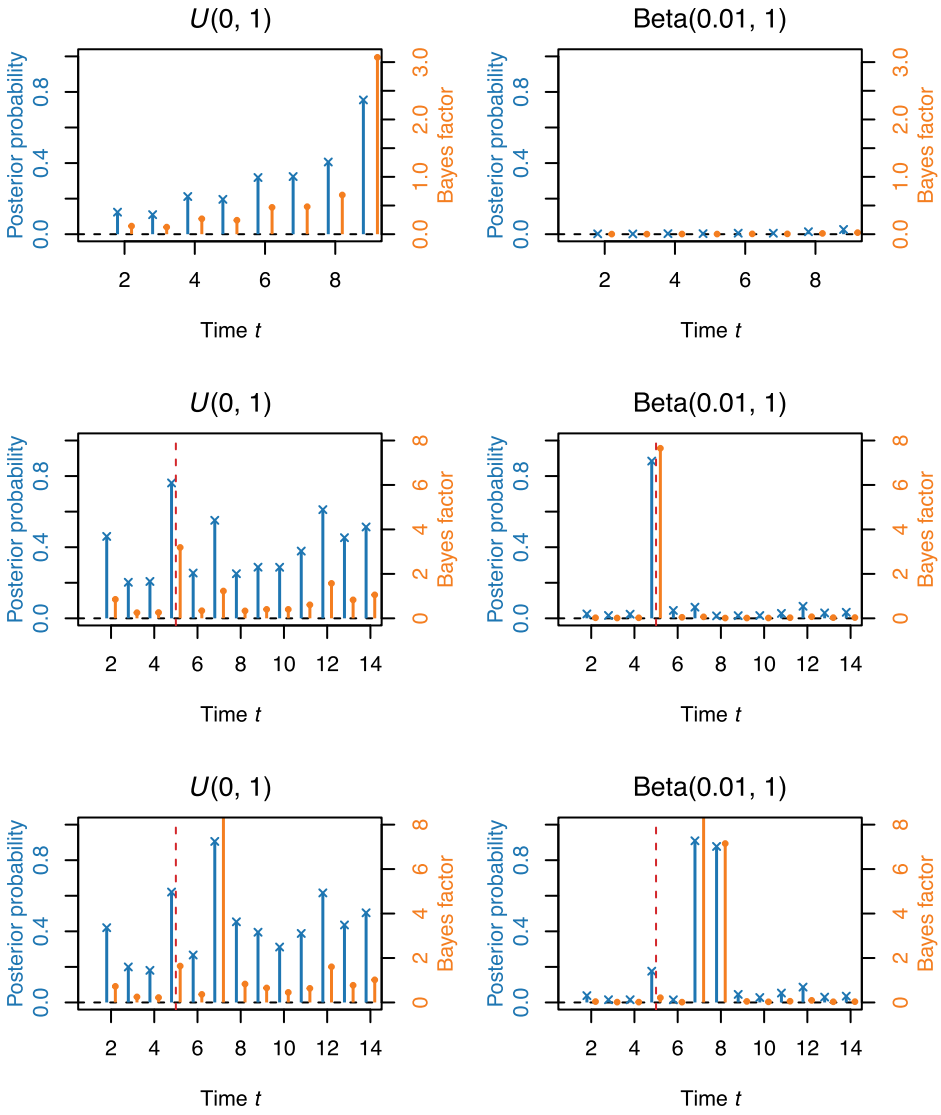

FIGURE 2 Time-varying geometric simulation: posterior probabilities (×) and Bayes factors (•) for the presence of a change point with a uniform prior (left column, $\phi_\ell^\star \sim \mathcal{U}(0, 1)$) and regularization toward zero (right column, $\phi_\ell^\star \sim \text{Beta}(0.01, 1)$) when simulating data without (top row) and with (middle and bottom rows) a change point. The bottom row uses the data from the middle row without the observations with $t_i = 7$. Some Bayes factors are outside the plotting range. The dashed lines are drawn in correspondence of the true change point.

the flexibility of the time-varying geometric, different combinations of $\{\phi_t\}$ can lead to a satisfactory fit of a data set, leading to identifiability problems. As such, we impose prior regularization by a priori shrinking the value of ϕ_t toward zero. We further motivate our prior choice in the following simulation study.

We simulate $n = 500$ times t_i from Equation 2 with $t_{\max} = 1000$ using 2 different settings for ϕ_t. We consider a scenario without change points with $\phi_t = 0.5$ and a scenario with a single change point given by $\phi_t = 0.5$ for $t \leq 4$ and $\phi_t = 0.25$ for $t \geq 5$. For this last scenario, we also consider the data after removal of observations with $t_i = 7$, which we discuss in Section 2.3.1 in relation to the prior constraints on change point location. To understand how the prior on ϕ_t can affect inference on change points, we compare 2 priors within a Bayesian change point model defined as follows: We specify a uniform prior over all possible change point configurations among the ϕ_t. Let $\phi_\ell^\star$ denote the unique value of ϕ_t over each time interval delimited by the change points. Conditionally on a change point configuration, we choose a prior on $\phi_\ell^\star$. Then, the likelihood in Equation 2 completes the model. We fit this model with $t_{\max} = \max_i t_i$, such that $t_{\max} = 9$ for the data with no change point and $t_{\max} = 14$ for the data with a change point. We compare posterior inference obtained with a uniform prior, $\phi_\ell^\star \sim \mathcal{U}(0, 1)$, and with a prior that shrinks the parameters toward zero, $\phi_\ell^\star \sim \text{Beta}(0.01, 1)$.

Figure 2 shows the inference on change points. The uniform prior (left column) leads to the detection of too many change points, especially at larger t. Regularization toward zero using $\phi_\ell^\star \sim \text{Beta}(0.01, 1)$ (right column) yields more accurate posterior inference without spurious change points.

2.6 Posterior computation using local-global MCMC

To devise an MCMC scheme to perform posterior inference, we exploit the representation of the likelihood in Equation 1 as a multinomial logistic regression (Tutz and Schmid, 2016), using a data augmentation trick. This results in the availability of conjugate updates, leading to more efficient mixing and preventing, at the same time, large changes in the configuration of change points, resulting in more effective local moves. The latent variables associated with the data augmentation are highly correlated with the change points, and, as such, it is difficult to explore the change points space conditionally on the latent variables. To counterbalance this drawback, we also devise global moves of change points conditionally on the observed data. Such moves are based on ideas from the Bayesian nonparametric literature (Dahl, 2005; Martínez and Mena, 2014; Creswell et al., 2020). Finally, from the MCMC output, we can derive Bayes factors to test for the presence of change points using the Savage–Dickey ratio (Dickey, 1971; Verdinelli and Wasserman, 1995) (see Web Appendix C for details).

We refer to the resulting hybrid algorithm as "local-global MCMC" borrowing the terminology from Samsonov et al. (2022). Here, we provide a brief explanation of our MCMC strategy in relation to previous work. Web Appendix D details the algorithm.

2.6.1 *Local MCMC with data augmentation*

We exploit the data augmentation representation of a multinomial logistic regression in terms of Gumbel latent variables by McFadden (1974) and Frühwirth-Schnatter and Frühwirth (2007). Then, the augmented likelihood is Gaussian, which enables convenient MCMC updates. Importantly, conditionally on the z_{rt}, the $\alpha^{\star}_{r\ell}$ have a Gaussian prior such that they can be integrated out from the augmented posterior, enabling efficient updates of z_{rt} and γ_t without having to specify Metropolis–Hastings proposals for $\alpha^{\star}_{r\ell}$.

More recently, other augmentations have been proposed in the literature (see, for instance, Held and Holmes, 2006; Frühwirth-Schnatter and Frühwirth, 2010; Polson et al., 2013; Linderman et al., 2015). We do not opt for them because they do not provide a convenient augmented likelihood in the presence of multiple risks.

2.6.2 *Global MCMC*

Augmented data can strongly reflect the change points of the current state of the MCMC chain, resulting in local MCMC updates to the change point parameters z_{rt} and γ_t. Therefore, we also consider MCMC moves without data augmentation, that is, based on the original data, to enable more global change point updates and explore better posterior space.

Specifically, we exploit the fact that change points induce a partition of time into intervals and apply ideas from Bayesian nonparametrics (Dahl, 2005; Martínez and Mena, 2014; Creswell et al., 2020) to deal with nonconjugate updates. This allows for more global moves at the cost of having to specify Metropolis–Hastings proposals for $\alpha^{\star}_{r\ell}$. Alternating between local and global MCMC steps allows for better mixing and convergence of the algorithm.

We demonstrate the performance of our approach in simulation studies. Web Appendix E presents simulation studies with a wide range of scenarios and comparison with alternative models. We find that the multivariate Bernoulli detector generally results in the most accurate estimation. Prior shrinkage of baseline hazards can result in estimation bias, which is a common feature of Bayesian shrinkage priors and the bias-variance trade-off they induce (eg, Polson and Sokolov, 2019).

3 APPLICATION TO ICU LENGTH OF STAY

3.1 Data description and analysis

We apply our model to data on ICU stays from the MIMIC-IV database (Johnson et al., 2023) with length of stay as outcome (Meir and Gorfine, 2023).

See Web Appendix A for a detailed data description. We consider $m = 3$ competing risks: discharge to home, transfer to another medical facility, and in-hospital mortality. Length of stay is recorded in days with the longest uncensored time being 28 days. We analyse $n = 25\,159$ ICU stays with 17 357 discharged to home, 5379 transferred, 1529 deaths, and 894 censored. We include the following covariates: demographics, variables related to the ICU stay (eg, whether it is a repeat admission), and lab tests from the first day. Most covariates are categorical with 2 or more levels. Representing them as dummy variables leads to a total of $p = 36$ predictors.

We fit our model with $t_{\max} = \max_i t_i = 28$ days using 200 000 MCMC iterations, discarding the first 50 000 as burn-in. The trace plots in Web Figure 14 suggest satisfactory convergence.

3.2 Posterior inference on the baseline hazards

The posterior probability of absence of overall change points is zero as well as the Bayes factor (see Web Appendix C). Figure 3 summarizes inference on the baseline hazards (see Web Figure 15 for corresponding Bayes factors). The hazard functions differ markedly between risks: the hazard of discharge to home is high in the first 2 weeks, but not on the first day in the ICU. The hazard of a transfer to another medical facility is lowest during the first few days. Finally, the hazard function for in-hospital mortality does not vary substantially across the length of stay.

3.3 Posterior inference on the regression coefficients

The regression coefficients β_{rj} capture cause-specific covariate effects on length of stay. The spike-and-slab prior provides explicit inference on whether there is a covariate effect through the posterior probability of $\beta_{rj} \neq 0$. Posterior inclusion probabilities for each risk are shown in Figure 4. In Figure 5, we report posterior inference on regression coefficients. Finally, we remark that the results on the baseline hazards and covariate effects are in line with those obtained in Meir and Gorfine (2023).

4 COMPARISON WITH OTHER MODELS

We compare our results on the ICU data to those obtained from maximum likelihood estimation and a more recent alternative, namely the model by King and Weiss (2021).

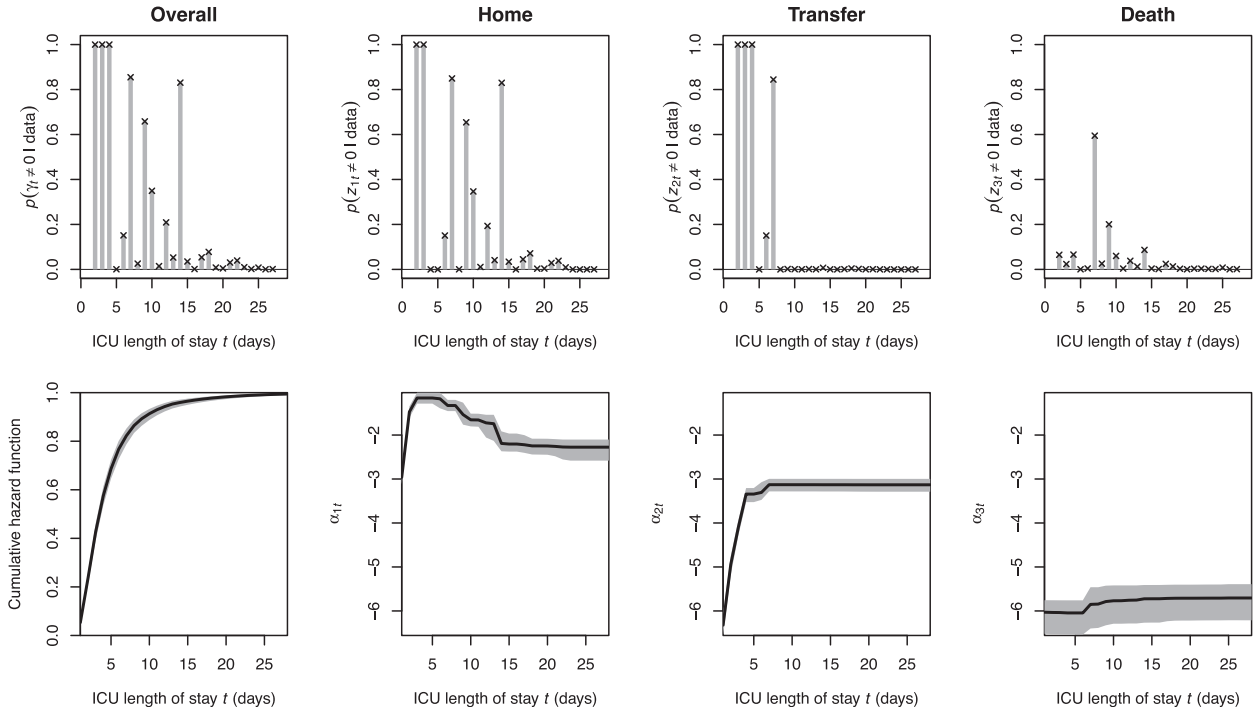

FIGURE 3 ICU data: posterior inference on the overall (left column) and cause-specific (other columns) baseline hazards. The top row displays the posterior probabilities for the presence of a change point. The bottom row shows posterior inference for the cumulative hazard function for $\boldsymbol{x}_i = \mathbf{0}$ and the baseline hazard parameter α_{rt}. Black lines represent posterior means and shaded areas correspond to 95% credible intervals.

4.1 Maximum likelihood estimation

We maximize the likelihood in Equation 1 using the R package `nnet` (Venables and Ripley, 2002) as the likelihood is equivalent to a multinomial logistic regression (Tutz and Schmid, 2016). The resulting inference is shown in Web Figures 16 and 17. Estimates of $\boldsymbol{\alpha}_t$ are less smooth than for our model but are otherwise similar. Estimates of $\boldsymbol{\beta}_r$ are in agreement with ours.

4.2 Semi-parametric model by King and Weiss (2021)

We also compare our model with the Bayesian semi-parametric model by King and Weiss (2021), which also involves multiple risk and a flexible model for the hazard function. For simplicity of explanation in what follows, we denote with x_{ij} a continuous covariate j for individual i and with d_{ik} a dummy variable corresponding to a level of a categorical covariate. King and Weiss (2021) specify a multinomial logit model for discrete survival analysis with competing risks with $\eta_{irt} = \alpha_{rt} + \sum_{j=1}^{p^c} f_{\beta rj}(x_{ij}) + \sum_{k=1}^{p^d} \beta_{rk}\, d_{ik}$, where p^c and p^d denote the number of continuous and dummy variables, respectively. Moreover, $\alpha_{tr} = \beta_{0r} + f_{\alpha r}(t)$ for intercepts β_{0r}, and functions $f_{\alpha r}$ and $f_{\beta rj}$, which are object of inference. Note that in their approach, King and Weiss (2021) include every level of a categorical covariate. The functions $f_{\alpha r}$ and $f_{\beta rj}$ are inferred using a Gaussian Markov random field prior. For prior specification and parameter choice, we follow the recommendations in Appendix C of King and Weiss (2021) for uninformative priors. We fit the model using the R package `brea` (King, 2017) using 50 000 burn-in MCMC iterations followed by 200 000 recorded iterations.

The resulting inference is shown in Web Figures 19-21. The estimates of baseline hazards are in line with our model, though not as smooth. The nonlinear covariate effect of age is consistent with the linear effect from our model, but positive association of age and transfer hazard only appears at an older age. The other covariate effects are also similar to the results from our model. See Figure 5.

5 DISCUSSION

In this work, we focus on the estimation of the hazard function of competing risks in the context of discrete survival. We assume a change point model for the hazard function, with cause-specific change points, introducing dependence among change point locations across risks. In our approach, both number and location of change points are random. We refer to our model as multivariate Bernoulli detector. Dependence across risks provides an attractive way for regularization of baseline hazards since changes to an individual's condition across time might affect multiple cause-specific hazards simultaneously. Our approach is widely applicable and interpretable. The data augmentation enables the use of any prior on regression coefficients making the MCMC updates more efficient. The simulation study and the real data application show that posterior inference on change points with dependence across risks is effective, with favorable comparisons with a frequentist approach and the Bayesian semi-parametric model by King and Weiss (2021).

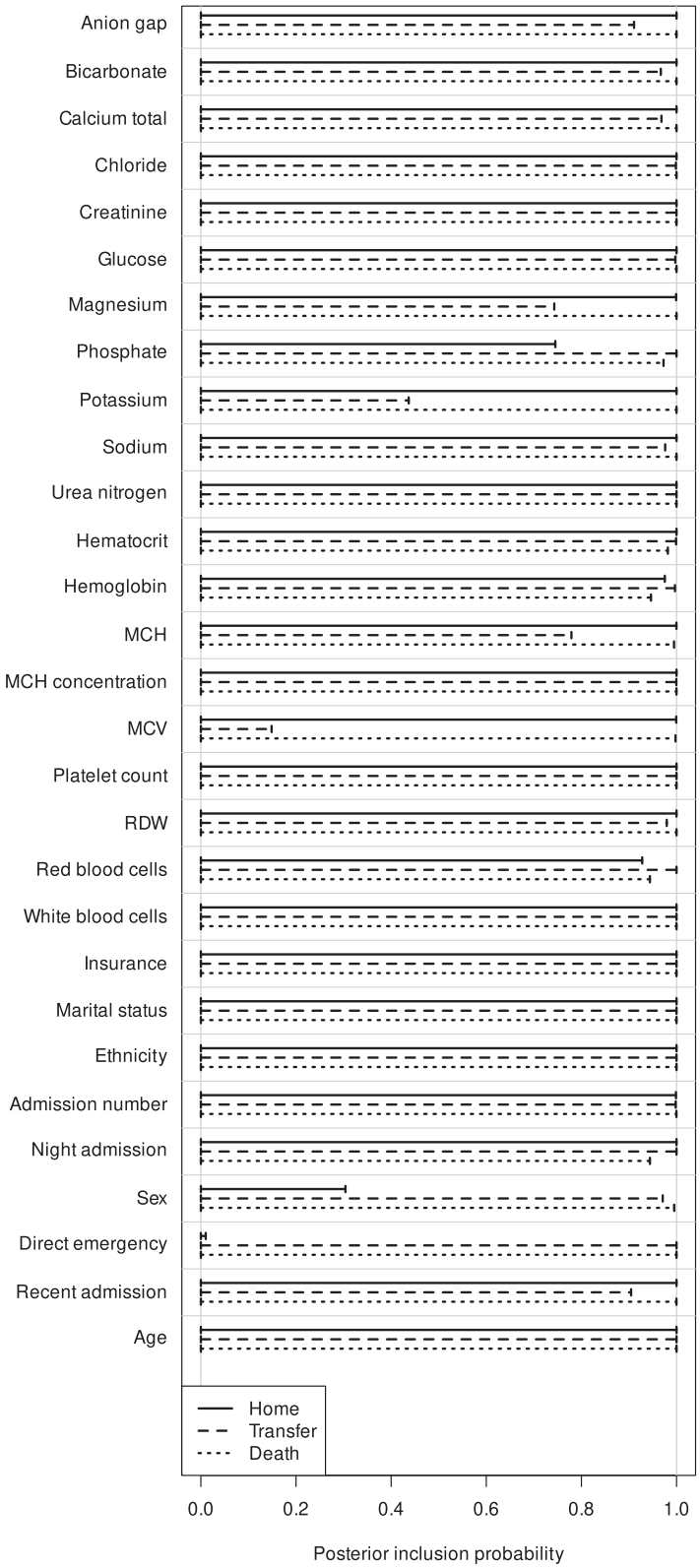

FIGURE 4 ICU data: posterior inclusion probabilities for each covariate and risk. MCH stands for mean cell hemoglobin, MCV for mean corpuscular volume, and RDW for red blood cell distribution width.

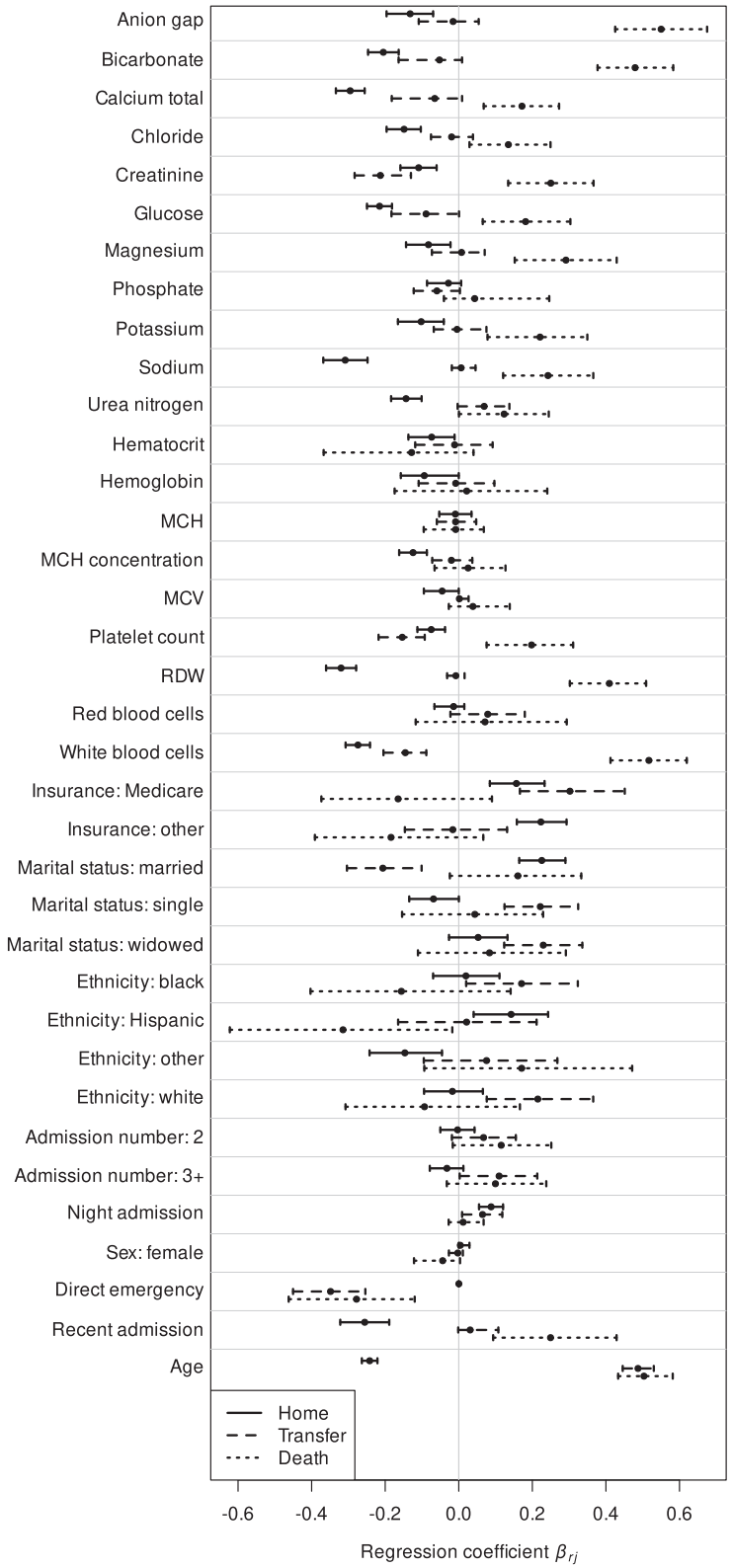

FIGURE 5 ICU data: posterior means (dot) and 95% credible intervals (lines) of the regression coefficients for each risk. The categorical predictors are coded as dummy variables as detailed in Web Appendix A. MCH stands for mean cell hemoglobin, MCV for mean corpuscular volume, and RDW for red blood cell distribution width.

The proposed model can be easily extended to accommodate more complex scenarios, for example, inclusion of recurrent event processes as outcomes (see, eg, King and Weiss, 2021), of time-varying covariates, or semi-competing risk structure. In this work, we employ the multinomial logit model, which is a popular choice for the analysis of discrete competing risks. It closely relates to multinomial logistic regression and offers computational advantages. Nevertheless, the multivariate Bernoulli detector can be used with other likelihoods, such as multinomial probit models or multiple time series. Finally, we note that we could apply the same computational strategy even for change point models in continuous time by restricting the split points to the locations of the events.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices and Figures referenced in Sections 2, 3, and 4, and the code to implement the model are available with this paper at the Biometrics website on Oxford Academic. The code is also available at https://github.com/willemvandenboom/mvb-detector.

FUNDING

This work is supported by the Singapore Ministry of Health's National Medical Research Council under its Open Fund—Young Individual Research Grant [OFYIRG19nov-0010]. A.G. has been partially supported by MUR—Prin 2022—grant no. 2022CLTYP4, funded by the European Union—Next Generation EU.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data that support the findings in this paper are available from PhysioNet. Restrictions apply to the availability of these data, which were used under license in this paper. Data are available at physionet.org with the permission of PhysioNet.

REFERENCES

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13, 61–98.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes. Springer Series in Statistics*. New York, NY: Springer.

Andersen, P. K., Geskus, R. B., de Witte, T. and Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology*, 41, 861–870.

Bhadra, A., Datta, J., Polson, N. G. and Willard, B. (2019). Lasso meets horseshoe: a survey. *Statistical Science*, 34, 405–427.

Creswell, R., Lambert, B., Lei, C. L., Robinson, M. and Gavaghan, D. (2020). Using flexible noise models to avoid noise model misspecification in inference of differential equation time series models, arXiv, arXiv:2011.04854v1.

Dahl, D. B. (2005). Sequentially-allocated merge-split sampler for conjugate and nonconjugate Dirichlet process mixture models. *Technical Report*. Department of Statistics, Texas A&M University.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42, 204–223.

Dobigeon, N., Tourneret, J.-Y. and Davy, M. (2007). Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach. *IEEE Transactions on Signal Processing*, 55, 1251–1263.

Fahrmeir, L. and Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association*, 91, 1584–1594.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, 51, 3509–3528.

Frühwirth-Schnatter, S. and Frühwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In: *Statistical Modelling and Regression Structures* (eds, Kneib, T. and Tutz, G.), 111–132, Berlin: Springer.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 40, 63–79.

Goodman, M. S., Li, Y. and Tiwari, R. C. (2011). Detecting multiple change points in piecewise constant hazard functions. *Journal of Applied Statistics*, 38, 2523–2532.

Harlé, F., Chatelain, F., Gouy-Pailler, C. and Achard, S. (2016). Bayesian model for multiple change-points detection in multivariate time series. *IEEE Transactions on Signal Processing*, 64, 4351–4362.

Held, L. and Holmes, C. C. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1, 145–168.

Heyard, R., Timsit, J.-F., Essaied, W. I. and Held, L. (2019). Dynamic clinical prediction models for discrete time-to-event data with competing risks—a case study on the OUTCOMEREA database. *Biometrical Journal*, 61, 514–534.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73, 387–396.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, 1, 255–273.

Hougaard, P., Myglegaard, P. and Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*, 50, 1178–1188.

Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S. et al. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10, 1.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. Hoboken, NJ: John Wiley & Sons.

King, A. J. (2017). brea: Bayesian recurrent event analysis. R package version 0.2.0. https://CRAN.R-project.org/package=brea [Accessed 1 June 2023].

King, A. J. and Weiss, R. E. (2021). A general semiparametric Bayesian discrete-time recurrent events model. *Biostatistics*, 22, 266–282.

Kozumi, H. (2000). Bayesian analysis of discrete survival data with a hidden Markov chain. *Biometrics*, 56, 1002–1006.

Landau, V. A. and Zachmann, L. J. (2019). tvgeom: an R package for the time-varying geometric distribution. R package version 1.0.1. https://CRAN.R-project.org/package=tvgeom [Accessed 1 June 2023].

Lee, M., Feuer, E. J. and Fine, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics*, 74, 1468–1481.

Linderman, S., Johnson, M. J. and Adams, R. P. (2015). Dependent multinomial models made easy: stick-breaking with the Pólya–Gamma augmentation. In: *Advances in Neural Information Processing Systems* (eds, Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R.), vol. 28. Redhook, NY: Curran Associates, Inc.

Luo, S., Kong, X. and Nie, T. (2016). Spline based survival model for credit risk modeling. *European Journal of Operational Research*, 253, 869–879.

Mandelbaum, M., Hlynka, M. and Brill, P. H. (2007). Nonhomogeneous geometric distributions with relations to birth and death processes. *TOP*, 15, 281–296.

Martínez, A. F. and Mena, R. H. (2014). On a nonparametric change point detection model in Markovian regimes. *Bayesian Analysis*, 9, 823–858.

Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, 38, 463–468.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics, Economic Theory and Mathematical Economics* (ed. Zarembka, P.), 105–142, New York, NY: Academic Press.

Meir, T. and Gorfine, M. (2023). Discrete-time competing-risks regression with or without penalization, arXiv, arXiv:2303.01186v2.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83, 1023–1032.

Möst, S., Pößnecker, W. and Tutz, G. (2016). Variable selection for discrete competing risks models. *Quality & Quantity*, 50, 1589–1610.

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108, 1339–1349.

Polson, N. G. and Sokolov, V. (2019). Bayesian regularization: from Tikhonov to horseshoe. *WIREs Computational Statistics*, 11, e1463.

Samsonov, S., Lagutin, E., Gabrié, M., Durmus, A., Naumov, A. and Moulines, E. (2022). Local-global MCMC kernels: the best of both worlds, arXiv, arXiv:2111.02702v3.

Schmid, M. and Berger, M. (2020). Competing risks analysis for discrete time-to-event data. *WIREs Computational Statistics*, 13, e1529.

Teugels, J. L. (1990). Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis*, 32, 256–268.

Tutz, G. (1995). Competing risks models in discrete time with nominal or ordinal categories of response. *Quality and Quantity*, 29, 405–420.

Tutz, G. and Schmid, M. (2016). *Modeling Discrete Time-to-Event Data. Springer Series in Statistics*. Switzerland: Springer.

Vallejos, C. A. and Steel, M. F. J. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180, 613–631.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S. Statistics and Computing*. 4th edition. New York, NY: Springer.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage–Dickey density ratio. *Journal of the American Statistical Association*, 90, 614–618.

Wang, C.-P. and Ghosh, M. (2007). Change-point diagnostics in competing risks models: two posterior predictive *p*-value approaches. *TEST*, 16, 145–171.

Received: August 21, 2023; Revised: May 23, 2024; Accepted: July 24, 2024