Title
A general semiparametric Bayesian discrete-time recurrent events model

Journal
Biostatistics, 22(2)

ISSN
1465-4644

Authors
King, Adam J
Weiss, Robert E

Publication Date
2021-04-10

DOI
10.1093/biostatistics/kxz029

Peer reviewed

Biostatistics (2021) **22**, 2, *pp.* 266–282
doi:10.1093/biostatistics/kxz029
Advance Access publication on August 2, 2019

A general semiparametric Bayesian discrete-time recurrent events model

ADAM J. KING*

Department of Mathematics & Statistics, California State Polytechnic University, Pomona, 3801 West Temple Ave., Pomona, CA 91768, USA

king@cpp.edu

ROBERT E. WEISS

Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, 650 Charles E. Young Dr. South, Los Angeles, CA 90095-1772, USA

SUMMARY

Event time variables are often recorded in a discrete fashion, especially in the case of patient-reported outcomes. This work is motivated by a study of illicit drug users, in which time to drug use cessation has been recorded as a number of whole months. Existing approaches for handling such discrete data include treating the survival times as continuous (with adjustments for inevitable tied outcomes), or using discrete models that omit important features like random effects. We provide a general Bayesian discrete-time proportional hazards model, incorporating a number of features popular in continuous-time models such as competing risks and frailties. Our model also provides flexible baseline hazards for time effects, as well as generalized additive models style semiparametric incorporation of other time-varying covariates. Our specific modeling choices enable efficient Markov chain Monte Carlo inference algorithms, which we provide to the user in the form of a freely available R package called `brea`. We demonstrate that our model performs better on our motivating substance abuse application than existing approaches. We also present a reproducible application of the `brea` software to a freely available data set from a clinical trial of anesthesia administration methods.

Keywords: Competing risks; Cox model; Discrete time; Generalized additive models; Recurrent events; Semiparametric models; Software; Substance abuse.

1. INTRODUCTION AND MOTIVATION

A common feature of survival data is the discrete recording of the timings of events (Allison, 1982). This may occur when time is truly discrete, as would be the case when analyzing the number of treatment administrations before some desired effect is reached. Alternatively, continuous time may be partitioned into non-overlapping intervals (often corresponding to whole days, weeks, or months), and only the interval in which an event occurs is recorded. This special case of interval censoring, called *grouped time*, is responsible for the occurrence of tied observations in nominally continuous survival data.

*To whom correspondence should be addressed.

Our work is motivated by examining the time until cocaine use cessation among subjects in the Treatment Utilization and Effectiveness Project (TUE), an observational study of illicit drug users conducted from 1992 to 1997 in Los Angeles County, California (Hser *and others*, 1998). The data we utilize come from a subsample who completed a Natural History Interview (NHI), which retrospectively recorded all lifetime episodes of cocaine use, incarceration, and several other traits related to drug use. Episodes are recorded as contiguous segments of whole calendar months. For example, if a subject began using cocaine in January of 1995 and ceased using in June of the following year, that would be recorded as one episode of cocaine use covering the 18 discrete-time points corresponding to all 12 months of 1995 and the first 6 months of 1996. The NHI interview protocol uses a timeline of life events (arrests, hospitalizations, marriages, etc.) constructed from official records to assist subjects in recalling episodes occurring many years prior to the interview.

In this application, it is critical to distinguish between two competing risks for drug use cessation. A study subject may cease use due to arrest and subsequent incarceration; we call this risk *incarceration* and label it risk $r = 1$. Alternatively, a subject may voluntarily cease use, which we refer to as *stop-use* for short and label as risk $r = 2$. In any discrete-time survival scenario, the discrete hazard rate $\lambda(t)$ is defined to be the probability of event occurrence at time point t (where conventionally $t = 1, 2, 3, \ldots$) given that the event had not already occurred earlier. In applications with two competing risks such as ours, we have two cause-specific discrete hazards $\lambda^{(1)}(t)$ and $\lambda^{(2)}(t)$, representing respectively the probabilities of incarceration and stop-use given that we are t months into an episode of cocaine use. Our goal is to model these hazard rates as functions of a variety of possibly time-varying covariates.

The most common way to relate probabilities of mutually exclusive outcomes to linear predictors encapsulating covariate effects is via the multinomial logit link function. Letting $\lambda^{(0)}(t) = 1 - \lambda^{(1)}(t) - \lambda^{(2)}(t)$ denotes the probability that neither competing risk occurs at t, we write $\eta^{(r)}(t) = \log(\lambda^{(r)}(t)/\lambda^{(0)}(t))$ for $r = 1, 2$. Then for example, the linear predictor $\eta^{(1)}(t)$ tells us how the incarceration rate varies on a log-odds scale. We would like each of these two linear predictors $\eta^{(r)}(t)$ to include random effects α to model association among repeated observations from the same subject, fixed effects τ to incorporate categorical covariates, and arbitrary smooth functions $f(\cdot)$ to model the nonlinear effects of quantitative covariates. These smooth functions offer a semiparametric extension of standard linear modeling, analogous to their use in generalized additive models (GAMs; Wood, 2011). We will adopt a Bayesian approach for modeling the presumed smooth $f(\cdot)$ using Gaussian Markov random fields (GMRF).

To illustrate, let i denotes the study subject and j denotes the episode number of the current episode of cocaine use from subject i (i.e., the number of previous episodes of drug use plus 1). Temporarily suppressing the competing risk superscript (r), we have the following additive formulation for each of our linear predictors

$$\eta(t) = \beta_0 + \alpha(i) + \tau_1(\text{sex}_i) + \tau_2(\text{race}_i) + \tau_3(j) + f_1(t) + f_2(\text{age}_{ijt}) + f_3(\text{calendar time}_{ijt}), \qquad (1.1)$$

where β_0 is an intercept, $\alpha(i)$ is a subject-specific random effect, the three τ terms model the effects of three categorical predictors, and the three functions $f(\cdot)$ model the effects of the quantitative variables discrete time t, current age, and current calendar time (e.g., June of 1995). The age and calendar time variables in (1.1) are indexed by i, j, and t, since these covariates vary between and during episodes of cocaine use (e.g., the age of a teenager advances nontrivially over the course of an 18-month drug use episode). We assume that each of the two linear predictors $\eta^{(1)}(t)$ and $\eta^{(2)}(t)$ utilize the same set of covariates specified in (1.1), but the corresponding effect parameters will differ. For instance, following the convention of superscripts denoting competing risk, the random effect $\alpha^{(1)}(i)$ captures subject i's tendency to become incarcerated while using cocaine, while $\alpha^{(2)}(i)$ captures subject i's tendency to voluntarily cease drug use. In addition, we allow these two random effects to be correlated within subjects.

We demonstrate in Section 5 of the article that all of these features we have introduced are essential for modeling the TUE data. Furthermore, many if not all of these features will prove necessary in a wide variety of discrete survival applications, such as examining the time to reemployment following job loss, where finding a job and leaving the labor force would be competing risks (Narendranathan and Stewart, 1993). However, up to this point no other models and software exist that can handle such complex discrete survival scenarios.

The rest of this article is organized as follows: In Section 2, we describe the GMRF approach to modeling the arbitrary smooth functions $f(\cdot)$ in our linear predictors, and we review relevant work on Markov chain Monte Carlo (MCMC) inference for models incorporating GMRF components. In Section 3, we detail a general version of the model described in the introduction, and Section 4 introduces algorithms for efficiently obtaining inferences from this general model. In Section 5, we compare our model to alternative approaches for analyzing the TUE data, and then provide practical inferences regarding substance abuse behaviors. We end the main article with discussion of future additions to our model in Section 6. Our models and algorithms are implemented in the R package `brea` (King, 2017), and specifications of this software's interface along with a reproducible application appear in the supplementary material available at *Biostatistics* online.

2. BAYESIAN SMOOTHING USING GMRF PRIORS

We model the arbitrary smooth functions $f(\cdot)$ in (1.1) as piecewise-constant functions, and we need a convenient way of introducing prior correlation among neighboring levels of these step functions. One approach, introduced by Besag (1974) and subsequent coauthors to model spatial correlation in geostatistical and imaging applications (see, e.g., Besag *and others*, 1991), specifies the distribution of a parameter vector $\theta = (\theta_1, \ldots, \theta_n)$ implicitly via its full conditional distributions $p(\theta_t|\theta_{-t})$, where $\theta_{-t} = \{\theta_s \mid s \neq t\}$. Such specifications are called conditional autoregressions or Markov random fields (Besag and Kooperberg, 1995). The Markov name refers to the feature that the distributions $p(\theta_t|\theta_{-t})$ often only depend on a small number of elements in θ_{-t}, which correspond to nodes s which are spatially or temporally close to node t. Typically, the full conditionals $p(\theta_t|\theta_{-t})$ are presumed Gaussian, in which case the resulting joint distribution is multivariate normal and called a GMRF. The distribution can then be specified using the inverse covariance matrix, or precision matrix, Q of the multivariate normal via $p(\theta|Q) \propto \exp(-\frac{1}{2}\theta' Q\theta)$. The Markov property of the full conditionals ensures that the matrix Q is sparse, which leads to fast algorithms for sampling θ (Rue, 2001; Rue and Held, 2005).

The high posterior correlation among the θ_t induced by the GMRF prior makes block MCMC updating of the θ_t necessary to avoid slow MCMC mixing (Carter and Kohn, 1994). In the case of a generalized linear model with a non-normal response variable, the full conditional of a GMRF θ whose elements appear in the linear predictors of the model will generally not be possible to sample from directly. Several authors have attempted to devise full conditional approximating (FCA) block Metropolis–Hastings (M–H) proposals for θ or its subvectors, but achieved very low acceptance rates for their proposed parameter updates (Shephard and Pitt, 1997; Gamerman, 1998). However, Rue (2001) and Knorr-Held and Rue (2002) constructed FCA block M–H proposals of θ for Poisson disease mapping problems by combining quadratic approximations of the exponential terms of the Poisson log-likelihood with the GMRF prior densities to produce a GMRF approximate full conditional density. They reported high acceptance rates of 20–80%, even when updating blocks of hundreds of parameters. Our M–H algorithms in Section 4 employ a similar strategy of log-likelihood approximation, but for a very different observation model with much more complex linear predictor specifications.

3. BAYESIAN DISCRETE SURVIVAL MODEL

In this section, we outline our discrete-time competing risks observation model, which is a generalization of the model presented in Section 1; here, the number of competing risks R is arbitrary, whereas in

our motivating application we had $R = 2$ event types. Our approach is a discrete-time analogue of the continuous-time cause-specific hazards approach of Prentice *and others* (1978). We then discuss the components of the linear predictors and their prior specification.

3.1. *Discrete-time competing risks observation model*

Let $i = 1, \ldots, I$ index subjects, $j = 1, \ldots, J_i$ index repeated at-risk episodes, or recurrent events, within subject i, and $t = 1, \ldots, T_{ij}$ index discrete-time points at which subject i is at risk for her jth recurrent event. The value t, often called *time-at-risk*, *duration*, or *gap time*, always resets to $t = 1$ at the beginning of each successive at-risk episode, so that T_{ij} is the discrete duration of episode j from subject i. We also let $r = 1, \ldots, R$ index the distinct competing risks under consideration, and let $\Delta_{ij} \in \{1, \ldots, R\}$ denotes the competing risk which terminates the jth at-risk episode from subject i at time $t = T_{ij}$.

The discrete-time cause-specific hazard $\lambda_{ij}^{(r)}(t)$ is the probability that the jth recurrent event from subject i occurs due to risk r at time t elapsed since the subject became at risk for the jth recurrence, given that the event had not yet occurred by time $t-1$, so $\lambda_{ij}^{(r)}(t) \equiv \Pr[T_{ij} = t, \Delta_{ij} = r \mid T_{ij} \geq t]$. For brevity, we often suppress the dependence of the hazards, linear predictors, and covariates on the indices i and j, writing $\lambda^{(r)}(t)$ for the rth cause-specific hazard.

For convenience, let $\lambda^{(0)}(t)$ denotes the probability that none of the competing risks occurs at time t given the subject was still at risk at that point in the episode, so that $\lambda^{(0)}(t) = 1 - \sum_{r=1}^{R} \lambda^{(r)}(t)$. For each time point t, the collection $\{\lambda^{(0)}(t), \ldots, \lambda^{(R)}(t)\}$ are probabilities of mutually exclusive events, representing the $R+1$ possible outcomes at time t. We relate these probabilities to R linear predictors $\{\eta^{(1)}(t), \ldots, \eta^{(R)}(t)\}$ using the multinomial logit link function,

$$\eta^{(r)}(t) \equiv \log\left(\frac{\lambda^{(r)}(t)}{\lambda^{(0)}(t)}\right), \quad r = 1, \ldots, R, \tag{3.2}$$

which allows the $\lambda^{(r)}(t)$ to assume any values subject to $0 < \lambda^{(r)}(t) < 1$ and $\sum_{r=0}^{R} \lambda^{(r)}(t) = 1$.

3.2. *Linear predictor and prior specification*

The linear predictors $\eta^{(r)}(t)$ can include covariates that vary between subjects, between at-risk episodes within subjects, and between time points within episodes. Hence, in general the value of a covariate will depend on i, j, and t, and we let $X_{m,ij}(t)$ denotes the value of covariate m ($m = 1, \ldots, M$) for subject i at time t during the jth at-risk episode. When it creates no confusion, we will suppress the dependence of the covariate values on some or all of these indices, using for example X_m to denote the mth predictor or just X to denote a generic predictor.

The way in which a covariate is incorporated into the linear predictors depends on its type. For a quantitative covariate X which we presume has a smooth relationship with the hazard of event occurrence, we use a function $f(X)$ as in (1.1). On the other hand, if X is categorical, then we use either a random effect $\alpha(X)$ or a fixed effect $\tau(X)$. In all three cases, the parameters representing the corresponding covariate effects (the levels of the function $f(\cdot)$ or the values of α or τ) are allowed to differ for each competing risk r, and when necessary we make this explicit by writing for example $f^{(r)}(X)$. We also let K_m denotes the number of parameters used to represent the effect of X_m on the hazard of each competing risk r, and we denote these parameters as $\{\beta_m^{(r)}[1], \ldots, \beta_m^{(r)}[K_m]\}$ for all three covariate types; we describe these representations in detail in the following three subsections, where we make use of the notational shorthand $\beta_k \equiv \beta_m^{(r)}[k]$.

Altogether, we have the following linear predictor formulation

$$\eta^{(r)}(t) \equiv \beta_0^{(r)} + \sum_{\text{random } m} \alpha_m^{(r)}(X_m) + \sum_{\text{fixed } m} \tau_m^{(r)}(X_m) + \sum_{\text{quantitative } m} f_m^{(r)}(X_m). \tag{3.3}$$

We address the lack of identifiability in the overall level of the terms in (3.3) by setting $\sum_{k=1}^{K} \beta_k = 0$, which is similar to the approach conventionally taken in GAMs. We assign flat priors for the intercepts, $p(\beta_0^{(r)}) \propto 1$. With our identifiability constraints, the intercepts will be identified as long as the data contain at least one observed event for each risk r as well as at least one subject observed at one time point where no event occurs. In these circumstances, which should be almost always true in practice, the posterior distribution will be proper.

3.2.1. *Quantitative covariates* For quantitative X, we use piecewise-constant functions $f(X)$ with GMRF priors on their levels to approximate arbitrary smooth functions. We may then, for example, incorporate duration dependence by setting $X(t) = t$, in which case $f^{(r)}(X)$ corresponds to a cause-specific baseline hazard for risk r on the log-odds scale.

Let $c_0 < c_1 < \cdots < c_K$ be a fixed collection of change points for the effect of covariate X. We assume $f(X)$ is constant and equal to the value β_k for all X in the kth interval defined by these points, i.e. $f(X) \equiv \beta_k$ for all $X \in [c_{k-1}, c_k)$. We then place a GMRF prior on $\beta_1, \ldots, \beta_K$,

$$\beta_k \mid \beta_1, \ldots, \beta_{k-1} \sim N(\beta_{k-1}, \sigma^2), \quad k = 2, \ldots, K, \tag{3.4}$$

with a flat prior for the first parameter β_1. We have defined this GMRF as an autoregressive process in a forward direction, with distributions conditional on parameters with lower index values k. However, this prior specification is mathematically equivalent to a full conditional prior specification of $\beta_k \mid \beta_{-k} \sim N(\frac{\beta_{k-1}+\beta_{k+1}}{2}, \frac{\sigma^2}{2})$, where $\beta_{-k} = \{\beta_1, \ldots, \beta_{k-1}, \beta_{k+1}, \ldots, \beta_K\}$. This is an improper, or *intrinsic*, GMRF, as it is invariant to level shifts added to all the β_k (Besag and Kooperberg, 1995). The posterior will be proper as long as there is at least one observed event for each competing risk in the data set as well as at least one person-time observation where no competing risks occur. Denoting by $IG(a, b)$ the inverse gamma distribution with shape a and scale b, we complete the specification by assigning priors $\sigma^2 \sim IG(a, b)$, so that $p(\sigma^2|a, b) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp\left(\frac{-b}{\sigma^2}\right)$. In most applications, $K = 50$ or 100 evenly spaced change points are sufficient to approximate a smooth function; more than 100 change points results in additional computational overhead with little benefit.

3.2.2. *Categorical covariates (fixed effects)* The second type of covariate is a categorical variable with a small or moderate number of categories. Suppose X may only take on values in the set $\{x_1, \ldots, x_K\}$. We let β_k denotes the value taken by the fixed effects term τ on the kth possible value of X, i.e., $\tau(x_k) = \beta_k$ for $k = 1, \ldots, K$.

We require a prior for the column vector of parameters $\beta \equiv (\beta_1, \ldots, \beta_K)'$, with three properties: exchangeability of the components β_k, borrowing of strength among the β_k, and invariance to level shifts. Letting I_K denotes the $K \times K$ identity matrix and J_K the $K \times K$ matrix of 1's, define the singular prior precision matrix $Q \equiv \phi \frac{K}{K-1}\left(I_K - \frac{1}{K}J_K\right)$, which then defines an improper prior for β as $p(\beta|Q) \propto \exp(-\frac{1}{2}\beta' Q \beta)$. As with the continuous covariate case, the posterior is proper if we observe at least one event for each risk as well as one person-time observation where no event occurs. The *a priori* specified conditional prior precision parameter ϕ controls the degree of borrowing of strength in the prior, which can be seen from the induced full conditional distributions $\beta_k \mid \beta_{-k} \sim N\left(\frac{1}{K-1}\sum_{l \neq k} \beta_l, \phi^{-1}\right)$.

3.2.3. *Categorical covariates (random effects)* Finally, we consider categorical variables with large numbers of categories, in which case it is desirable to allow the data to inform the between-category variability. For example, we may use the variable X to model associations among recurrent events within subjects by setting $X = i$, where i indexes study subjects. The random effects $\alpha(X)$ are then often called *shared frailties* (Hougaard, 2000, Chap. 9).

As in the fixed effects case, we assume X takes values in $\{x_1, \ldots, x_K\}$ and write $\alpha^{(r)}(x_k) = \beta_k^{(r)}$, where we make explicit these parameters' dependence on the competing risk r. Let $\beta_k = (\beta_k^{(1)}, \ldots, \beta_k^{(R)})'$ denotes the vector of the kth category's random effects for all competing risks r; we allow these effects to be correlated within categories by assigning a multivariate normal distribution $\beta_k \mid \Sigma \overset{iid}{\sim} N_R(0, \Sigma)$ for $k = 1, \ldots, K$. Letting $IW(\nu, S)$ denotes the inverse Wishart distribution with ν degrees of freedom and scale S, we assign priors $\Sigma \sim IW(\nu, S)$, so that $p(\Sigma|\nu, S) \propto |\Sigma|^{-(\nu+R+1)/2} \exp\left(-\frac{1}{2}\text{trace}\left(S\Sigma^{-1}\right)\right)$.

4. Computing

We now derive useful formulations of the likelihood for the model of the previous section, and then use approximations of the log-likelihood to construct block M–H algorithms for obtaining posterior inferences from this model.

4.1. *Data structures and likelihood evaluation*

4.1.1. *Likelihood* First we establish some notation. Let the pair (t_{ij}, δ_{ij}) be the data observed for the jth recurrent event from subject i, where $\delta_{ij} = r \geq 1$ means risk r was observed at time $t = t_{ij}$ and $\delta_{ij} = 0$ denotes right censoring at t_{ij}. Our notation assumes that right censoring occurs at the end of the corresponding time interval; censoring at the beginning requires recoding. For $r = 0, \ldots, R$, set $\delta_{ij}^{(r)} = 1$ if $\delta_{ij} = r$ and $\delta_{ij}^{(r)} = 0$ otherwise, and let $y_{ij}^{(r)}(t)$ be person-time level event indicators, where $y_{ij}^{(r)}(t) = 1$ if $t = t_{ij}$ and $r = \delta_{ij}$, with $y_{ij}^{(r)}(t) = 0$ otherwise.

Letting θ denotes all model parameters and $D = \{(t_{ij}, \delta_{ij})\}$ denotes the data, the likelihood is

$$\begin{aligned} L(\theta|D) &= \prod_{i=1}^{I}\prod_{j=1}^{J_i}\left(\prod_{t=1}^{t_{ij}-1}\lambda_{ij}^{(0)}(t)\right)\left(\prod_{r=0}^{R}\left(\lambda_{ij}^{(r)}(t_{ij})\right)^{\delta_{ij}^{(r)}}\right) \\ &= \prod_{i=1}^{I}\prod_{j=1}^{J_i}\left(\prod_{t=1}^{t_{ij}}\lambda_{ij}^{(0)}(t)\right)\left(\prod_{r=1}^{R}\left(\exp\left(\eta_{ij}^{(r)}(t_{ij})\right)\right)^{\delta_{ij}^{(r)}}\right) \\ &= \prod_{i=1}^{I}\prod_{j=1}^{J_i}\prod_{t=1}^{t_{ij}}\left[\left(1+\sum_{r=1}^{R}\exp\left(\eta_{ij}^{(r)}(t)\right)\right)^{-1}\prod_{r=1}^{R}\exp\left(y_{ij}^{(r)}(t)\cdot\eta_{ij}^{(r)}(t)\right)\right], \end{aligned} \tag{4.5}$$

where in the second line we have used (3.2) to replace $\lambda_{ij}^{(r)}(t_{ij})$ with $\lambda_{ij}^{(0)}(t_{ij})\exp(\eta_{ij}^{(r)}(t_{ij}))$ and moved the term $\lambda_{ij}^{(0)}(t_{ij})$ into the first product. In the third line, we have replaced the indicators $\delta_{ij}^{(r)}$ with $y_{ij}^{(r)}(t_{ij})$ and included the exponential terms for $t < t_{ij}$, which is possible since for those t we have $y_{ij}^{(r)}(t) = 0$. In addition, in the third line we have used the substitution $\lambda_{ij}^{(0)}(t) = (1 + \sum_{r=1}^{R}\exp(\eta_{ij}^{(r)}(t)))^{-1}$ implied by the link function (3.2).

Unlike the continuous-time cause-specific hazards model, the discrete-time model likelihood (4.5) does not factor into separate components for each cause-specific hazard. Thus separate hazard estimation is

not possible with the discrete model, which makes sense as otherwise this could result in incompatible estimates of the $\lambda_{ij}^{(r)}(t)$, which sum over r to greater than 1.

4.1.2. *Person-time indexing* The likelihood (4.5) is a product over all person-time observations of the expression in large brackets. Currently, we are using three indices i, j, and t to specify a discrete person-time observation. For convenience, we now pass to a single index $n = 1, \ldots, N$ for all person-time observations, where $N = \sum_{i=1}^{I} \sum_{j=1}^{J_i} t_{ij}$ is the total number of person-time observations. The order in which n indexes these observations can be arbitrary, but for optimal computer memory access performance, we recommend *row major order*, where $n = 1$ corresponds to $(i,j,t) = (1,1,1)$, $n = 2$ corresponds to $(i,j,t) = (1,1,2)$,..., $n = t_{11} + 1$ corresponds to $(i,j,k) = (1,2,1)$, and so on. We now replace the ij subscript and function argument t with the single subscript n in all our notation; e.g., $\lambda_{ij}^{(r)}(t)$ and $X_{m,ij}(t)$ become simply $\lambda_n^{(r)}$ and $X_{m,n}$. Thus the likelihood (4.5) may be rewritten

$$L(\theta|D) = \prod_{n=1}^{N} \left[\left(1 + \sum_{r=1}^{R} \exp\left(\eta_n^{(r)}\right) \right)^{-1} \prod_{r=1}^{R} \exp\left(y_n^{(r)} \cdot \eta_n^{(r)}\right) \right]. \tag{4.6}$$

4.1.3. *Discretized predictors and multinomial aggregation* The effect of the mth covariate X_m on the hazard of risk r is represented in the linear predictor by a smooth function $f^{(r)}(\cdot)$, a fixed effects term $\tau^{(r)}$, or a random effects term $\alpha^{(r)}$. Each of these terms, in turn, is represented in our model by the finite collection of possible values of the term, denoted $\{\beta_m^{(r)}[1], \ldots, \beta_m^{(r)}[K_m]\}$, regardless of the covariate type. For each entry $x = X_{m,n}$ in our set of covariate values, let $k_{m,n}$ denotes the integer k such that $\beta_m^{(r)}[k]$ equals $f^{(r)}(x)$, $\tau^{(r)}(x)$, or $\alpha^{(r)}(x)$, depending on the covariate type. Using these *discretized predictors* $k_{m,n}$ and the person-time index n, the linear predictor specification (3.3) may be rewritten

$$\eta_n^{(r)} \equiv \beta_0^{(r)} + \sum_{m=1}^{M} \beta_m^{(r)}[k_{m,n}]. \tag{4.7}$$

Depending on the number and types of covariates, as well as the discretizations chosen for any continuous covariates, for certain subsets $\{n_1, \ldots, n_b\}$ of person-time observations we will have $k_{m,n_1} = \cdots = k_{m,n_b}$ for all m. These observations will then always have the same values for the corresponding linear predictors (4.7). In this case, we can store only one copy of the covariate vector $(k_{1,n_1}, \ldots, k_{M,n_1})$ along with the size b of the subset and the total number of observed events $y^{(r)} \equiv y_{n_1}^{(r)} + \cdots + y_{n_b}^{(r)}$ for each competing risk r. The likelihood contribution of these b person-time observations can then be replaced by the single multinomial likelihood term

$$\left(1 + \sum_{r=1}^{R} \exp\left(\eta_{n_1}^{(r)}\right) \right)^{-b} \prod_{r=1}^{R} \exp\left(y^{(r)} \cdot \eta_{n_1}^{(r)}\right), \tag{4.8}$$

which may speed evaluation of the likelihood (4.6) substantially; we call this replacement *multinomial aggregation*. For example, the model described in Section 5.1 has $M = 7$ covariates, including subject-specific random effects and three quantitative time scales. Despite using discretizations of the time variables with around 50 change points each, multinomial aggregation in this case reduced the number of terms in the likelihood by 17%, yielding a substantial reduction in computation time per MCMC iteration. We recommend performing this aggregation in practice, and this feature is available in the `brea` R package via the optional function argument `n` to `brea_mcmc()`, as described in Appendix C of the

supplementary material available at *Biostatistics* online. However, for sake of simplicity we ignore this feature's implementation in subsequent algorithm descriptions.

Additional data structures for enabling fast software implementation of the likelihood evaluation are given in Appendix A of the supplementary material available at *Biostatistics* online.

4.2. *Block random walk Metropolis step*

We propose two M–H schemes for obtaining posterior inferences. The first is a block random walk Metropolis (RWM) algorithm which uses multivariate normal proposals centered at the current parameter values. We describe this method here, and then provide a block FCA M–H algorithm in the following subsection.

Given a contiguous subset $k_1 : k_2 = \{k_1, k_1 + 1, \ldots, k_2\}$ of the discretized values $\{1, \ldots, K\}$ of the predictor X and its associated block of parameters $B \equiv \{\beta^{(r)}[k] \mid k_1 \leq k \leq k_2,\ 1 \leq r \leq R\}$, we need to find a covariance matrix Σ_B which ensures that the RWM proposal

$$B^* \sim N(B, \Sigma_B) \tag{4.9}$$

provides efficient MCMC mixing. Gelman *and others* (1996) show that we should choose Σ_B proportional to the covariance Σ_C of the full conditional distribution $p(B \mid \theta \backslash B, D)$; specifically, we should set $\Sigma_B \equiv (2.4/\sqrt{|B|})^2 \Sigma_C$, where $|B| = R \cdot (k_2 - k_1 + 1)$ is the dimension of the block B. Because Σ_C is difficult to calculate directly, we approximate the corresponding precision matrix $Q_C \equiv \Sigma_C^{-1}$ using the Hession matrix of the log full conditional density with respect to the parameters in B. The second derivatives of the log conditional prior density $p(B \mid \theta \backslash B)$ are available immediately, since for all three covariate classes described in Section 3.2, the conditional prior is normal with precision Q_P which is a simple function of conditioned upon parameters.

We approximate the second derivatives of the logarithm $l(\theta|D)$ of the likelihood (4.6) as follows. Letting $\beta_1 \equiv \beta^{(r_1)}[\tilde{k}_1]$ and $\beta_2 \equiv \beta^{(r_2)}[\tilde{k}_2]$ denote two parameters in B, we first have

$$\frac{\partial^2 l(\theta|D)}{\partial \beta_1^2} = \sum_{n \in n_{\tilde{k}_1}} -\lambda_n^{(r_1)}(1 - \lambda_n^{(r_1)}) \approx \sum_{n \in n_{\tilde{k}_1}} -\lambda_n^{(r_1)} \approx -y^{(r_1)}, \tag{4.10}$$

where $n_{\tilde{k}_1}$ denotes the collection of person-time observations whose hazards depend on $\beta_1 = \beta^{(r_1)}[\tilde{k}_1]$ and $y^{(r_1)}$ denotes the number of observed events for competing risk r_1 whose hazards depend on the parameter $\beta_1 = \beta^{(r_1)}[\tilde{k}_1]$. The first approximate equality in (4.10) follows from the fact that the hazards $\lambda_n^{(r)}$ are generally small probabilities, and the second is due to the fact that the aggregate hazard $\lambda^{(r)} \equiv \sum_{n \in n_k} \lambda_n^{(r)}$ for a collection n_k of person-time observations should be roughly equal to the total number of events $y^{(r)} \equiv \sum_{n \in n_k} y_n^{(r)}$ observed among that collection. For the mixed partial derivatives, when $\tilde{k}_1 = \tilde{k}_2$ and $r_1 \neq r_2$, we have

$$\frac{\partial^2 l(\theta|D)}{\partial \beta_1 \partial \beta_2} = \sum_{n \in n_{\tilde{k}_1}} \lambda_n^{(r_1)} \lambda_n^{(r_2)} \approx 0, \tag{4.11}$$

because the product of the hazards $\lambda_n^{(r_1)} \lambda_n^{(r_2)}$ is small relative to the sizes of the terms in (4.10). Finally, the mixed partials are identically 0 when $\tilde{k}_1 \neq \tilde{k}_2$, since $\beta^{(r_1)}[\tilde{k}_1]$ and $\beta^{(r_2)}[\tilde{k}_2]$ have no person-time observations in common.

We combine these log-likelihood approximations with the conditional prior precision matrix Q_P of the block B by adding the numbers of observed events $y^{(r)}$ along the diagonal of Q_P to produce an approximation

$\hat{Q}_C$ to the full conditional precision Q_C. Setting $\Sigma_B \equiv (2.4/\sqrt{|B|})^2 \hat{Q}_C^{-1}$, we then sample from the proposal (4.9) using an efficient method which takes advantage of the sparsity of $\hat{Q}_C$ (Rue and Held, 2005, Chap. 2). Importantly, computing $\hat{Q}_C$ and sampling from (4.9) does not require any computation which scales with N, so computation time is still dominated by log-likelihood evaluation. As we show in Appendix B of the supplementary material available at *Biostatistics* online, these proposals result in nearly optimal acceptance rates with no tuning or adaptation required and with little increase in computation time per iteration compared to fixed proposals.

4.3. *Block FCA M–H step*

It is critical for fast MCMC mixing to update groups of highly correlated parameters within the same block. When using GMRF priors with fine discretizations of quantitative covariates, these groups of highly correlated parameters can be large, necessitating large blocks B. However, the maximum efficiency of RWM algorithms is roughly $0.3/|B|$, and thus declines with increasing block size $|B|$ (Gelman *and others*, 1996). Hence, as $|B|$ increases, the advantages of block updating of highly correlated parameters are often negated by the RWM block size efficiency penalty, leading to poor performance with MCMC efficiencies on the order of 1%. For these circumstances, we have developed block M–H proposals which seek to approximate the entire shape of the full conditional distributions, not just their dispersion as with RWM. These FCA M–H steps should be nearly as efficient as exact draws from the full conditionals.

Our strategy for creating such proposals is to combine the Gaussian conditional prior with a multivariate second-order Taylor series approximation of the log-likelihood. In particular, we expand $l(B|\theta\backslash B, D)$ as a function of the block parameters $B \equiv \{\beta^{(r)}[k] \mid k_1 \le k \le k_2,\ 1 \le r \le R\}$ about their current values in the chain $\tilde{B} \equiv \{\tilde{\beta}^{(r)}[k] \mid k_1 \le k \le k_2,\ 1 \le r \le R\}$. Letting $=_B$ and $\approx_B$ denote equality and approximate equality as a function of B up to an additive constant gives

$$\begin{aligned} l(B|\theta\backslash B, D) &\approx_B \sum_{r=1}^{R} \sum_{k=k_1}^{k_2} \left[\frac{\partial l(\tilde{B})}{\partial \beta^{(r)}[k]} \left(\beta^{(r)}[k] - \tilde{\beta}^{(r)}[k] \right) + \frac{1}{2} \frac{\partial^2 l(\tilde{B})}{\partial \beta^{(r)}[k]^2} \left(\beta^{(r)}[k] - \tilde{\beta}^{(r)}[k] \right)^2 \right] \\ &=_B \sum_{r=1}^{R} \sum_{k=k_1}^{k_2} \left[\left(\frac{\partial l(\tilde{B})}{\partial \beta^{(r)}[k]} - \frac{\partial^2 l(\tilde{B})}{\partial \beta^{(r)}[k]^2} \cdot \tilde{\beta}^{(r)}[k] \right) \beta^{(r)}[k] + \frac{1}{2} \left(\frac{\partial^2 l(\tilde{B})}{\partial \beta^{(r)}[k]^2} \right) \beta^{(r)}[k]^2 \right]. \end{aligned} \quad (4.12)$$

We omit the terms of the Taylor expansion with mixed partial derivatives, since by (4.11) and the associated discussion these derivatives are all either close to zero or identically zero. For the two fixed effects covariate cases, this allows sampling of the parameters $B^{(r)} \equiv \{\beta^{(r)}[k] \mid k_1 \le k \le k_2\}$ separately for each risk r, since the elements of $B^{(r)}$ are then conditionally independent.

The first derivatives of the log-likelihood evaluated at the current parameter values $\tilde{B}$ are

$$\frac{\partial l(\tilde{B})}{\partial \beta^{(r)}[k]} = \sum_{n \in n_k} y_n^{(r)} - \tilde{\lambda}_n^{(r)} = y^{(r)} - \sum_{n \in n_k} \tilde{\lambda}_n^{(r)}, \quad (4.13)$$

where $y^{(r)}$ denotes the number of observed events of type r among observations whose hazards depend on $\beta^{(r)}[k]$. Computation of these derivatives requires evaluating the hazards $\tilde{\lambda}_n^{(r)}$ under the current values of the model parameters; the second derivatives may then either be approximated as in (4.10), or calculated exactly from the $\tilde{\lambda}_n^{(r)}$. Once the necessary derivatives are calculated, the coefficients of the linear terms $\beta^{(r)}[k]$ and quadratic terms $\beta^{(r)}[k]^2$ in (4.12) are combined with the linear and quadratic terms from the log prior density to produce a quadratic approximation to the log full conditional density. Because quadratic log density functions correspond to normal distributions, we can then sample from this density as our M–H

Table 1. *Covariate specification used for modeling the durations of repeated episodes of cocaine use*

m	Name	Covariate type	Categories K_m	Values/discretization
1	Subject ID	Random effect	408	1,…,408
2	Sex	Categorical	2	Male, Female
3	Race	Categorical	3	Black, Hispanic, White
4	Episode number	Categorical	6	1,…,5, ≥ 6
5	Episode duration	Quantitative	66	Months 1,…,60, years 6,…,10, ≥ 10 years
6	Age	Quantitative	46	10,…,55 years of age
7	Calendar time	Quantitative	35	Calendar years 1963,…,1997

proposal for the block B. The main difficulty with this approach is that it requires computing the hazards $\tilde{\lambda}_n^{(r)}$ under both the current and proposed parameter values. However, we demonstrate via simulation in Appendix B of the supplementary material available at *Biostatistics* online that the dramatic improvement in mixing more than makes up for the additional computation time per M–H update.

5. Drug use cessation application

We now use the models and algorithms developed in the previous sections to analyze the discrete durations of repeated episodes of cocaine use from the TUE data set described in the introduction. First, we describe the sample and covariates used, and then we address the insufficient smoothing of the duration effect. Next, we compare and contrast our model with the closest available existing approach, which is found in the `mgcv` R package (Wood, 2011). Finally, we provide some practical inferences regarding substance abuse behaviors.

5.1. *Data and model*

The 408 TUE subjects contributed 1527 episodes and 29 645 person-month observations of cocaine use. Hence, on average each subject had $1527/408 = 3.74$ cocaine use episodes with mean observed duration $29\,645/1527 = 19.4$ months. We consider the $R = 2$ competing causes for cocaine episode termination described in the introduction: *incarceration* ($r = 1$) and voluntary cessation, or *stop-use* for short ($r = 2$). Of the 1527 cocaine use episodes, 689 ended with incarceration, 738 ended with stop-use, and 100 were right censored at the time of interview.

We consider $M = 7$ covariates in total. Sex and race are included as categorical fixed effects and given the exchangeable priors described in Section 3.2.2, with small conditional prior precisions ϕ, making the priors relatively noninformative. Number of episodes of cocaine use up to and including the current episode is also treated as categorical, with the effect remaining the same for the sixth and subsequent episodes for a total of six categories. Current duration of the episode is included using the GMRF smoothing prior given in Section 3.2.1. We allow the duration effect to vary at the ends of each of the first 60 months of cocaine use and at the ends of each of the sixth through tenth years of use; this discretization of the duration time scale results in 66 categories. Similarly, we include current age and current absolute calendar year. Finally, we include subject-specific random effects. Table 1 summarizes these covariate specifications.

5.2. *Nonhomogeneous GMRF duration effect prior*

The first panel of Figure 1 depicts posterior medians and 95% credible intervals for the duration effect on the stop-use competing risk. We observe a roughly 4-fold decrease in the risk of voluntary use cessation

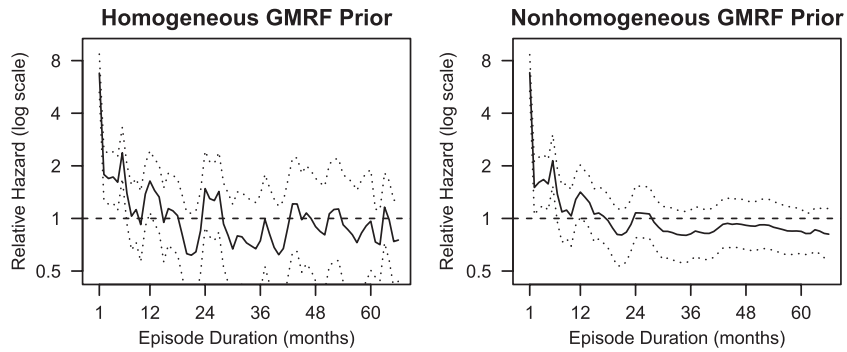

Fig. 1. Posterior medians (solid lines) and point-wise 95% credible intervals (dotted lines) of the duration effects on the stop-use competing risk, using the homogeneous GMRF prior (left panel) and nonhomogeneous prior (right panel).

following the first month of cocaine use. This large, well-estimated first-month change results in a high inferred value for the increment variance σ^2 of the GMRF prior (3.4), which in turn leads to the inadequate smoothing evident in the left panel of Figure 1.

The source of the problem is that our GMRF prior (3.4) is *homogeneous*; that is, the increment variance σ^2 is the same for all k. A more sensible prior would reflect our prior knowledge that the hazards should change more rapidly earlier in the cocaine use episodes. For example, the chances of a subject voluntarily stopping use should change more between months 1 and 2 of cocaine use than between months 50 and 51. We model this by introducing constant scaling factors s_k, $k = 1, \ldots, K_m - 1$, in front of σ^2, yielding the *nonhomogeneous* random walk GMRF prior

$$\beta_k \mid \beta_1, \ldots, \beta_{k-1} \sim N(\beta_{k-1}, s_{k-1}\sigma^2), \quad k = 2, \ldots, K_m. \tag{5.14}$$

For the duration effect, we choose the constants s_k so that $s_k \propto 1/k$ and $\sum_k s_k = K_m - 1$. With this choice, for a given value of σ^2, the aggregate variance $\sum_k s_k \sigma^2$ of the random walk prior across the entire time scale is the same as for the homogeneous prior (3.4), making it sensible to compare inferred values of σ^2 between models using the two priors.

The right panel of Figure 1 shows posterior summaries of the duration effect on the stop-use risk using the nonhomogeneous prior. At earlier durations of cocaine use, the estimates are similar to those obtained with the homogeneous prior. At later durations, more borrowing of strength occurs, leading to a smoother inferred relationship with narrower credible intervals. The posterior mean of σ^2 decreases 4-fold from 0.163 to 0.042 when passing from the homogeneous to nonhomogeneous prior, which also shows the nonhomogeneous prior induces much more smoothing across the time scale. Meanwhile, for the incarceration competing risk, we observe no dramatic change in hazard during the first month, and the two priors give similar results, with the posterior mean of σ^2 virtually unchanged when passing between the two formulations. We adopt the nonhomogeneous prior to the duration effects on both competing risks and retain the homogeneous formulations for the age and calendar time effects when giving inferences in Section 5.4.

5.3. *Comparison with alternative models*

Our model relates cause-specific discrete hazards to linear predictors using the multinomial logit link (3.2), with linear predictors given a semiparametric formulation as in (3.3). Available software implementing

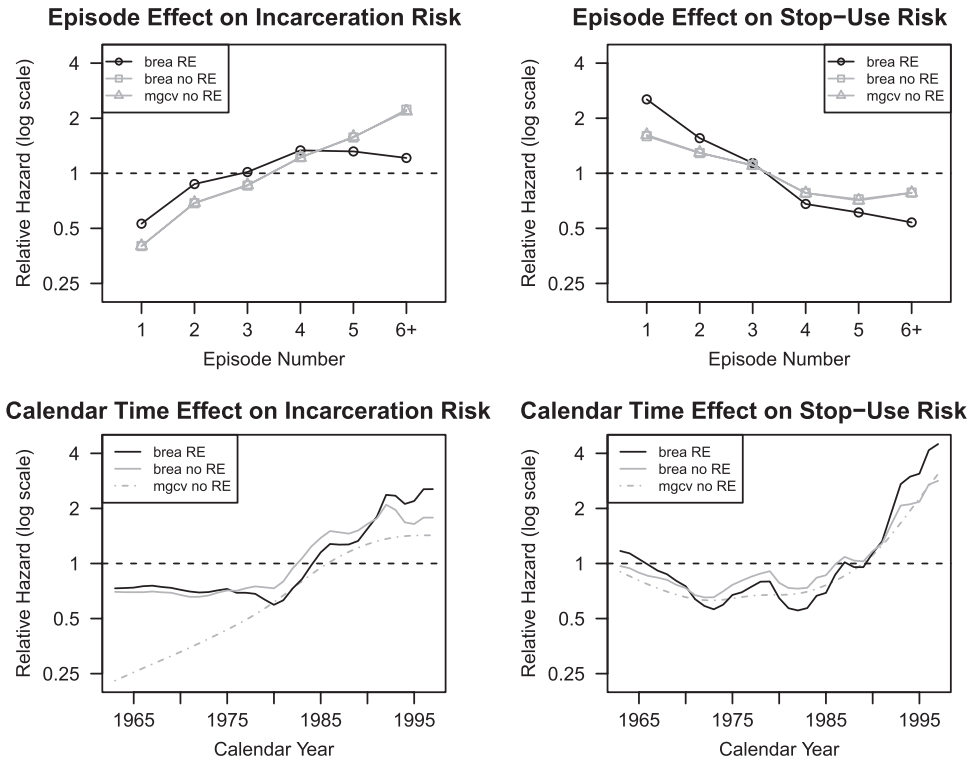

Fig. 2. Estimated relative hazards corresponding to episode number (top row) and absolute calendar time (bottom row) for three models: the full `brea` model including random effects, the `brea` model excluding random effects, and the `mgcv` model excluding random effects; `brea` estimates are posterior medians, while `mgcv` estimates use restricted maximum likelihood.

GAMs allow the user to include some, though not all, of the features we have described. The closest available approach to ours is found in the `mgcv` R package, which provides routines for fitting multinomial-response GAMs with automatic estimation of the amount of smoothness of the arbitrary functions (Wood, 2011, 2019). However, this package does not allow random effects to be used with multinomial models.

We compare our full model implemented in our `brea` R package to two alternatives. First, we use the `brea` package to fit our model with the random effects excluded. Second, we use the `mgcv` package to fit the same multinomial GAM formulation we are using with `brea`, but once again excluding random effects, as they are not available with multinomial responses in `mgcv`. Comparison of our original full model with the `brea` model without random effects provides direct evidence of the consequences of excluding random effects, while comparing the second `brea` model to the `mgcv` model allows us to see the effects of the very different under-the-hood implementations of these two superficially equivalent GAM models.

Figure 2 provides covariate effect estimates for the episode number and calendar time variables on both competing risks for all three models. The `brea` estimates are posterior medians obtained from 20 000 MCMC iterations, which required 4 min running time on a Mac laptop. The `mgcv` results use restricted maximum likelihood estimation, and required just over 30 s on the same laptop. In the top two panels, we see that the estimates of the `brea` and `mgcv` models excluding random effects are almost identical, which shows that the differing implementations of the smooth functions for quantitative covariates are having little impact on the categorical fixed effects estimates.

On the other hand, exclusion of random effects has a substantial impact on the episode number effect estimates, resulting in increased hazard estimates during later episodes for both competing risks. This is due to the fact that we observe later episodes more frequently from subjects with greater tendency to move in and out of a cocaine using state frequently. Random effects explicitly model such tendencies, and when they are excluded, we end up with estimates of the durations of later episodes that are biased towards being too short (i.e., hazards of cessation too high). The full `brea` model predicts that the hazard of voluntary cessation drops by a factor of 4.7 between the first and sixth episodes of cocaine use, whereas when random effects are excluded, we estimate that the hazard of stop-use drops by only a factor of 2.0. This dramatic underestimation of the within-subject decrease in the stop-use rate demonstrates the necessity of random effects.

The bottom two panels of Figure 2 plot the estimated calendar time effects on the risks of incarceration and stop-use. In the first of these panels, we see dramatically different amounts of smoothness inferred by our approach versus that used by `mgcv`. The `mgcv` results appear almost linear over much of the covariate range, and miss important features of the data. In particular, the sharp increase in the incarceration rate beginning in the early 1980s, which corresponds to the beginning of the crack cocaine epidemic in Los Angeles, is completely missed, whereas it is quite apparent in the `brea` results. On the other hand, the smoother `mgcv` curve in the bottom right panel appears to do as good a job fitting the data as the more irregular `brea` curves. Finally, in both panels, we note that exclusion of random effects from the `brea` model appears to make the calendar time effect appear more modest.

5.4. *Substance abuse behavior inferences*

Figure 3 presents inference plots for the three categorical covariates included in our model; covariate values appear along the horizontal axes, while parameters β_k representing corresponding linear predictor contributions appear on the vertical axis; the β_k are first centered for identifiability, and the vertical axis is then labeled with the exponentials of these values, so that these labels correspond to approximate hazard ratios comparing subjects with a specific covariate value to the average hazard. We see in the first panel that the hazard of incarceration among active cocaine users does not depend strongly on the subject's sex, whereas in the second panel we observe that women is more likely to voluntarily stop cocaine use in any given month when compared to comparable men. Hispanic active cocaine users are incarcerated at a rate roughly 40% higher than for black or white users, which are incarcerated at similar rates, though 95% credible intervals for the groups overlap substantially. White subjects stopped using cocaine voluntarily at a rate around 75% higher than black or Hispanic users, who stopped use at similar rates.

As discussed in the previous subsection, number of episodes of cocaine use up to and including the current episode appears strongly related to both competing risks, though in opposite ways. The hazard of becoming incarcerated during an episode of cocaine use for subjects having had three or more previous episodes of use is roughly 2.5 times higher than for subjects during their first episode of use, all else equal. On the other hand, the hazard of a subject voluntarily stopping use drops by almost 80% between the first and sixth episodes of cocaine use. These findings underscore the importance of intervening as early as possible in drug users' lives.

Time effects are summarized in Figure 4. Incarceration risk appears elevated during the first year of a cocaine use episode, and then steadily drops off thereafter, decreasing by almost half between month 12 and month 60 of continuous cocaine use. The hazard of a subject voluntarily ceasing cocaine use is 4.5 times lower in the second month of use than in the first month. Over the subsequent 30 months of use, the stop-use hazard drops by another factor of 2, leveling off thereafter. Spikes in the hazard at months 6, 12, and 24 may represent recall errors due to subjects rounding the durations of episodes they report. These spikes do not show up in the incarceration hazard plot, because official arrest records were used to aid recall.

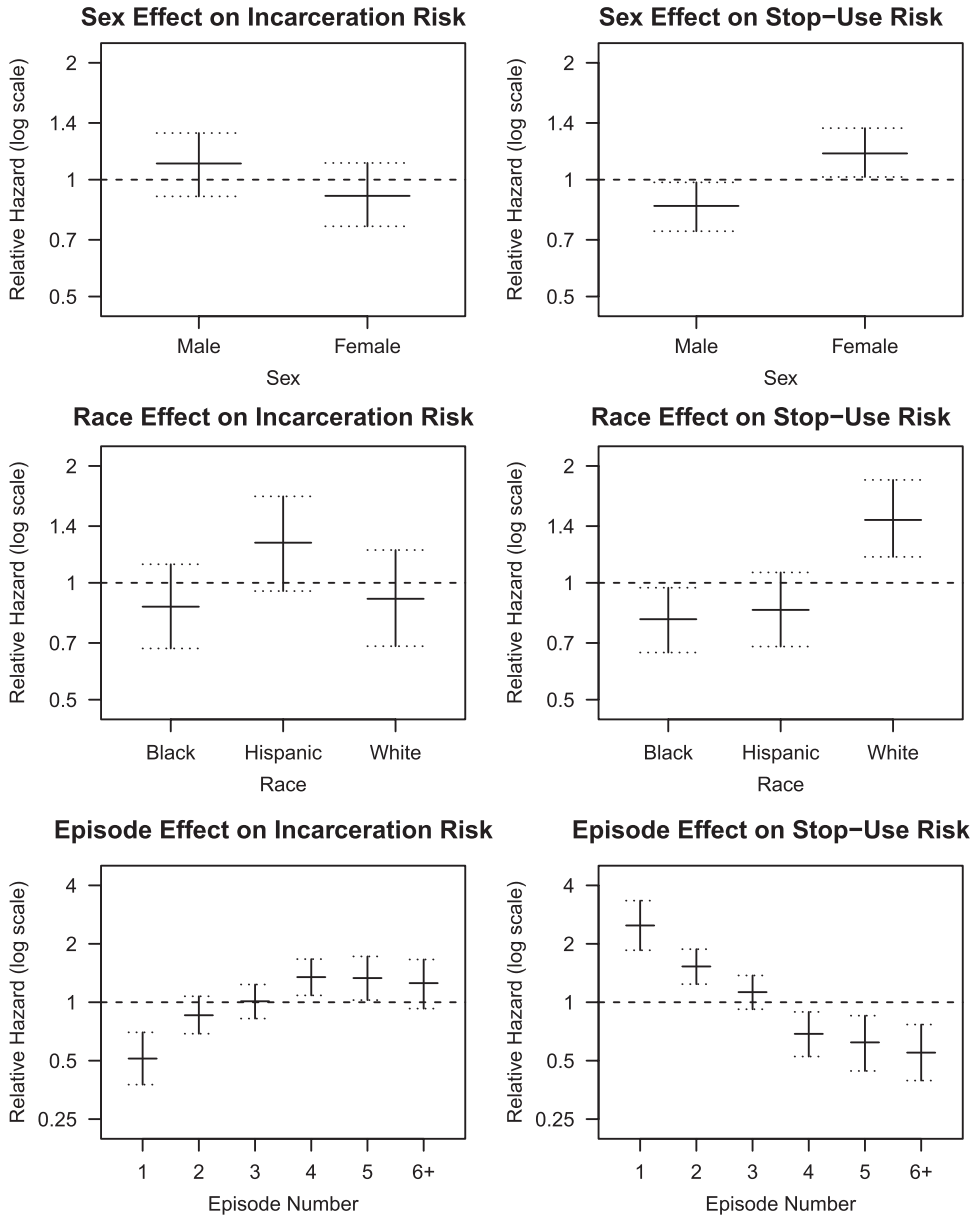

Fig. 3. Posterior medians (solid lines) and point-wise 95% credible intervals (dotted lines) of the categorical covariate effects on the risk of incarceration (left column) and stop-use (right column).

Our model does not provide clear evidence of associations between age and the risks of either incarceration or voluntary use cessation. In contrast, absolute calendar time has a very strong relationship with both competing risks. Incarceration rates for active cocaine users increased 4-fold between 1980 and 1995, coinciding with the crack cocaine epidemic in the Los Angeles region. The chances of a subject voluntarily stopping use appear to have increased even more during the same timeframe. However, this could be

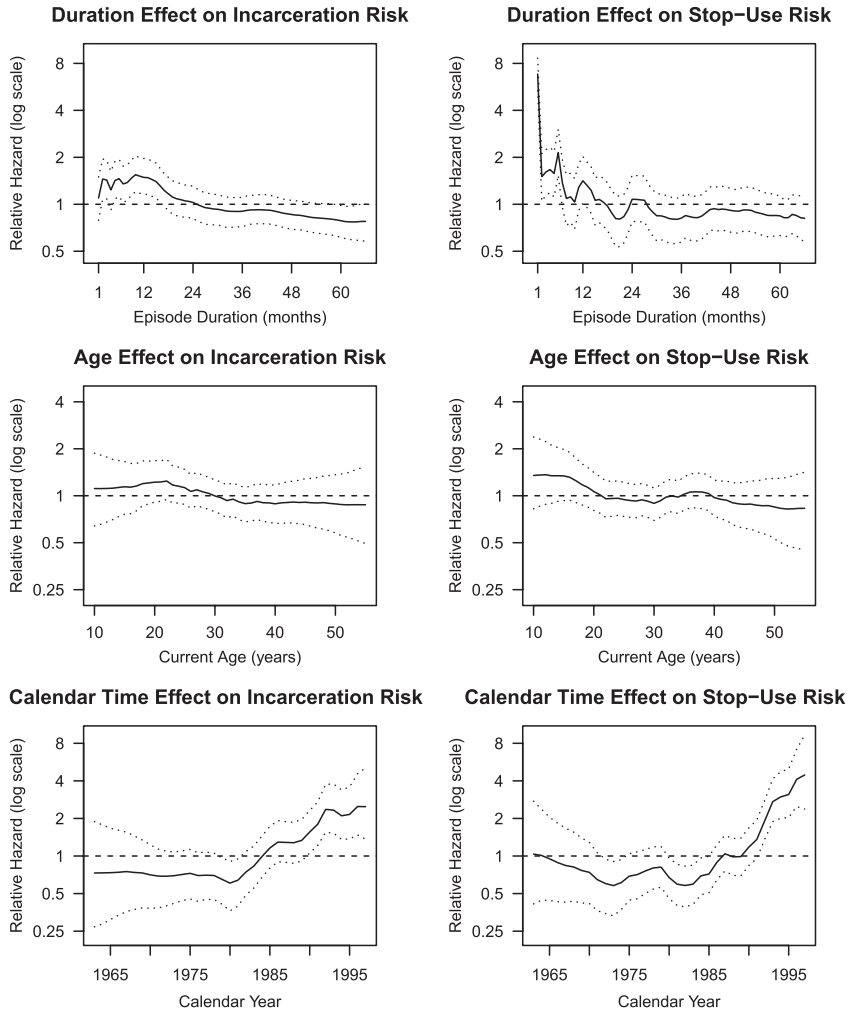

Fig. 4. Posterior medians (solid lines) and point-wise 95% credible intervals (dotted lines) of the time scale effects on the risk of incarceration (left column) and stop-use (right column).

partially attributable to recall bias, with subjects having a more fine-grained recollection of cocaine use episodes (and thus a greater frequency of reported voluntary termination events) in the years just before the interviews were conducted.

6. DISCUSSION

Motivated by data on recurrent episodes of substance abuse, we constructed a general Bayesian discrete-time survival model that allows competing risks, random effects shared across recurrent events and correlated across competing risks, and GAM-style semiparametric incorporation of time-varying covariates. We derived efficient M–H algorithms for posterior inference and have made these available via the CRAN R package `brea`. This complex modeling approach proved essential for our cocaine use application, where omitting key features not available in competing packages, such `mgcv`, results in substantial bias of the fixed effects estimates.

While many quantitative covariates have uniformly smooth relationships with hazard rates, others may have irregular relationships. We observed a dramatic decrease in the hazard of voluntary drug use cessation following the first month of use, with much more gradual changes in the hazard in subsequent months. We addressed this irregularity by allowing the increment variance of our GMRF prior to vary, creating a nonhomogeneous random walk. An alternative approach to our nonhomogeneous formulation would be to allow the change points in the covariate discretization to be random, as in Haneuse *and others* (2008). In this case, a change point for the episode duration effect on voluntary cessation would then be inferred to be between the first and second months of drug use, with change points inferred to be more dispersed in later months.

Interactions between quantitative covariates are not currently implemented in the `brea` package, but are a natural extension of our GMRF approach to smooth functions in a single dimension. If two quantitative covariates are each given fine discretizations, then a GMRF prior on a 2D lattice consisting of pairs of the discretized values of the two covariates may be used to approximate a smooth function of two quantitative predictors. Block M–H updates of the corresponding parameters would then proceed just as in the 1D case. Such interactions may also be used to introduce nonproportionality into our discrete proportional hazards model by including interactions between the time variable and other covariates.

7. Software

The models and algorithms described in this article have been implemented in an R package called `brea`, which is available on the Comprehensive R Archive Network (CRAN) repository (King, 2017). In addition, Appendix C of the supplementary material available at *Biostatistics* online includes an example of applying the `brea` package to a randomized controlled trial of different types of anesthesia.

Supplementary material

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

Funding

The Center for HIV Identification, Prevention, and Treatment (CHIPTS) National Institute of Mental Health (P30MH058107); the UCLA Center for AIDS Research (CFAR) (5P30AI028697), Core H.

Acknowledgments

The authors thank Dr. Yih-Ing Hser of the Integrated Substance Abuse Programs (ISAP) of the UCLA School of Medicine for providing the TUE data. *Conflict of Interest*: None declared.

References

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

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–553.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika* **85**, 215–227.

GELMAN, A., ROBERTS, G. O. AND GILKS, W. R. (1996). Efficient Metropolis jumping rules. In: Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (editors), *Bayesian Statistics 5*. Oxford: Oxford University Press, pp. 599–607.

HANEUSE, S. J.-P. A., RUDSER, K. D. AND GILLEN, D. L. (2008). The separation of timescales in Bayesian survival modeling of the time-varying effect of a time-dependent exposure. *Biostatistics* **9**, 400–410.

HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. New York, NY: Springer.

HSER, Y.-I., BOYLE, K. AND ANGLIN, M. D. (1998). Drug use and correlates among sexually transmitted disease patients, emergency room patients, and arrestees. *Journal of Drug Issues* **28**, 437–454.

KING, A. J. (2017). *brea: Bayesian Recurrent Event Analysis*. R Package Version 0.2.0. https://CRAN.R-project.org/package=brea.

KNORR-HELD, L. AND RUE, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics* **29**, 597–614.

NARENDRANATHAN, W. AND STEWART, M. B. (1993). Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards. *Journal of the Royal Statistical Society, Series C* **42**, 63–83.

PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V., FLOURNOY, N., FAREWELL, V. T. AND BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.

RUE, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B* **63**, 325–338.

RUE, H. AND HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman & Hall/CRC.

SHEPHARD, N. AND PITT, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–667.

WOOD, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B* **73**, 3–36.

WOOD, S. (2019). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. R Package Version 1.8-27. https://CRAN.R-project.org/package=mgcv.

[*Received September 4, 2018; revised July 6, 2019; accepted for publication July 7, 2019*]