Biostatistics (2022) **00**, 00, *pp.* 1–19
https://doi.org/10.1093/biostatistics/kxac025

Bayesian design of clinical trials using joint models for recurrent and terminating events

JIAWEI XU, MATTHEW A. PSIODA, JOSEPH G. IBRAHIM*
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
ibrahim@bios.unc.edu

Summary

Joint models for recurrent event and terminating event data are increasingly used for the analysis of clinical trials. However, few methods have been proposed for designing clinical trials using these models. In this article, we develop a Bayesian clinical trial design methodology focused on evaluating the effect of an investigational product (IP) on both recurrent event and terminating event processes considered as multiple primary endpoints, using a multifrailty joint model. Dependence between the recurrent and terminating event processes is accounted for using a shared frailty. Inferences for the multiple primary outcomes are based on posterior model probabilities corresponding to mutually exclusive hypotheses regarding the benefit of IP with respect to the recurrent and terminating event processes. We propose an approach for sample size determination to ensure the trial design has a high power and a well-controlled type I error rate, with both operating characteristics defined from a Bayesian perspective. We also consider a generalization of the proposed parametric model that uses a nonparametric mixture of Dirichlet processes to model the frailty distributions and compare its performance to the proposed approach. We demonstrate the methodology by designing a colorectal cancer clinical trial with a goal of demonstrating that the IP causes a favorable effect on *at least one* of the two outcomes but no harm on either.

Keywords: Bayesian design; Clinical trials; Joint frailty models; Multiple endpoints; Recurrent events; Sampling prior.

1. Introduction

Recurrent event data are increasingly common in clinical trials. For example, studies may follow patients for new lesions in a metastatic colorectal cancer study (González *and others*, 2005), platelet transfusion and bleeding for myelodysplastic syndromes (Chen *and others*, 2014), or post-stroke hospital readmissions (Duncan *and others*, 2017). Designs based on more time-to-event measurements (i.e., recurrent events) provide greater efficiency and power compared to designs using only one time-to-event endpoint (Chen *and others*, 2014). In situations with recurrent events and a terminating event (e.g., death), a patient's clinical experience will be characterized by both event processes, and the two processes are generally dependent on one another. For example, the relapse of a disease could lead to a high risk of death. The relapse of disease can then be used as a prediction of death risk (Schmoor *and others*, 2013; Conlon *and others*, 2014). Similarly, a terminating event can also preclude the further occurrence of the recurrent events (Liu *and others*, 2004). Under such scenarios, the censorship of recurrent events is no longer

*To whom correspondence should be addressed.

© The Author 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

noninformative but instead depends on the terminating event (Rondeau *and others*, 2007). Therefore, it is imperative to study the recurrent and terminating event processes simultaneously in order to account for their dependence, as well as to evaluate an investigational product's (IP's) effectiveness on both event processes.

Marginal models and frailty models are the two most commonly used approaches to analyze recurrent event process in the presence of a terminating event. Marginal models treat the terminating event as the competing risk for each recurrent event (Li and Lagakos, 1997) or investigate the mean rate function of recurrence given survival (Cook and Lawless, 1997). Ghosh and Lin (2000) developed a nonparametric estimation method based on the marginal mean of the cumulative recurrent event number, and the two events were connected through a linear combination of their weighted log-rank statistics. They also extended the method to semiparametric regression models using inverse probability of censoring/survival weighting (Ghosh and Lin, 2002).

The use of joint frailty models provides a framework for analyzing recurrent event data with informative censoring and modeling the treatment effect on both the recurrent and terminating event processes. Approaches that jointly model recurrent and terminating event data offer possibilities for increased efficiency in the analysis of these types of data. Lancaster and Intrator (1998) proposed the use of a common frailty for modeling the joint distribution of recurrent and terminating events. The two event processes are then independent conditional on the frailty. Liu *and others* (2004) proposed a semiparametric joint model for the intensity function where the shared frailty was allowed to have different effects on the two event processes through a power parameter on the frailty for the terminating event. The same shared frailty model was also used by Rondeau *and others* (2007) with hazard functions estimated by maximizing the penalized likelihood. Paulon *and others* (2020) developed a Bayesian nonparametric joint-modeling approach based on the Dirichlet process mixture prior but treated only the terminating event as the primary outcome when modeling the association between survival times and recurrence of events.

In this article, we develop a Bayesian design for clinical trials involving recurrent and terminating event data, using a multifrailty joint model based on gap times (i.e., time-between-events) for the recurrent event data. Though our development is based on modeling gap times, an alternative development based on, for example, modeling calendar time would be analogous. Our purpose in this article is not to argue for or against a particular strategy for modeling the recurrent event times but rather to develop a framework for how joint models of the type considered can be effectively deployed in design contexts. For ease of exposition, we focus on the design of a parallel two-group randomized, controlled trial. We assume the recurrent event (e.g., readmission to hospital) and terminating event (e.g., death) to be multiple primary endpoints, which correspond to multiple chances to "win" as defined in FDA guidance "Multiple Endpoints in Clinical Trials Guidance for Industry" (Food and Drug Administration, 2017). A superiority test is developed to assess whether the IP has a favorable effect compared to the control, on at least one of the two event processes (i.e., the IP shows benefit with respect to at least one and no harm with respect to either). We propose the use of a joint model with two frailties, where one is specific to the recurrent event hazard to account for the dependence between recurrent event times and the other accounts for the dependence between the recurrent and terminating events. The proposed joint model provides a framework for the characterization of treatment effects on two dependent event processes when both are primary outcomes of interest. In the proposed design, the frailties are modeled via a parametric distribution (i.e., Gamma). We also consider a generalization of the proposed approach that uses a nonparametric mixture of Dirichlet processes to model the frailty distributions. A comparison of the two approaches is conducted through simulation studies.

We develop a simulation-based approach to identify the necessary sample size required to obtain a desired level of Bayesian power while controlling a Bayesian type I error rate. The Bayesian type I error rate and power are defined with respect to sampling prior distributions which are based on the null and alternative hypotheses, respectively (Psioda and Ibrahim, 2018, 2019). We evaluate the operating

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

characteristics of designs based on the multifrailty joint model and a frequentist approach. The frequentist approach is conducted based on a one-sided superiority test using the Cox model, with the recurrent and terminating events as a coprimary outcome. We note that the frequentist approach does not perfectly align with the proposed approach due to the difference in testing structure, since the proposed Bayesian design evaluates whether the IP is effective on at least one of the event processes and not harmful on either, as discussed in detail in Section 3. Our results demonstrate that the proposed design provides similar or higher power estimates compared to the frequentist approach. The type I error rates are well-controlled when there is no effect or a non-negligible harmful effect on either event process.

The rest of this article is organized as follows: In Section 2, we introduce a multifrailty joint model and define a superiority test using both event processes as multiple primary endpoints. We also develop a Bayesian design and sample size determination (SSD) strategy. In Section 3, we use simulations to compare the design based on the proposed joint model to the frequentist approach. We also generalize the proposed approach to use a nonparametric mixture of Dirichlet processes to model the frailty distributions. A comparison of the estimated operating characteristics based on the two approaches is also conducted. We close the article with some discussion in Section 5.

2. Methods

2.1. *Multifrailty joint model*

Using notation similar to Rondeau *and others* (2007), assume the ith patient has a terminating event at time d_i or is censored at time C_i, and X_{ij} is the jth recurrent event for $j = 1, ..., n_i$, where n_i is the total number of observed recurrent events or censoring for patient i. Define $T_{ij} = \min(X_{ij}, d_i, C_i)$ as the observed recurrent event times during follow-up, with $\delta_{ij} = I(t_{ij} = X_{ij})$ indicating whether the jth recurrent event occurred. Similarly, define $T_i^* = \min(d_i, C_i)$ as the last observed time for the ith patient, which is either a time of terminating event or time of censoring. Let $\delta_i^* = I(T_I^* = d_i)$ denote the indicator for whether the patient is censored or not. The gap time (i.e., time-between-events) is then given by $S_{ij} = T_{ij} - T_{i,j-1}$ with $T_{i0} = 0$ for the recurrent event hazard function. Let z_i be the IP indicator, x_{ir} and $x_{i\lambda}$ are the vectors of covariates of interest for the recurrent and terminating event hazard ratio regression models, respectively. For the ith patient, the multifrailty joint model of hazard functions for the recurrent event and terminating event using gap times can be written as follows in equations (2.1) and (2.2), respectively,

$$r_i(s|\mu_i, \nu_i) = \mu_i \nu_i r_0(s) \exp(z_i\gamma_r + x_{ir}'\beta_r) = \mu_i \nu_i r_i(s) \tag{2.1}$$

$$\lambda_i(t|\mu_i) = \mu_i \lambda_0(t) \exp(z_i\gamma_\lambda + x_{i\lambda}'\beta_\lambda) = \mu_i \lambda_i(t) \tag{2.2}$$

where γ_r, β_r and $r_0(t)$ are the treatment effect, vector of coefficient parameters and piecewise constant baseline hazard function for the recurrent event model, while γ_λ, β_λ, and $\lambda_0(t)$ are the analogous quantities for the terminating event model. The frailty $\mu_i \sim \text{Gamma}(1/\theta, 1/\theta)$ accounts for the dependence between the recurrent and terminating event processes, and the frailty $\nu_i \sim \text{Gamma}(1/\eta, 1/\eta)$ accounts for dependence between recurrent event times. The two frailties are assumed independent from each other. Conditional on the frailty ν_i, the gap times for the same patient are mutually independent. When μ_i is fixed, the two event processes are independent. The association parameter θ controls the strength of dependence between the two event processes and, conditional on the variance parameter η, a larger θ reflects a stronger dependence between the recurrent and terminating event processes.

2.2. *Likelihood*

Let $\mathbf{D}$ be the observed data for n patients. Denote $\psi = (\gamma, \beta, \lambda_0, r_0, \theta, \eta)$ as the full set of fixed effect parameters, where $\boldsymbol{\gamma} = (\gamma_r, \gamma_\lambda)$ and $\boldsymbol{\beta} = (\beta_r, \beta_\lambda)$. For the ith patient, the likelihood contribution associated

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

with the time-to-event component of the distribution is given by:

$$\begin{aligned} L_i(\psi|\mu_i,\nu_i,\mathbf{D}) &= \prod_{j=1}^{n_i}\left\{ r_i(T_{ij}|\mu_i,\nu_i)^{\delta_{ij}} \exp(-\mu_i\nu_i \int_{T_{ij-1}}^{T_{ij}} r_i(t)dt)\right\} \\ &\quad \times \lambda_i(T_i^*|\mu_i)^{\delta_i^*} \exp(-\mu_i \int_0^{T_i^*} \lambda_i(t)dt) \\ &= \prod_{j=1}^{n_i}\left\{\mu_i\nu_i r_0(T_{ij}) \exp(z_i\gamma_r + x_{ir}'\beta_r)\right\}^{\delta_{ij}} \exp\{-\mu_i\nu_i R_0(T_i^*)\exp(z_i\gamma_r + x_{ir}'\beta_r)\} \\ &\quad \times \{\mu_i\lambda_0(T_i^*)\exp(z_i\gamma_\lambda + x_{i\lambda}'\beta_\lambda)\}^{\delta_i^*} \exp\{-\mu_i\Lambda_0(T_i^*)\exp(z_i\gamma_\lambda + x_{i\lambda}'\beta_\lambda)\}, \end{aligned} \tag{2.3}$$

where $R_0(t)$ and $\Lambda_0(t)$ are the cumulative piecewise constant baseline hazard functions corresponding to $r_0(t)$ and $\lambda_0(t)$, respectively. The complete data likelihood contribution for the ith patient is obtained by multiplying the likelihood in (2.3) by the distribution for the frailties.

2.3. *Study design*

We consider a design for demonstrating the superiority of an IP compared to a control with respect to recurrent and terminating events as multiple primary outcomes of interest. We follow patients for both recurrent and terminating events starting at baseline. For each patient, recurrent events are documented until a fixed time or the occurrence of the terminating event. The model makes the implicit assumption that, assuming no terminating event occurs, the (stochastic) occurrence of recurrent events will continue indefinitely. If the recurrent and terminating events occur at the same time, only the terminal event is recorded.

2.3.1. *Superiority test* In order to test whether the IP has a favorable effect on at least one of the two event processes (i.e., benefit to at least one and no harm to either), we consider the following group of hypotheses:

$$\begin{aligned} H_1 &: \quad \exp(\gamma_r) > \delta_r \quad \text{or} \quad \exp(\gamma_\lambda) > \delta_\lambda \\ H_2 &: \quad \exp(\gamma_r) > \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) = \delta_\lambda \\ H_3 &: \quad \exp(\gamma_r) = \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) > \delta_\lambda \\ H_4 &: \quad \exp(\gamma_r) = \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) = \delta_\lambda \\ H_5 &: \quad \exp(\gamma_r) < \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) = \delta_\lambda \\ H_6 &: \quad \exp(\gamma_r) = \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) < \delta_\lambda \\ H_7 &: \quad \exp(\gamma_r) < \delta_r \quad \text{and} \quad \exp(\gamma_\lambda) < \delta_\lambda, \end{aligned}$$

where δ_r and δ_λ are prespecified thresholds (e.g., $\delta_r = \delta_\lambda = 1$), $\exp(\gamma_r)$ and $\exp(\gamma_\lambda)$ are hazard ratios of the IP compared to the control for the recurrent event and terminating event, respectively. We consider the union of H_1, H_2, H_3, and H_4 as the null hypothesis (i.e., $H_0 = H_1 \cup H_2 \cup H_3 \cup H_4$) with the alternative as the union of H_5, H_6, and H_7 (i.e., $H_a = H_5 \cup H_6 \cup H_7$). Hypotheses H_1, H_2 and H_3 are consistent with scenarios where the IP has an inferior effect on either one or both event processes. Hypotheses H_5 and H_6 assume that the IP is only effective on one of the event processes but is not harmful with respect to the

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

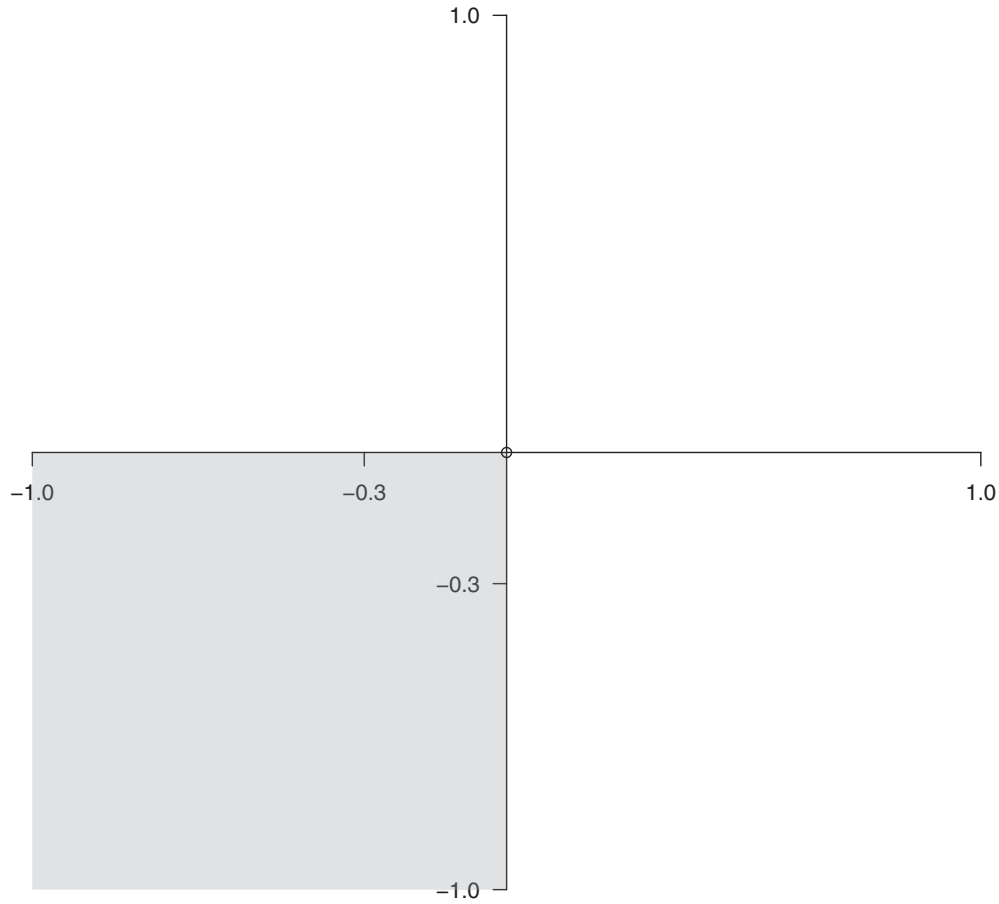

Fig. 1. Depiction of the null and alternative hypothesis space. The gray-shaded region (including the coordinate axis boundaries but not the origin) corresponds to the alternative and the complement of the aforementioned space (including the origin) corresponds to the null hypothesis. The hypothesized treatment effects are −0.3 on both recurrent and terminating event processes.

other. Hypothesis H_7 assumes the IP has a favorable effect on both event processes. Figure 1 provides an overview of the hypothesis space. The shaded area (including the coordinate axis boundaries but not the origin) corresponds to the alternative H_a, and the complement of the aforementioned space (including the origin) corresponds to the null hypothesis H_0. The coordinate axis boundaries of the alternative hypothesis corresponds to the subhypotheses H_5 and H_6.

2.4. *Computation of posterior model probabilities*

Consider the model space of the treatment parameter vector $(\gamma_r, \gamma_\lambda)$, let B_1 denote the full model (i.e., neither treatment parameter constrained), B_2 and B_3 define the models with γ_λ and γ_r fixed, respectively, and B_4 the model having both parameters fixed. We refer to models $\{B_1, B_2, B_3, B_4\}$ as the *Basis* models.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

Table S2 in Appendix C of the Supplementary material available at *Biostatistics* online gives an overview of the relationship between the *Basis* models and the hypotheses.

Let E_j be an indicator that γ takes a value in the parameter space associated with hypothesis H_j, for $j = 1, ..., 7$. One can show that the posterior probabilities for the null and alternative hypotheses, respectively, take the form

$$\begin{aligned} P(H_0|\mathbf{D}) &= P(H_1|\mathbf{D}) + P(H_2|\mathbf{D}) + P(H_3|\mathbf{D}) + P(H_4|\mathbf{D}) \\ &= P(B_1|\mathbf{D}) \cdot P(E_1|B_1, \mathbf{D}) + P(B_2|\mathbf{D}) \cdot P(E_2|B_2, \mathbf{D}) + P(B_3|\mathbf{D}) \cdot P(E_3|B_3, \mathbf{D}) + P(B_4|\mathbf{D}), \\ P(H_a|\mathbf{D}) &= P(H_5|\mathbf{D}) + P(H_6|\mathbf{D}) + P(H_7|\mathbf{D}) \\ &= P(B_2|\mathbf{D}) \cdot P(E_5|B_2, \mathbf{D}) + P(B_3|\mathbf{D}) \cdot P(E_6|B_3, \mathbf{D}) + P(B_1|\mathbf{D}) \cdot P(E_7|B_1, \mathbf{D}), \end{aligned} \tag{2.4}$$

where $P(E_j|B_k, \mathbf{D})$ can be easily obtained by computing the proportion of Monte Carlo Markov chain (MCMC) samples that satisfy H_j given B_k. The proof of (2.4) is given in Appendix D of the Supplementary material available at *Biostatistics* online.

We adapt the methods of Chen (1994) and Chen and Shao (1997) to estimate $P(B_k|\mathbf{D}), k = 1, 2, 3, 4$. Denote $\boldsymbol{\phi} = (\psi, \mu, \nu)$, where $\mu = (\mu_1, ..., \mu_n)$ and $\nu = (\nu_1, ..., \nu_n)$. Let $\phi^{(B_k)}$ be the vector of parameters in *Basis* model B_k that are free to vary. Write $\boldsymbol{\phi} = (\boldsymbol{\phi}^{(-B_k)}, \boldsymbol{\phi}^{(B_k)})$ where $\boldsymbol{\phi}^{(-B_k)}$ is the complementary set of parameters that are fixed under model B_k. Following the method of Chen and Shao (1997), based on the MCMC sample $\{\boldsymbol{\phi}_{(i)}, i = 1, ..., N\}$ from the full model B_1, the posterior probability of model B_k can be estimated as

$$\hat{p}(B_k|\mathbf{D}) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{L\left(\boldsymbol{\phi}_{(i)}^{(B_k)}\right)\pi(\boldsymbol{\phi}_{(i)}^{(B_k)})w(\boldsymbol{\phi}_{(i)}^{(-B_k)}|\boldsymbol{\phi}_{(i)}^{(B_k)})}{L(\boldsymbol{\phi}_{(i)})\pi(\boldsymbol{\phi}_{(i)})}\right)p(B_k)}{\sum_{p=1}^{2^P}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{L(\boldsymbol{\phi}_{(i)}^{(Bp)})\pi(\boldsymbol{\phi}_{(i)}^{(Bp)})w(\boldsymbol{\phi}_{(i)}^{(-Bp)}|\boldsymbol{\phi}_{(i)}^{(Bp)})}{L(\boldsymbol{\phi}_{(i)})\pi(\boldsymbol{\phi}_{(i)})}\right)p(B_p)} \tag{2.5}$$

for $k = 1, 2, 3, 4$, where $P = 2$ (the number of treatment effect parameters) in our case and $\boldsymbol{\phi}_{(i)} = (\boldsymbol{\phi}_{(i)}^{(-B_k)}, \boldsymbol{\phi}_{(i)}^{(B_k)})$.

Following Chen (1994), the weight function $w(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)})$ is a completely known conditional density of $\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)}$, with the optimal choice of $w(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)}) = p(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)}, \mathbf{D})$. Since a closed form for $p(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)}, \mathbf{D})$ is typically not available, an empirical procedure is used to select $w(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)})$. Specifically, we first compute the sample mean and covariance matrix $(\tilde{\boldsymbol{\phi}}, \tilde{\Sigma})$ based on the MCMC samples $\{\boldsymbol{\phi}_{(i)}, i = 1, ..., N\}$. Then, $w(\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)})$ can be approximated using the conditional density of $\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)}$ based on a normal approximation. Specifically, $\boldsymbol{\phi}^{(-B_k)}|\boldsymbol{\phi}^{(B_k)} \sim N(\tilde{\boldsymbol{\phi}}_k, \tilde{\Sigma}_k)$ where $\tilde{\boldsymbol{\phi}}_k = \tilde{\boldsymbol{\phi}}^{(-B_k)} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(\boldsymbol{\phi}^{(B_k)} - \tilde{\boldsymbol{\phi}}^{(B_k)})$ and $\tilde{\Sigma}_k = \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{12}'$.

For example, under model B_3, if $\boldsymbol{\phi}^{(-B_3)} = \gamma_r$ and $\boldsymbol{\phi}^{(B_3)} = (\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mu, \nu)$, the corresponding weight function is approximated as

$$\begin{aligned} w(\boldsymbol{\phi}^{(-B_3)}|\boldsymbol{\phi}^{(B_3)}) &= p(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mu, \nu, \mathbf{D}) \\ &= \frac{p(\mu, \nu|\gamma_r, \gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})}{p(\mu, \nu|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})} p(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D}) \end{aligned}$$

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

$$
\begin{aligned}
&= \frac{\prod_i^n p(\mu_i, \nu_i|\gamma_r, \gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})}{\prod_i^n p(\mu_i, \nu_i|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})} p(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D}) \\
&= \prod_i^n \frac{p(\mu_i, \nu_i|\gamma_r, \gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})}{p(\mu_i, \nu_i|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D})} \times p(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D}) \\
&\approx p(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D}) \\
&\approx N(\gamma_r|\gamma_\lambda, \boldsymbol{\beta}, r_0, \lambda_0, \theta, \eta, \mathbf{D}).
\end{aligned}
$$

The conditional density of (μ_i, ν_i) given the fixed effects except γ_r (denominator) is approximately equal to the conditional density given all fixed effects (numerator) because the frailties are patient-specific. Therefore, the posterior model probability (PMP) can be computed based on the conditional density of the fixed effects ψ. Elicitation of the prior model probability used in (2.5) is discussed in Section 2.5.

2.4.1. *Algorithm for computing PMPs* The PMPs can be computed following the algorithm

i Draw MCMC samples $\boldsymbol{\phi}_{B_1}, \boldsymbol{\phi}_{B_2}, \boldsymbol{\phi}_{B_3}$ from models B_1, B_2, B_3, respectively, with each sample vector having N elements. Note that B_1 is the full model.
ii Compute the posterior probability $P(E_j|B_k, \mathbf{D})$ based on the MCMC samples, where

$$P(E_j|B_k, \mathbf{D}) \approx \frac{1}{N} \sum_{i=1}^{N} 1\{\boldsymbol{\phi}_{B_k,i} \in E_j\}$$

and $1\{A\}$ is an indicator that A is true.
iii Using $\boldsymbol{\phi}_{B_1}$, $P(B_k|\mathbf{D})$ can be estimated for $k = 1, 2, 3, 4$ following the method given in (2.5).
iv We estimate $P(H_j|\mathbf{D})$ for $j \in \{0, a\}$ using (2.4).

2.5. *Prior model probabilities*

Define $\Delta_{0r} = \log(\delta_r)$ and $\Delta_{0\lambda} = \log(\delta_\lambda)$. Under the assumption that $\gamma_r \perp\!\!\!\perp \gamma_\lambda$ *a priori*, we have the prior distribution $\pi(\boldsymbol{\gamma}) = \pi(\gamma_r)\pi(\gamma_\lambda)$. Both $\pi(\gamma_r)$ and $\pi(\gamma_\lambda)$ are specified as mixture distributions with $\pi(\gamma_r) = \pi_r \cdot 1(\gamma_r = \Delta_{0r}) + (1 - \pi_r) \cdot f_r(\gamma_r)$ and $\pi(\gamma_\lambda) = \pi_\lambda \cdot 1(\gamma_\lambda = \Delta_{0\lambda}) + (1 - \pi_\lambda) \cdot f_\lambda(\gamma_\lambda)$, where $f_r(\cdot)$ and $f_\lambda(\cdot)$ are Normal(ω_r, σ_r^2) and Normal$(\omega_\lambda, \sigma_\lambda^2)$, respectively. This prior formulation *induces* the following prior model probabilities for the *Basis* models as

$$
\begin{aligned}
P(B_1) &= (1 - \pi_r)(1 - \pi_\lambda), \\
P(B_2) &= (1 - \pi_r)\pi_\lambda, \\
P(B_3) &= \pi_r(1 - \pi_\lambda), \\
P(B_4) &= \pi_r\pi_\lambda.
\end{aligned}
$$

The induced prior probabilities for the hypotheses are then defined as

$$
\begin{aligned}
P(H_1) &= P(B_1)\{1 - F_\lambda(\Delta_{0\lambda})F_r(\Delta_{0r})\}, \\
P(H_2) &= P(B_2)\{1 - F_r(\Delta_{0r})\},
\end{aligned}
$$

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

$$
\begin{aligned}
P(H_3) &= P(B_3)\{1 - F_\lambda(\Delta_{0\lambda})\},\\
P(H_4) &= P(B_4),\\
P(H_5) &= P(B_2)F_r(\Delta_{0r}),\\
P(H_6) &= P(B_3)F_\lambda(\Delta_{0\lambda}),\\
P(H_7) &= P(B_1)F_r(\Delta_{0r})F_\lambda(\Delta_{0\lambda}),
\end{aligned}
$$

where $F_r(\Delta_{0r}) = \int 1(\gamma_r < \Delta_{0r}) f_r(\gamma_r) d\gamma_r = \Phi(\frac{\Delta_{0r} - \omega_r}{\sigma_r})$ and $F_r(\Delta_{0\lambda}) = \int 1(\gamma_\lambda < \Delta_{0\lambda}) f_\lambda(\gamma_\lambda) d\gamma_\lambda = \Phi(\frac{\Delta_{0\lambda} - \omega_\lambda}{\sigma_\lambda})$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution. As shown above, the priors for the treatment effects $\pi(\gamma_r)$ and $\pi(\gamma_\lambda)$ together determine the prior model probabilities for the *Basis* models, which further determine the induced prior probabilities for the hypotheses. Therefore, one only needs to specify the prior distributions for treatment effects on the recurrent and terminating event processes in order to elicit the prior model probabilities for the *Basis* models and the prior probabilities for the hypotheses.

We provide a general scheme for the elicitation of priors when there is no prior information available. Without loss of generality, we assume $\pi_r = \pi_\lambda = \pi$ with both $f_r(\cdot)$ and $f_\lambda(\cdot)$ centered at 0 (i.e., $\omega_r = \omega_\lambda = 0$). Thus, we have $F_r(\Delta_{0r} = 0) = F_r(\Delta_{0\lambda} = 0) = \frac{1}{2}$ and for the alternative hypothesis,

$$
\begin{aligned}
P(H_5) + P(H_6) + P(H_7) &= \pi(1-\pi)/2 + \pi(1-\pi)/2 + (1-\pi)^2/4\\
&= -\frac{3}{4}\pi^2 + \frac{1}{3}\pi + \frac{1}{4},
\end{aligned}
$$

which is maximized at $(\pi_r, \pi_\lambda) = (\frac{1}{3}, \frac{1}{3})$. Therefore, the maximum weight the alternative hypothesis can obtain is $\frac{1}{3}$ when $\pi_r = \pi_\lambda = \frac{1}{3}$. For the standard deviation in $f_r(\cdot)$ and $f_\lambda(\cdot)$, we propose using $\sigma_r = \rho \cdot |\Delta_{1r}|$ and $\sigma_\lambda = \rho \cdot |\Delta_{1\lambda}|$, where Δ_{1r} and $\Delta_{1\lambda}$ are the hypothesized treatment effects on the recurrent and terminating event processes, respectively. Details and further discussion about the choice of ρ can be found in Section 3 and Appendix A of the Supplementary material available at *Biostatistics* online.

2.6. *Estimation and inference*

Following Psioda and Ibrahim (2018, 2019), we define the Bayesian type I error rate and power using user-specified null and alternative sampling prior distributions for $\boldsymbol{\psi}$, respectively. Sampling priors have also been referred to as design priors (O'Hagan and Stevens, 2001) and their use can be traced back to the work of Brown *and others* (1987). We use the term sampling prior throughout this article. A sampling prior specifies a probability distribution for the model parameters conditional on a particular hypothesis being true (Psioda and Ibrahim, 2018). In our context, the null (alternative) sampling prior gives nonzero weight to values of $\boldsymbol{\psi}$ such that H_0 (H_a) is satisfied. The sampling prior distributions are referred to as such because they are used to sample parameter values in the simulation-based estimation procedure for the Bayesian type I error rate and power. For the special case where the sampling priors place a point-mass on a fixed value of the model parameters such that $\pi_0^{(s)}(\boldsymbol{\psi}) = 1(\boldsymbol{\psi} = \boldsymbol{\psi}_0)$ and $\pi_a^{(s)}(\boldsymbol{\psi}) = 1(\boldsymbol{\psi} = \boldsymbol{\psi}_a)$, which is our focus in this article, the Bayesian type I error rate and power closely align with the frequentist versions. The superscript (s) indicates that the prior is a sampling prior.

Let $\alpha^{(s)}$ and $\beta^{(s)}$ denote the Bayesian type I and II error rates. Prespecify p_0 as the threshold for substantial evidence such that we reject the null hypothesis if $P(H_a|\mathbf{D}) \geq p_0$. For a fixed value of $\boldsymbol{\psi}$, the null hypothesis rejection rate is defined as $r(\boldsymbol{\psi}) = E[1\{P(H_a|\mathbf{D}) \geq p_0\}|\boldsymbol{\psi}]$. The Bayesian type I error rate and power are

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

defined as $\alpha^{(s)} = E[r(\boldsymbol{\psi})|\pi_0^{(s)}]$ and $1-\beta^{(s)} = E[r(\boldsymbol{\psi})|\pi_a^{(s)}]$, which, for nondegenerate sampling priors, are weighted averages of $r(\boldsymbol{\psi})$ with weights determined by $\pi_0^{(s)}(\boldsymbol{\psi})$ and $\pi_a^{(s)}(\boldsymbol{\psi})$, respectively.

The posterior distribution satisfies $\pi(\psi,\mu,\nu|\mathbf{D}) \propto L(\gamma,\beta,r_0,\lambda_0|\mu,\nu,\mathbf{D}) \times f(\mu|\theta) \times f(\nu|\eta) \times \pi^{(f)}(\psi)$ where $\pi^{(f)}(\psi)$ is the fitting prior. We assume the initial fitting prior is obtained by taking

$$\pi^{(f)}(\psi) = \pi_0(\gamma)\pi_0(\beta)\pi_0(\theta)\pi_0(\eta)\pi_0(\lambda_0)\pi_0(r_0)$$

and assume the following proper priors:

$$\begin{aligned}
\pi_0(\beta) &\propto \text{Normal}(0, \Sigma_0 = 5^2 \cdot I)\\
\pi_0(\theta) &\propto \text{Gamma}(a_\theta = 1.1, b_\theta = 1.1)\\
\pi_0(\eta) &\propto \text{Gamma}(a_\eta = 1.1, b_\eta = 1.1)\\
\pi_0(\lambda_0) &\propto \prod_{q=1}^{Q} \text{Gamma}(a_q = 0.1, b_q = 0.1)\\
\pi_0(r_0) &\propto \prod_{p=1}^{P} \text{Gamma}(a_p = 0.1, b_p = 0.1).
\end{aligned}$$

The gamma priors specified above use the shape and rate parameterization. For example, taking $a_q = b_q = 0.1$ gives mean 1 and variance of 10 (noninformative prior) and $a_\theta = b_\theta = 1.1$ gives mean 1 and variance of 0.91 (weakly informative prior to help stabilize the MCMC sampling of frailties).

2.7. *Sample size determination*

We propose using simulations to identify the required number of terminating events (effectively the sample size in the trials) such that the trial design has sufficiently high Bayesian power. The number of patients enrolled in the trial may be chosen to obtain a specified number of terminating events in a specified interval of time *on average*. Let n and v denote the sample size and number of terminating events, respectively. We consider an approach that fixes the ratio $r = \frac{n}{v}$ but varies the number of terminating events. If n_1 patients result in obtaining v_1 terminating events in a specific time frame, then to obtain $v_2 \geq v_1$ terminating events in the same time frame, one should increase n_2 proportionally.

We want to determine the smallest v such that the Bayesian power for the design is at least $1-\beta^{(s)}$. A simulation-based SSD procedure is given below:

S1. Let $v_1, ..., v_K$ denote the potential terminating event totals at which the trial might be stopped.
S2. Initialize $k = 1$.
S3. Compute the Bayesian power $1-\beta_k^{(s)}$ based on v_k.
S4. If $1-\beta_k^{(s)} \geq 1-\beta^{(s)}$ then set $v = v_k$ and stop; otherwise, increment k and return to S3.

Note that the approximate Bayesian type I error rate will be $\alpha^{(s)}$ when one takes $p_0 = 1-\alpha^{(s)}$ for the case where point-mass null sampling priors are used (along with a noninformative fitting prior). Thus, for the identified choice of v, it will generally be the case that $\alpha^{(s)} \approx 1-p_0$ and so specific efforts to control the Bayesian type I error rate at level $\alpha^{(s)}$ are not generally needed when p_0 is chosen in this way. Nonetheless, one can always compute the exact Bayesian type I error rate via simulation to ensure it is sufficiently close to the desired nominal level.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

Now we expound more on Step S3 from the simple algorithm given above. Letting R be the number of simulation studies to be performed, to estimate the Bayesian power $1-\beta_k^{(s)}$ associated with terminating event total v_k, one does the following:

S3.1 Sample $\boldsymbol{\psi}^{(b)}$ from the alternative sampling prior $\pi_1^{(s)}(\boldsymbol{\psi})$.

S3.2 Simulate the *observed* data $\mathbf{D}^{(b)}$ as defined below.

S3.3 Estimate the PMP $P(H_a|\mathbf{D})$ using an approach described in Section 2.4 and compute the null hypothesis rejection indicator

$$r^{(b)} = 1\{P(H_a|\mathbf{D}^{(b)}) \geq p_0\}.$$

S3.4 Approximate the Bayesian power as:

$$1-\beta_k^{(s)} \approx \frac{1}{R}\sum_{b=1}^{R} r^{(b)}.$$

The observed trial data $\mathbf{D}^{(b)}$ as required for Step S3.2 can be simulated using the following steps. For patient $i = 1, ..., n$, we do the following:

1. Simulate the enrollment time r_i using a chosen enrollment distribution.
2. Simulate x_i based on a treatment distribution and $\boldsymbol{z}_{ir}$ and $\boldsymbol{z}_{i\lambda}$ based on the covariate distribution.
3. Simulate μ_i from Gamma$(1/\theta, 1/\theta)$ and ν_i from Gamma$(1/\eta, 1/\eta)$.
4. Simulate o_i from a uniform distribution and compute the time-to-event $d_i = S_i^{-1}(o_i)$, where $S_i(t)$ is the survival function with hazard $\lambda_i(t) = \mu_i\lambda_0(t)\exp(z_i\gamma_\lambda + x'_{i\lambda}\beta_\lambda)$.
5. Simulate e_i from a uniform distribution and compute the gap times of recurrent events $y_{ik} = R_i^{-1}(e_i)$, where $R_i(t)$ is the survival function with hazard $r_i(t) = \mu_i\nu_i r_0(t)\exp(z_i\gamma_r + x'_{ir}\beta_r)$.
6. Simulate the time-to-censorship C_i based on the chosen censorship distribution.
7. Set $T_i^* = \min(d_i, C_i)$ and $\delta_i^* = I(d_i \leq C_i)$.

Let $t_{\max}$ be the duration of time from the first enrollment to the time at which the target number of terminating events is reached (or the maximum study duration is reached). Then, for patient $i = 1, ..., n$,

1. Remove any patient with $r_i \geq t_{\max}$ (Patients whose simulated enrollment time occurs after the study terminates).
2. If $T_i^* > t_{\max}$, set $T_i^* = t_{\max} - r_i$ and $\delta_i^* = 0$.
3. Compute $N_i(T_i^*) = \max_j(j : \sum_{k=1}^{j} y_{ik} \leq T_i^*)$ and record all gap times for $j \leq N_i(T_i^*)$ (i.e., $T_{ij} = \sum_{k=1}^{j} y_{ik}$ and $\delta_{ij} = 1$ for $j \leq N_i(T_i^*)$).

3. EXAMPLE APPLICATION: BAYESIAN CLINICAL TRIAL FOR COLORECTAL CANCER

Our design methodology is motivated by a colorectal cancer study conducted at Hospital Universitary in L'Hospitalet, Spain (González *and others*, 2005). The study investigated sex-based inequalities in hospital readmission among patients diagnosed with colorectal cancer. There were 403 patients diagnosed between January 1996 and December 1998, and they were actively followed up until 2002. Hospital readmission times related to colorectal cancer after surgery were collected, with mortality also recorded during follow-up.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

In the example application, we consider a design evaluating an IP (e.g., chemotherapy) with respect to hospital readmission times and mortality as multiple primary outcomes. For each patient, readmission to a hospital is recorded whenever it occurs until some fixed time (e.g., 6 years) or the occurrence of the terminating event. If hospital readmission and mortality occur at the same time (i.e., the patient dies during hospitalization), only the terminal event will be recorded. In the simulated trials, patients were randomized to two treatment arms using a 1:1 allocation scheme and the accrual rate was simulated to be uniform over a 1-year period. Censoring (i.e., dropout) was assumed to follow a mixture distribution whereby patients had a 0.05 probability of dropping out of the trial early and, conditional on being a dropout, the time to dropout was uniform over a 6-year period. All patients were administratively censored when the total number of terminating events for the simulated trial was reached. In the colorectal cancer study, gender was prognostic of readmission to a hospital and so we included a binary covariate (male vs female) in our hypothetical trial simulations, where $x_{ir} = x_{i\lambda}$ is an indicator that the patient's gender is female. The covariate was simulated such that approximately 50% of the subjects were females.

We assumed piecewise constant baseline hazards for both the recurrent and terminating hazard functions. In the design simulations, we considered five-component piecewise constant functions with knots at times (days) {12.0,56.5,179.0,418.0} and {128.5,260.5,488.0,791.5} for recurrent and terminating hazards, respectively. Knot placement was determined by fitting the proposed joint model to the colorectal cancer data such that each time interval of the hazards had approximately the same number of events.

For the treatment effects, we not only proposed the hypothesized effects on both hospital readmission and mortality as $\Delta_{1r} = \Delta_{1\lambda} = -0.3$ but also allowed various sampling priors of the treatment effects to study the trends in Bayesian type I error rate and power. For treatment effects on both event processes, we considered point-mass priors $\gamma_r, \gamma_\lambda \in \{-0.60, -0.50, -0.40, -0.30, -0.20, -0.10, 0.00\}$ for a favorable or no effect and $\gamma_r, \gamma_\lambda \in \{0.02, 0.04, 0.06\}$ for a harmful effect. The nuisance parameter values were taken to equal the approximate posterior modes based on our analysis of the colorectal cancer data as shown in Table S3 in Appendix C of the Supplementary material available at *Biostatistics* online. We implemented the prior distributions proposed in Section 2.5 with $\rho = 2$ which leads to prior probabilities for the null and alternative hypothesis equal to $\frac{2}{3}$ and $\frac{1}{3}$, respectively. The prior distributions for the treatment effects on both event processes are the same as $\pi(\gamma) = \frac{1}{3} \cdot 1(\gamma = 0) + \frac{2}{3} \cdot \mathrm{N}(0, 0.6)$, where γ is used hereto to represent the treatment effect on the recurrent or terminating event process. Further discussion about the different choices of ρ can be found in Appendix A of the Supplementary material available at *Biostatistics* online. To identify the desired number of terminating events required to achieve a Bayesian power equal to 0.8, we considered $v = 350$ to 500 in increments of 25. A total of 4000 simulated trials were performed to estimate the operating characteristics for each choice of sampling prior and each sample size considered.

We evaluate the performance of the proposed approach against a frequentist approach through type I error rate and power estimates. The frequentist approach is implemented based on a one-sided superiority test using the Cox model, with both the recurrent and terminating event hypotheses taken to be coprimary. For that approach, we test whether the IP has a favorable effect on both event processes, and the dependence between recurrent event times is accounted for using the marginal approach of Wei *and others* (1989). Note that the frequentist approach does not perfectly align with the proposed Bayesian approach due to the difference in how the alternative hypothesis is defined. It can be seen from Figure 1 that for the alternative hypothesis, the Bayesian approach includes the coordinate axis boundaries but the frequentist approach excludes them. Therefore, the frequentist approach does not have the capability to evaluate whether the IP is beneficial to at least one of the event processes and not harmful to either, whereas the Bayesian approach does.

In Sections 3.1 and 3.2, we compared the performances of the Bayesian and frequentist approaches in terms of type I error rate and power, respectively, when the number of terminating events $v = 400$. Section 3.3 presents the estimated type I error rate and power curves for the proposed approach with respect to different sample sizes.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

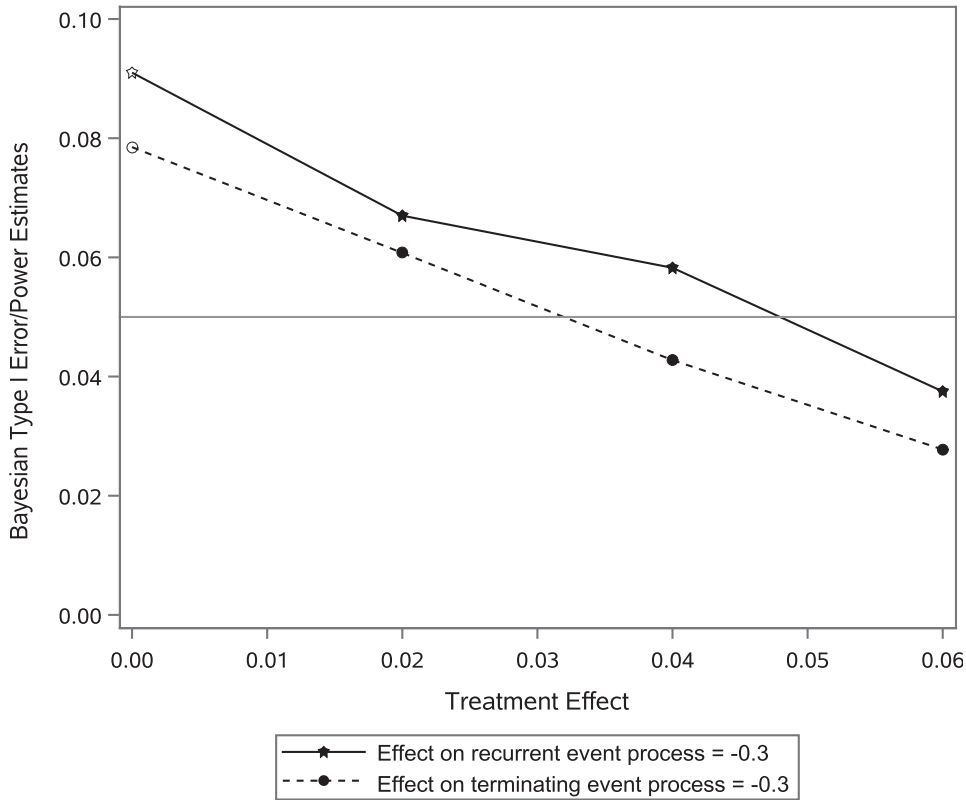

Fig. 2. Estimated Bayesian type I error rate curves when the IP has a favorable effect on one event process but has no effect or causes a modest degree of harm on the other.

3.1. *Evaluating the Bayesian type I error rates*

In this section, we compare the type I error rate estimates based on the proposed design and the frequentist approach. We first considered the scenario where there is no treatment effect on either the recurrent or terminating event hazard (i.e., $\gamma_r = \gamma_\lambda = 0$). Fitting the Cox model to the two time-to-event endpoints (recurrent and terminal) separately produces type I error rates equal to the nominal level (i.e., 0.05) individually. Both the proposed Bayesian approach and the frequentist approach have a conservative type I error rate (< 0.01) for their respective joint hypotheses regarding both effects. Figure 2 presents the estimated Bayesian type I error rate curves for the proposed Bayesian approach when the IP has a favorable effect on one event process but has no effect or causes a modest degree of harm on the other. The favorable effect is assumed to equal the hypothesized level (i.e., $\gamma_r = -0.3$ or $\gamma_\lambda = -0.3$). For the nonbeneficial effect, we varied its value from no effect to modest harm to study how the type I error rate changes near the hypothesis boundary. Note that a favorable effect for one outcome coupled with no effect on the other corresponds to the alternative hypothesis under the Bayesian approach and hence represents Bayesian power. According to the curves, the estimated Bayesian type I error rate was slightly inflated when the treatment effect was close to 0 (e.g., $\gamma_r = -0.3$ with $\gamma_\lambda = 0.02$ resulting in a type I error rate around 0.07). However, the estimates decreased quickly to/below the nominal level of 0.05 when the effects on either the recurrent or terminating event process is modestly harmful (e.g., effect equal to 0.03 to 0.05 or approximately 10–17% of the magnitude of the *hypothesized* effect). The estimated Bayesian type I

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

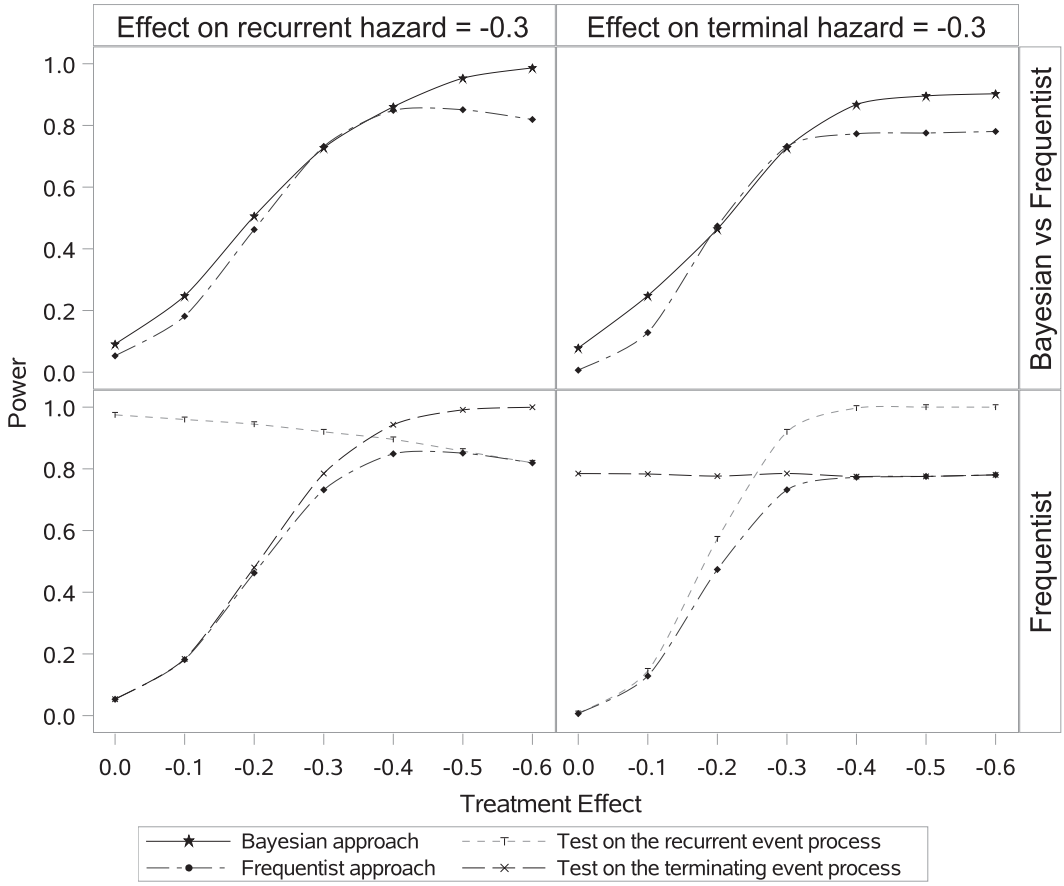

Fig. 3. Estimated power curves based on various point-mass sampling priors for the treatment effects. The prior for the treatment effect for one event process was taken equal the hypothesized level, and it was varied for the other event process.

error curves where treatment effect values were very close to the hypothesis boundaries are presented in Appendix B of the Supplementary material available at *Biostatistics* online.

3.2. *Evaluating power*

Figure 3 compares the power curves estimated based on the Bayesian and frequentist approaches with various point-mass sampling priors of treatment effects. We took the prior for the treatment effect on the recurrent event process equal to the hypothesized level (i.e., $\gamma_r = -0.3$) and varied the prior of the effect on the terminating event process $\gamma_\lambda \in \{0.0, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6\}$. The corresponding curves are shown in column 1 of Figure 3. We also considered the prior for the treatment effect on the terminating event process equal the hypothesized level and varied the prior of the effect on the recurrent event process $\gamma_\lambda \in \{0.0, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6\}$. The corresponding curves are presented in column 2. Compared with the frequentist approach, the Bayesian design produced comparable power estimates when the favorable effect on the terminal event process is moderate but gave significantly higher

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

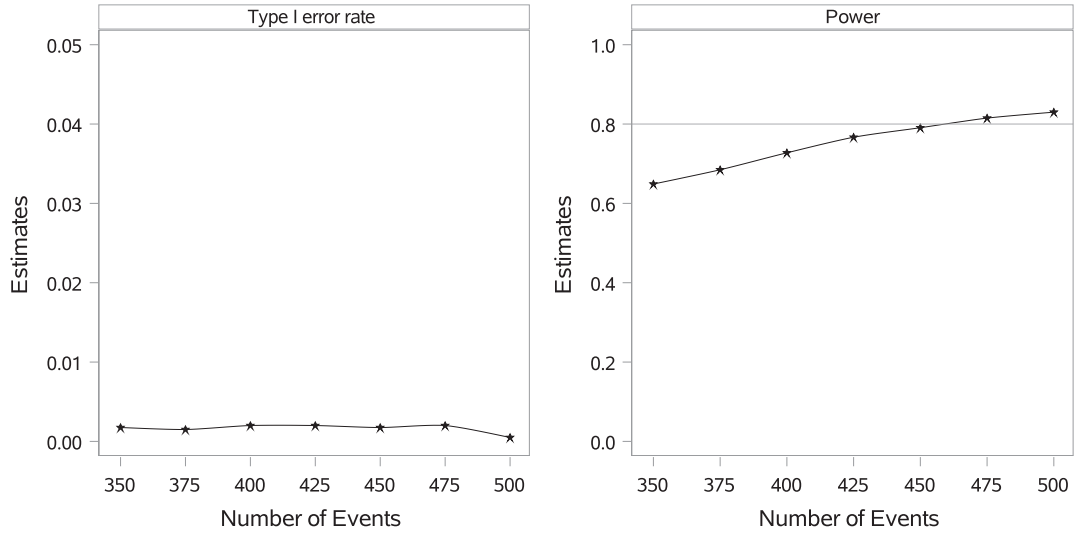

Fig. 4. Bayesian type I error rate and power curves based on varying numbers of required terminal events.

estimates when the effect is small or large. Similar patterns can also be observed when a favorable effect on the terminating event process was fixed equal to the hypothesized level but varied on the recurrent event process. The power curves based on the frequentist approach are presented in the row 2 of Figure 3, including the curves estimated using the Cox model on the recurrent and terminating event data separately. Because the frequentist power was computed based on a one-sided superiority test with respect to both events as coprimary outcomes, its value was then bounded by the smaller value between the two curves.

3.3. *SSD example*

Figure 4 demonstrates the estimated Bayesian type I error rate and power curves based on the proposed Bayesian approach regarding different sample sizes (number of terminating events). Treatment effects on both event processes were assumed to equal the hypothesized level. Based on the type I error curve, the proposed approach controlled the Bayesian type I error rate conservatively below the nominal level of 0.05, regardless of the sample size. The estimated Bayesian power reached the targeted 0.8 when the number of terminating event is around 460.

4. JOINT MODEL WITH MIXTURE OF DIRICHLET PROCESSES

In Section 2, we developed a multifrailty joint model with both frailties having parametric distributions (i.e., Gamma), and we demonstrated the proposed design's properties using simulation studies in Section 3. In this section, we extend the proposed model to a joint model with nonparametric frailty distributions, using mixtures of Dirichlet processes, to investigate the performance of our design strategy in the presence of patient heterogeneity. The only difference between the proposed and nonparametric approaches is how we modeled the frailty distributions. For the proposed approach, we assumed for the ith patient,

$$\mu_i \sim \text{Gamma}(1/\theta, 1/\theta),$$
$$\nu_i \sim \text{Gamma}(1/\eta, 1/\eta),$$

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

which was described in Section 2. For the nonparametric approach, mixtures of Dirichlet process priors were implemented for the estimation of the frailties with details given below.

4.1. *Multifrailty joint model with mixtures of Dirichlet processes*

For the ith patient, we assume

$$\begin{aligned} \mu_i &\sim G_1, \qquad [G_1|c_1,\psi] \sim DP(c_1, G_{01}(\theta_0)) \\ \nu_i &\sim G_2, \qquad [G_2|c_2,\psi] \sim DP(c_2, G_{02}(\eta_0)), \end{aligned}$$

where G_1 and G_2 are general distributions, c_1 and c_2 are positive scalars. The probability measures $G_{01}(\cdot)$ and $G_{02}(\cdot)$, also called base measures, are "guesses" at the parametric distributions for μ_i and ν_i, respectively. We took $G_{01}(\cdot)$ and $G_{02}(\cdot)$ to be gamma distributions with means 1 and variances θ_0 and η_0, respectively, where θ_0 and η_0 were taken to equal the approximate posterior modes based on our analysis of the colorectal cancer data as shown in Table S3 in Appendix C of the Supplementary material available at *Biostatistics* online. The scalar c_k is a confidence parameter reflecting a prior belief about how similar the distribution G_k is to the base measure $G_{0k}(\cdot)$, for $k = 1, 2$ (Kleinman and Ibrahim, 1998). Following Escobar (1994), the full conditional distributions of μ_i and ν_i are derived as:

$$p(\mu_i|\mathbf{D}_i,\psi,\mu_{-i},\nu_i) \propto \sum_{j\neq i} q_j^1 \delta^1_{\mu_j} + c_1 \cdot g_{01}(\mu_i) \cdot p(\mathbf{D}_i|\psi,\nu_i,\mu_i), \tag{4.6}$$

$$p(\nu_i|\mathbf{D}_i,\psi,\nu_{-i},\mu_i) \propto \sum_{j\neq i} q_j^2 \delta^2_{\nu_j} + c_2 \cdot g_{02}(\nu_i) \cdot p(\mathbf{D}_i|\psi,\mu_i,\nu_i), \tag{4.7}$$

where $p(\mathbf{D}_i|\psi,\nu_i,\mu_i)$ and $p(\mathbf{D}_i|\psi,\mu_i,\nu_i)$ are the sampling distributions of $\mathbf{D}_i$, with $\mathbf{D}_i$ defined as the observed data for the ith patient. δ^1_s and δ^2_s are degenerate distributions with point mass at s, g_{01} and g_{02} are the densities corresponding to the base measures G_{01} and G_{02}, respectively. Lastly, $q_j^1 = p(\mathbf{D}_i|\psi,\nu_i,\mu_j)$ and $q_j^2 = p(\mathbf{D}_i|\psi,\mu_i,\nu_j)$ for $j = 1,...,i-1,i+1,...,n$. Note that as $c_k \longrightarrow \infty$, $G_k \longrightarrow G_{0k}(\cdot)$ for $k = 1, 2$, so that the joint model with the mixture of Dirichlet processes converges to the proposed joint model (fully parametric) defined in Section 2. The algorithm for sampling the full set of parameters under the mixture of Dirichlet processes is provided in Appendix F of the Supplementary material available at *Biostatistics* online.

4.2. *Example application*

We demonstrate the nonparametric approach using the same data generated in Section 3. The simulation settings (e.g., ρ, $\delta_{1\cdot}$, etc.) were taken to be the same as in Section 3, and we only considered scenarios where the treatment effects on both the recurrent and terminating event processes equaled the hypothesized level (i.e., $\gamma_r = \gamma_\lambda = -0.3$) for power calculations and no effect on either event process (i.e., $\gamma_r = \gamma_\lambda = 0$) for type I error estimation. We took the confidence parameters $c_1 = c_2 = c$ and varied its value $c \in \{10, 50, 100, 500, 1000, \infty\}$. Results based on $c < 10$ were not presented here because the sampled values of both frailties were highly discrete for such small c's. More details and discussion about using smaller values of c can be found in Appendix E of the Supplementary material available at *Biostatistics* online. For a fair comparison, we defined both base measures G_{01} and G_{02} using the variance parameters estimated by fitting the fully parametric joint model to the colorectal data. The estimated parameter values can be found in Table S3 in Appendix C of the Supplementary material available at *Biostatistics* online. We considered $v = 400$ and due to the computational burden, a total of 1000 simulated trials were performed to estimate

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

the operating characteristics for each choice of sampling prior. The estimated Bayesian type I error and power are shown in Table 2, as well as the estimates based on the proposed (parametric) joint model.

According to Table 2, the type I error and power estimated when $c = \infty$ are highly similar to the estimates based on the proposed parametric joint model, an intuitive result given that the joint model with the mixture of Dirichlet processes converges to a parametric joint model with known frailty distributions as $c \longrightarrow \infty$. Among the results based on the nonparametric frailty joint model, for a certain choice of π, the power estimates increase as c decreases to 100 and then remain approximately constant. These similar estimates across small c's are also attributable to the fact that sampled values of both frailties are highly discrete.

5. Discussion

In Section 3 for the colorectal cancer study, we compared the performance of the proposed approach to a frequentist approach that treated both events as coprimary. An alternative frequentist approach would be to consider the events as multiple primary endpoints with an appropriate multiplicity correction. The hypothesis test then becomes whether the IP shows a beneficial effect on either one of the recurrent and terminating event processes (regardless of whether there is a harmful effect on the other). The proposed Bayesian approach could be reformulated for this setting but is beyond the scope of this article. Furthermore, an IP that provides benefit with respect to one event process but harm on the other is not easily interpretable as beneficial. Extending the proposed framework to allow for multiple primary endpoints of both Bayesian and frequentist approaches is a potential topic for future research for the authors

For the sampling priors used in the example application, we used point-mass sampling prior distributions based on parameter estimates from an analysis of the colorectal cancer data. Given a choice of $(\delta_r, \delta_\lambda)$, any combination of $(\gamma_r, \gamma_\lambda)$ that determines a set of $(e^{\gamma_r}, e^{\gamma_\lambda})$ that satisfies $(e^{\gamma_r} > \delta_r$ or $e^{\gamma_\lambda} > \delta_\lambda)$ or $(e^{\gamma_r} > \delta_r, e^{\gamma_\lambda} = \delta_\lambda)$ or $(e^{\gamma_r} = \delta_r, e^{\gamma_\lambda} > \delta_\lambda)$ or $(e^{\gamma_r} = \delta_r, e^{\gamma_\lambda} = \delta_\lambda)$, which corresponds to hypotheses H_1, H_2, H_3, and H_4, will define a null sampling prior and any combination that determines $(e^{\gamma_r} < \delta_r, e^{\gamma_\lambda} = \delta_\lambda)$ or $(e^{\gamma_r} = \delta_r, e^{\gamma_\lambda} < \delta_\lambda)$ or $(e^{\gamma_r} < \delta_r, e^{\gamma_\lambda} < \delta_\lambda)$, which corresponds to hypotheses H_5, H_6, and H_7, will define an alternative sampling prior. For example, given $(\delta_r = \delta_\lambda = 1)$, $(\gamma_r = \gamma_\lambda = 0)$ defines a null sampling prior and $(\gamma_r = \gamma_\lambda = -0.3)$ defines an alternative sampling prior. General advice for how to choose the point-mass sampling priors in the joint-modeling setting is also given in Xu *and others* (2020). More generally, the Bayesian framework for power and type I error evaluation is applicable for nondegenerate sampling priors on the parameters as well. For more extensive discussion on the use of nondegenerate sampling priors for computing Bayesian power and type I error rates, we refer the readers to the recent work of Psioda and Ibrahim (2018, 2019) and the references cited therein.

In Section 2.5, we proposed $\pi_r = \pi_\lambda = \pi = \frac{1}{3}$ to specify the prior distributions for treatment effects on the two event processes which help to indirectly elicit the prior probabilities for *Basis* models and the prior probabilities for the respective hypotheses. This choice was proposed as the default when there is little information suggesting a more appropriate choice for π_r and π_λ. Recall that the alternative space for the proposed approach includes (i) parameter values consistent with IP benefit on *both* event processes and (ii) parameter values consistent with IP benefit on one event process and no effect on the other. If case (i) is believed to be most likely, comparatively smaller values of π are suggested. In contrast, if case (ii) is viewed as equally likely, comparative larger values of π are suggested. Our choice to take $\pi = \frac{1}{3}$ reflects a compromise between these two scenarios and reflects *a priori* uncertainty regarding on which event process the IP will have an effect in case (2). Table 1 presents the estimated operating characteristics with $\pi_r = \pi_\lambda = \pi \in \{0.5, 1/3, 0.05\}$ under hypotheses specified on the boundary (i.e., IP has no effect on at least one of the event processes) and off the boundary (e.g., IP has favorable effects on both event processes). Through the power for scenarios where there is only an IP effect on one event process (e.g., rows 2–5) one can see that taking relatively larger values of π produces greater power.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

Table 1. *Type I error rate and power estimates with different choices of* π

γ_r	γ_λ	$\pi = 0.5$	$\pi = 1/3$	$\pi = 0.05$
0	0	<0.01	<0.01	0.01
0	−0.3	0.13	0.08	0.06
0	−0.6	0.49	0.21	0.08
−0.3	0	0.18	0.09	0.05
−0.6	0	0.27	0.12	0.06
−0.3	−0.3	0.67	0.73	0.79

γ_r, Treatment effect on recurrent event process.
γ_λ, Treatment effect on terminating event process.

Table 2. *Bayesian type I error rate and power estimates with different values of c*

	$\gamma_r = \gamma_\lambda = 0$			$\gamma_r = \gamma_\lambda = -0.3$		
c	$\pi = 0.5$	$\pi = 1/3$	$\pi = 0.05$	$\pi = 0.5$	$\pi = 1/3$	$\pi = 0.05$
Parm	<0.01	<0.01	0.01	0.67	0.73	0.79
∞	<0.01	<0.01	0.01	0.66	0.72	0.79
1000	<0.01	<0.01	0.01	0.68	0.73	0.79
500	<0.01	<0.01	0.01	0.71	0.74	0.80
100	<0.01	<0.01	0.01	0.73	0.75	0.79
50	<0.01	<0.01	0.01	0.74	0.75	0.79
10	<0.01	<0.01	0.01	0.74	0.75	0.78

Parm, Parametric frailty distribution from Section 2.

Lastly, we considered the case where $\pi_r = \pi_\lambda$ based on the assumption that, for case (2), it is not likely that it would be known upon which of the two event processes the IP is likely to have an effect. If such information were available, using that event to define the primary outcome, rather than having multiple primary outcomes, would seem to be the most sensible approach.

In Section 4, we extended our proposed model to a joint model with mixture of Dirichlet processes so the frailties can be modeled in a relatively flexible way. Results show that the designs based on joint models with parametric/nonparametric frailty distributions produce comparable Bayesian type I error and power estimates. In practice, one can always choose a less restrictive model (e.g., joint model with nonparametric frailty distribution) for analysis purposes. However, the proposed approach with parametric frailties serves as a good fit, particularly for design purposes, since it provides comparable estimates but saves substantially more computational time.

When studying the recurrent and terminating event processes, there are other approaches to model the two event processes jointly as described in Section 1. For example, van den Boom *and others* (2021) recently proposed a model in which the number of recurrences before termination is a random variable and conditionally on this number, a joint distribution for recurrence and survival was specified. However, our focus here is on a design strategy and comparisons with other joint models could be a topic for future research but is beyond the scope of this article.

Supplementary material

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

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

ACKNOWLEDGMENTS

The authors wish to thank the editor, associate editor, and referees for helpful comments and suggestions, which have led to improvements of this article. A GitHub repository (https://github.com/JiaweiXu-code/Joint-Models2) contains the programs and other resources needed to reproduce the results presented in this article.

Conflict of Interest: None declared.

REFERENCES

BROWN, B. W., HERSON, J., ATKINSON, E. N. AND ROZELL, M. E. (1987). Projection from previous studies: a Bayesian and frequentist compromise. *Controlled Clinical Trials* **8**, 29–44.

CHEN, M.-H. (1994). Importance-weighted marginal Bayesian posterior density estimation. *Journal of the American Statistical Association* **89**, 818–824.

CHEN, M.-H., IBRAHIM, J. G., ZENG, D., HU, K. AND JIA, C. (2014). Bayesian design of superiority clinical trials for recurrent events data with applications to bleeding and transfusion events in myelodyplastic syndrome. *Biometrics* **70**, 1003–1013.

CHEN, M.-H. AND SHAO, Q.-M. (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics* **25**, 1563–1594.

CONLON, A. S. C., TAYLOR, J. M. G. AND SARGENT, D. J. (2014). Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in Medicine* **33**, 1750–1766.

COOK, R. J. AND LAWLESS, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16**, 911–924.

DUNCAN, P. W., BUSHNELL, C. D., ROSAMOND, W. D., JONES BERKELEY, S. B., GESELL, S. B., D'AGOSTINO, R. B., AMBROSIUS, W. T., BARTON-PERCIVAL, B., BETTGER, J. P., COLEMAN, S. W. *and others*. (2017). The Comprehensive Post-Acute Stroke Services (COMPASS) study: design and methods for a cluster-randomized pragmatic trial. *BMC Neurology* **17**, 133.

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

FOOD AND DRUG ADMINISTRATION. (2017). *Multiple Endpoints in Clinical Trials Guidance for Industry*. Food and Drug Administration.

GHOSH, D. AND LIN, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554–562.

GHOSH, D. AND LIN, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663–688.

GONZÁLEZ, J. R., FERNANDEZ, E., MORENO, V., RIBES, J., PERIS, M., NAVARRO, M., CAMBRAY, M. AND BORRÀS, J. M. (2005). Sex differences in hospital readmission among colorectal cancer patients. *Journal of Epidemiology and Community Health* **59**, 506 LP – 511.

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

LANCASTER, T. AND INTRATOR, O. (1998). Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93**, 46–53.

LI, Q. H. AND LAGAKOS, S. W. (1997). Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16**, 925–940.

LIU, L., WOLFE, R. AND HUANG, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747–756.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022

O'Hagan, A. and Stevens, J. W. (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Medical Decision Making* **21**, 219–230.

Paulon, G., De Iorio, M., Guglielmi, A. and Ieva, F. (2020). Joint modeling of recurrent events and survival: a Bayesian non-parametric approach. *Biostatistics* **21**, 1–14.

Psioda, M. A. and Ibrahim, J. G. (2018). Bayesian design of a survival trial with a cured fraction using historical data. *Statistics in Medicine* **37**, 3814–3831.

Psioda, M. A. and Ibrahim, J. G. (2019). Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics* **20**, 400–415.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics (Oxford, England)* **8**, 708–721.

Schmoor, C., Schumacher, M., Finke, J. and Beyersmann, J. (2013). Competing risks and multistate models. *Clinical Cancer Research* **19**, 12 LP – 21.

van den Boom, W., De Iorio, M. and Tallarita, M. (2021). Bayesian inference on the number of recurrent events: a joint model of recurrence and survival. *Statistical Methods in Medical Research* **31**, 139–153.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.

Xu, J., Psioda, M. A. and Ibrahim, J. G. (2020). Bayesian design of clinical trials using joint models for longitudinal and time-to-event data. *Biostatistics* **23**, 591–608.

[*Received June 10, 2021; revised June 19, 2022; accepted for publication June 22, 2022*]

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac025/6646084 by University of North Carolina - Chapel Hill Libraries user on 13 December 2022