Lifetime Data Analysis
https://doi.org/10.1007/s10985-022-09581-5

Bayesian Design of Clinical Trials Using Joint Cure Rate Models for Longitudinal and Time-to-Event Data

Jiawei Xu[1] · **Matthew A. Psioda**[1] · **Joseph G. Ibrahim**[1]

Received: 4 November 2021 / Accepted: 24 October 2022

Abstract

For clinical trial design and analysis, there has been extensive work related to using joint models for longitudinal and time-to-event data without a cure fraction (i.e., when all patients are at risk for the event of interest), but comparatively little treatment has been given to design and analysis of clinical trials using joint models that incorporate a cure fraction. In this paper, we develop a Bayesian clinical trial design methodology focused on evaluating the treatment's effect on a time-to-event endpoint using a promotion time cure rate model, where the longitudinal process is incorporated into the hazard model for the promotion times. A piecewise linear hazard model for the period after assessment of the longitudinal measure ends is proposed as an alternative to extrapolating the longitudinal trajectory. This may be advantageous in scenarios where the period of time from the end of longitudinal measurements until the end of observation is substantial. Inference for the time-to-event endpoint is based on a novel estimand which combines the treatment's effect on the probability of cure and its effect on the promotion time distribution, mediated by the longitudinal outcome. We propose an approach for sample size determination such that the design has a high power and a well-controlled type I error rate with both operating characteristics defined from a Bayesian perspective. We demonstrate the methodology by designing a breast cancer clinical trial with a primary time-to-event endpoint where longitudinal outcomes are measured periodically during follow up.

Keywords Bayesian design · Clinical trials · Joint cure rate models · Longitudinal outcomes · Sampling prior

✉ Jiawei Xu
jiawei@live.unc.edu

Matthew A. Psioda
matt_psioda@unc.edu

Joseph G. Ibrahim
ibrahim@bios.unc.edu

[1] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Published online: 10 November 2022

1 Introduction

In clinical trials with time-to-event endpoints, it is often observed that some proportion of patients never experience the event of interest or the event will not occur for an exceedingly long period. These patients are considered *cured* from disease. For example, a radiation therapy can cure patients if it kills all the tumor cells (Withers et al. 1995). In such cases patients will have indefinite relapse-free survival (RFS), i.e., an indefinite period with no signs or symptoms of the cancer. The proportion of cured patients is referred to as the *cure rate* or *cure fraction*. Survival models accommodating a cure fraction, known as cure rate models, have been developed for this setting. A mixture cure rate model was proposed by Berkson and Gage (1952), which assumed a certain fraction of the population is cured. Farewell (1982) used a binary distribution to model the cure probability and a parametric failure time distribution to model the event times. Chen et al. (1999) developed a parametric promotion time cure rate model with a proportional hazards structure. A semiparametric extension was developed by Chen et al. (2002).

The cure rate model has been used to study time-to-event data for various types of cancers, including breast cancer (Woods et al. 2009), leukemia (Tsodikov et al. 1998) and melanoma (Kirkwood et al. 2000). In such trials, outcomes associated with the time-to-event, such as quality of life (QOL) scores or immune response measures (e.g., CD4 counts), are often measured intermittently at potentially different times with a potentially different number of measurements for each patient (Wulfsohn and Tsiatis 1997). The use of joint models for clinical trials with longitudinal and time-to-event data has then become popular. See for example work of De Gruttola and Tu (1994), Wulfsohn and Tsiatis (1997), Xu and Zeger (2001) and Chi and Ibrahim (2006). There are relatively few approaches to study longitudinal and time-to-event data in the presence of a cure fraction. Brown and Ibrahim (2003) developed a joint cure rate model based on a mixture distribution with a point mass at zero for cases where the longitudinal outcome has an excess of zeros (e.g., in cases where some participants exhibit no immune response to an immunotherapy). Chen et al. (2004) also proposed a joint cure rate model in the context of a cancer vaccine study that allows for incorporating multiple longitudinal markers. Kim et al. (2013) proposed a class of transformed promotion time cure models to account for long-term plateaus and conducted inference based on nonparametric maximum likelihood estimation. However, all approaches mentioned above were developed for analysis, with virtually no treatment given to how such models could be used for clinical trial design.

In this paper, we develop a Bayesian design framework using a joint cure rate model for clinical trials involving longitudinal and time-to-event data in a general context. For ease of exposition, we focus on the design of a parallel two-group randomized, controlled trial. We assume the primary endpoint is a time-to-event endpoint (e.g., relapse-free survival) and that a single, continuous longitudinal outcome (e.g., quality of life) is measured repeatedly during the follow up period, and that it potentially provides predictive or prognostic information about the time-to-event endpoint. We propose the use of a piecewise linear time trajectory that incorporates patient-specific random effects to account for patient-level heterogeneity which impacts the longitudinal outcome distribution and the distribution for the promotion times (i.e., the event

times for those uncured). The proposed joint cure rate model allows the treatment to have effects on both the probability of cure and the promotion time distribution, where the effect on the promotion time distribution is characterized by the treatment's effect on the longitudinal outcome and the longitudinal outcome's effect on the promotion times. After collection of longitudinal measures ceases, the trajectory is often assumed to maintain its functional form and is therefore extrapolated until the end of follow-up, which may sometimes be problematic. Conditional on the promotion time hazard at the time at which the measurement of longitudinal outcomes ends, we develop a method to model the hazard as a piecewise linear function of time so that the overall hazard model can accommodate potential changes in the trajectory function after the end of longitudinal assessment (i.e., for the period over which there is no data to directly estimate it). This approach provides added robustness for analysis of the clinical trial data even if a simplified model (i.e., one with extrapolation) is used for sample size determination.

We develop a simulation-based approach whereby one can identify the necessary sample size required to obtain a desired level of Bayesian power while controlling a Bayesian type I error rate. Bayesian (i.e., average) 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). For the special case where the sampling priors place a point-mass on a fixed value of the model parameters, which is our focus in this paper, the Bayesian type I error rate and power for a design closely align with the frequentist versions. We evaluate the operating characteristics of designs based on the proposed joint cure rate model, a semiparametric cure rate model, and a joint model without a cure fraction.

The rest of this paper is organized as follows: In Sect. 2, we introduce a joint cure rate model, construct the superiority hypothesis test, and discuss the use of an approach that avoids extrapolation of the trajectory function after the measurement of longitudinal outcomes ends. We also develop the study design and Bayesian sample size determination strategy. In Sect. 3, we provide a simulation study comparing the Bayesian design based on the proposed joint cure rate model to designs based on a semiparametric cure rate model and a joint model without a cure fraction. We close the paper with some discussion in Sect. 4.

2 Method

2.1 Joint Cure Rate Models

Let $y_i(t)$ be the longitudinal outcome at time t for patient i where $y_i(t) = \mu_i(t) + \epsilon_i(t)$ with $\epsilon_i(t) \sim N(0, \sigma^2)$. We refer to the conditional expectation of $y_i(t)$, denoted by $\mu_i(t)$, as the *longitudinal process* and consider a design model based on a trajectory function given as follows:

$$\begin{aligned}\mu_i(t) &= E[y_i(t)|\boldsymbol{\theta}_i] = \boldsymbol{g}(t)^T\boldsymbol{\theta}_i + \boldsymbol{X}_i(t)\boldsymbol{\gamma} \\ &= \boldsymbol{g}(t)^T\boldsymbol{\theta}_i + \boldsymbol{g}(t)^T\boldsymbol{\gamma}_t + x_i\boldsymbol{g}(t)^T\boldsymbol{\gamma}_x + \boldsymbol{z}_i^T\boldsymbol{\gamma}_z, \end{aligned} \tag{1}$$

where $\boldsymbol{g}(t)$ is a function of time t, $\boldsymbol{\theta}_i \sim N(\mathbf{0}, \Sigma_\theta)$ is a mean zero random effect with positive definite covariance matrix Σ_θ, $\boldsymbol{X}_i(t) = [\boldsymbol{g}(t)^T, x_i\boldsymbol{g}(t)^T, \boldsymbol{z}_i^T]$ represents the covariate process with treatment indicator x_i and baseline covariate vector $\boldsymbol{z}_i$, and $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_t, \boldsymbol{\gamma}_x, \boldsymbol{\gamma}_z]$ is a vector of regression coefficients with $\boldsymbol{\gamma}_t$, $\boldsymbol{\gamma}_x$, and $\boldsymbol{\gamma}_z$ corresponding to $\boldsymbol{g}(t)$, $x_i\boldsymbol{g}(t)$, and $\boldsymbol{z}_i$, respectively. On the basis of randomization and assuming that the longitudinal measurement at time $t = 0$ is pre-treatment, the main treatment effect can be set to 0 and excluded from the model. In our development, we keep the main effect simply for ease of exposition. More details about the trajectory function and the importance of including an interaction term between the time covariate and treatment indicator can be found in Xu et al. (2020).

Using similar notation as in Ibrahim et al. (2001), for the ith patient, let N_i be the unobserved number of metastasis competent tumor cells which follows a Poisson distribution with mean η_i. Define P_{ij}, $j = 1, \ldots, N_i$ to be the time to tumor progression (i.e., time to a detectable tumor arising from the individual cell) for the jth metastasis competent cell in the ith patient. The P_{ij} are referred as *promotion times*. Conditional on N_i, the P_{ij} are assumed to be independent and follow a model with hazard $\lambda_i(t)$ and corresponding distribution function $F_i(t) = 1 - S_i(t)$. The time to an event is then defined as $Y_i = \min\{P_{ij}, 0 \le j \le N_i\}$ with $P(P_{i0} = \infty) = 1$. The marginal probability of survival past time y for the ith patient is given as

$$\begin{aligned} S_{ip}(y) &= P(N_i = 0) + P(Y_i > y | N_i \ge 1) P(N_i \ge 1) \\ &= \exp(-\eta_i) + \sum_{k=1}^{\infty} S_i(y)^k \frac{e^{-\eta_i}\eta_i^k}{k!} \\ &= \exp(-\eta_i + \eta_i S_i(y)). \end{aligned} \tag{2}$$

Since $S_{ip}(\infty) \equiv P(N_i = 0) = \exp(-\eta_i) \neq 0$, (2) is not a proper survival function. The probability of survival beyond time y for an uncured individual is given by $S_i^*(y) = P(Y_i > y | N_i \ge 1) = \frac{S_{ip}(y) - \exp(-\eta_i)}{1 - \exp(-\eta_i)}$, which is a proper survival function. Model (2) can be rewritten as $S_{ip}(y) = \exp(-\eta_i) + \{1 - \exp(-\eta_i)\} S_i^*(y)$ and therefore it is related to the standard mixture cure rate model (Berkson and Gage 1952). The sub-density and sub-hazard function for the ith patient take the forms of $f_{ip}(y) = \eta_i f_i(y) \exp(-\eta_i F_i(y))$ and $h_{ip}(y) = \eta_i f_i(y)$, respectively.

Because treatment may work on the tumor cells either to eliminate them or to prevent their proliferation, as described in Brown and Ibrahim (2003), we model the treatment effect as occurring through two possible mechanisms: (i) impacting the number of metastasis competent tumor cells and (ii) impacting the promotion time distribution for those cells. For mechanism (i), we propose a hazard model given by

$$\log \lambda_i(t) = \log \lambda_0(t) + \beta \left\{ \boldsymbol{g}(t)^T \boldsymbol{\theta}_i + \boldsymbol{g}(t)^T \boldsymbol{\gamma}_t + x_i \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x \right\} \tag{3}$$

with the corresponding cure faction model given by

$$\eta_i = \exp\{\psi_x x_i + \psi_z z_i\}, \tag{4}$$

where $\psi = \{\psi_x, \psi_z\}$ are effects of the treatment and baseline covariates on the probability of cure. For mechanism (ii), we propose a hazard model given by

$$\log \lambda_i(t) = \log \lambda_0(t) + \beta \left\{ \boldsymbol{g}(t)^T \boldsymbol{\theta}_i + \boldsymbol{g}(t)^T \boldsymbol{\gamma}_t + x_i \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x \right\} + \alpha_x x_i + \alpha_z z_i \quad (5)$$

with the corresponding cure faction model given by

$$\eta_i = \eta_0, \quad (6)$$

where $\alpha = \{\alpha_x, \alpha_z\}$ are analogous to the direct effect of the treatment and baseline covariates as defined in Xu et al. (2020) for joint models without a cure faction. For both mechanistic models, β is an association parameter that controls the influence of the longitudinal process on the promotion time hazard and $\boldsymbol{\lambda} = (\lambda_{01}, \ldots, \lambda_{0K})^T$ is a K-component piecewise constant baseline hazard associated with a fixed partition of the time axis. The effects of treatment and baseline covariates are included as "main" effects in either the hazard model or cure fraction model but not in both. Extensive simulation studies (data not shown) support that these two effects are only weakly identifiable relative to one another, and thus it is not necessary to model both.

Let $\boldsymbol{\xi} = (\psi, \alpha, \eta_0, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \beta, \Sigma_\theta)$ denote the complete set of fixed effect parameters and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ denote the collection of random effect vectors for the set of n patients enrolled in the trial. We denote the observed data for the complete set of n patients by $\mathbf{D}$. Suppose patient $i = 1, \ldots, n$ has the longitudinal outcome measured m_i times, denoted by $t_{i1}, \ldots, t_{im_i}$. Let $y_{ij} = y_i(t_{ij})$ and $\mathbf{X}_{ij} = \boldsymbol{X}_i(t_{ij})$ denote the observed longitudinal outcome and covariate process at time t_{ij}, respectively, for $j = 1, \ldots, m_i$. Define $\boldsymbol{N} = (N_1, N_2, \ldots, N_n)$, the complete data likelihood $L(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{N}|\mathbf{D})$ is written as

$$L(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{N}|\mathbf{D}) = \prod_{i=1}^{n} \left[\prod_{j=1}^{m_i} f_L(y_{ij}|\mathbf{X}_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\gamma}, \sigma^2) \right] f_S(t_i, v_i, N_i|x_i, \mathbf{z}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) f_\theta(\boldsymbol{\theta}_i|\Sigma_\theta),$$

where $f_L(\cdot)$ and $f_\theta(\cdot)$ are the distributions of longitudinal outcomes and the random effects, respectively. The joint distribution $f_S(\cdot)$ for the observation time and number of metastasis competent tumor cells has the form

$$f_S(t_i, v_i, N_i|x_i, \mathbf{z}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) = S_i(t_i)^{N_i - v_i} \{N_i f_i(t_i)\}^{v_i} \frac{e^{-\eta_i} \eta_i^{N_i}}{N_i!},$$

where t_i is the observation time and v_i is an indicator for whether an event is observed for the i^{th} patient. We can obtain the likelihood $L(\boldsymbol{\xi}, \boldsymbol{\theta}|\mathbf{D})$ by summing over the latent N_i,

$$L(\boldsymbol{\xi}, \boldsymbol{\theta}|\mathbf{D}) = \prod_{i=1}^{n} \left[\prod_{j=1}^{m_i} f_L(y_{ij}|\mathbf{X}_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\gamma}, \sigma^2) \right] f_S(t_i, v_i|x_i, \mathbf{z}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) f_\theta(\boldsymbol{\theta}_i|\Sigma_\theta)$$

with the density function

$$f_S(t_i, \delta_i | x_i, \mathbf{z}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) = \{\eta_i f_i(t_i)\}^{v_i} \exp\{-\eta_i F_i(t_i)\}.$$

Further integrating over the random effects $\boldsymbol{\theta}_i$ for $i = 1, \ldots, n$ gives the observed data likelihood

$$\begin{aligned} L(\boldsymbol{\xi}|\mathbf{D}) &= \prod_{i=1}^{n} \int_{\theta_i} \left[\prod_{j=1}^{m_i} f_L(y_{ij}|\mathbf{X}_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\gamma}, \sigma^2) \right] f_S(t_i, v_i|x_i, \mathbf{z}_i, \boldsymbol{\theta}_i, \boldsymbol{\xi}) f_\theta(\boldsymbol{\theta}_i|\Sigma_\theta) d\theta_i \\ &= \prod_{i=1}^{n} \int_{\theta_i} \left[\prod_{j=1}^{m_i} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2\sigma^2}(y_{ij}-\mu_{ij})^2} \right] \{\eta_i e^{-\int \lambda_i(s)ds} \lambda_i(t_i)\}^{v_i} \\ &\quad \exp\{-\eta_i(1 - e^{-\int \lambda_i(s)ds})\} f_\theta(\boldsymbol{\theta}_i|\Sigma_\theta) d\theta_i \end{aligned}$$

where μ_{ij} and $\lambda_i(\cdot)$ are defined in (1) and (3), respectively. Since μ_{ij}, $\lambda_i(\cdot)$ and $f_\theta(\cdot)$ all include random effects, the observed data likelihood does not exist in closed form.

2.2 Piecewise Linear Time Trajectory Function

For the longitudinal process, a common practice is to take $\boldsymbol{g}(t)$ to be a polynomial vector (Chen et al. 2011; Crowther et al. 2013). For our approach, following Xu et al. (2020), we assume a continuous, semiparametric, piecewise linear time trajectory function for $\boldsymbol{g}(t)$, which is constructed using a pre-specified number of segments with specified knots. This approach is advantageous because it allows for a more flexible shape for the time trajectory (e.g., a trajectory that levels off at some time point).

We assume a piecewise linear trajectory function with M components and $M-1$ knots denoted by $k = (k_1, \ldots, k_{M-1})$ which satisfy $k_0 = 0 < k_1 < \cdots < k_{M-1} < k_M = \infty$. Define the $M+1$ dimensional vector $\boldsymbol{g}(t)$ to have 1 as its first component and $f_m(t) = \max\{\min\{t, k_m\} - k_{m-1}, 0\}$ as its $(m+1)^{\text{th}}$ component for $m = 1, \ldots, M$ (i.e., $\boldsymbol{g}(t)^T = [1, f_1(t), \ldots, f_M(t)]$). Thus, the first components of $\boldsymbol{\theta}_i$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\gamma}_x$ correspond to intercept parameters and the remaining M components of each vector combine to determine each trajectory's slope over the M time intervals in the time axis partition.

2.3 Treatment Effects on the Cure Fraction and Promotion Times

The proposed joint cure rate model allows the treatment to have effects on the probability of cure and/or the promotion time distribution. For models (3) and (4), we refer to ψ_x as the effect on the probability of cure. The effect on the promotion time hazard is $\beta \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x$ which can be decomposed as the effect of treatment on the longitudinal outcome $\boldsymbol{g}(t)^T \boldsymbol{\gamma}_x$ multiplied by the effect of the longitudinal outcome on the promotion time hazard β. Figure 1 provides a simple illustration for how treatment affects the time-to-event endpoint. There are several possibilities for how the treatment could

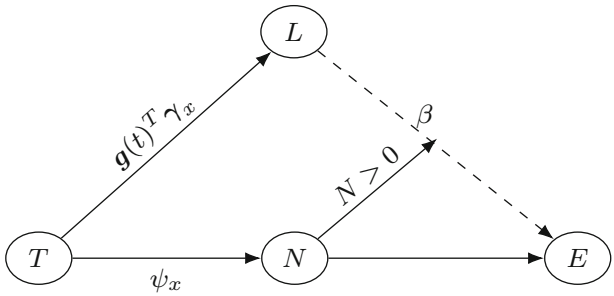

Fig. 1 Diagram for the treatment effect on the time-to-event outcome. T = Treatment effect; L = Longitudinal outcome; E = Time-to-event endpoint; N = Number of metastasis competent tumor cells. The dashed line reflects a pathway that is only open conditional on $N > 0$

impact the time-to-event distribution. The treatment could have an effect only on the probability of cure ($\psi_x \neq 0, \beta \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x = 0$), only on the promotion time distribution ($\psi_x = 0, \beta \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x \neq 0$), or both. The effects in models (5) and (6), however, are only on the promotion time distribution and are analogous to the direct/indirect effects defined by Xu et al. (2020). We refer the reader to the paper for more details.

The proposed formulation of the joint cure rate model is ideal for cases where effects on both the probability of cure and on the promotion time distribution for those uncured are plausible and it is of interest to design a trial with a goal of identifying their contribution to the total treatment effect.

2.4 Hypothesis Testing

We consider a study design where the goal is to demonstrate superiority of an investigational product to a control with respect to a time-to-event endpoint such as relapse-free survival (RFS). While it is of interest to estimate the effectiveness of the investigational product, quantification of effects on both the probability of cure and on the promotion time distribution for those uncured is also of interest. We assume that the longitudinal outcome is measured at baseline and follow-up time points scheduled at regular intervals until some fixed time t_{traj}. Patients are followed for the time-to-event endpoint starting at baseline and up to time $t_{\max}$, which corresponds to a random point in time when a specified number of events have accrued in the trial or a maximum duration for the trial has been reached. Let $h_1(t)$ and $h_0(t)$ be the promotion time hazards for comparable patients in the treatment and control groups. Here, *comparable* implies that although the patients receive different treatments, their random effects and baseline covariate vectors are equal (i.e., $h_1(t) = \log \lambda(t, x = 1, z)$ and $h_0(t) = \log \lambda(t, x = 0, z)$, where $\log \lambda(t)$ is defined in (3)).

We define the one-sided null and alternative hypotheses for superiority as: $H_0 : \Delta(t_0, \omega_0) \geq 1$ versus $H_1 : \Delta(t_0, \omega_0) < 1$, where $\Delta(t_0, \omega_0) = \{e^{\psi_x}\}^{\omega_0} \cdot \phi(t_0)^{(1-\omega_0)}$ for fixed t_0 with $t_{traj} \leq t_0 \leq t_{max}$. Further, e^{ψ_x} is the ratio of the probability of cure in the control patients compared to a comparable treated group, and $\phi(t_0) = G^{-1}\left(\int_0^{t_0} G\left(\frac{h_1(t)}{h_0(t)}\right) \Omega(t) dt\right)$ is the average time-varying hazard ratio of the promotion

time distribution where $G(x)$ is a strictly increasing function and $\Omega(t)$ is a non-negative weight function such that $\int_0^{t_0} \Omega(t)dt = 1$. Under models (5) and (6), we have $\psi_x = 0$ and so logically $\omega_0 = 0$ so that $\Delta(t_0, \omega_0) = \phi(t_0)$. For models (3) and (4), $\Delta(t_0, \omega_0)$ is a weighted function of treatment effects on the probability of cure and on the promotion time distribution with ω_0 prespecified between 0 and 1, where taking $\omega_0 = 1$ results in a test of the treatment effect on the probability of cure and taking $\omega_0 = 0$ results in a test of the effect on the promotion times. If one believes the treatment has impact only on the promotion times, ω_0 can be taken close to 0; if treatment affects only the probability of cure, ω_0 should be close to 1. We suggest a default value as $\omega_0 = 0.5$ which results in an estimand that weights the ratio for the probability of cure and the average time-varying hazard ratio equally. More details and discussion about how to choose a value of ω_0 can be found in Appendix A of the Supplementary Materials. For the time-varying hazard ratio $\phi(t_0)$, we take $G(x) = \log(x)$ and $\Omega(t) \propto |\beta \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x|$, where $\beta \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x$ can be viewed as the difference in log hazards for two comparable patients. More details on this choice of weight function can be found in Xu et al. (2020). In addition to the specific choices for $G(x)$ and $\Omega(t)$ discussed above, one can approximate $\phi(t_0)$ for arbitrary choices using a trapezoidal approximation to the integral. We refer the reader to Xu et al. (2020) for more details.

2.5 Avoiding Extrapolation of the Longitudinal Process

In use of joint models, it is often assumed that the longitudinal process maintains the same trajectory functional form after collection of the longitudinal measurements ceases (i.e., a linear trajectory continues to hold). In such cases, the trajectory is effectively extrapolated from the point of the last measurement of the longitudinal outcome until the observation period ends, which may sometimes be problematic. In an ideal setting, longitudinal outcomes will be collected over the period $[0, t_0]$, which avoids extrapolation altogether. However, this solution may not be practical in all settings. Thus, we propose an approach to avoid extrapolation which can be used for analysis of the trial data even if simplified models that use extrapolation are assumed for the purposes of sample size determination.

Similar to the piecewise linear time trajectory function proposed in Sect. 2.2, for the promotion time hazard after the end of longitudinal measurements, we propose a model for the hazard that has a piecewise linear form conditional on the patient-specific hazard at the assessment end time t_{traj}. The promotion time hazard model after t_{traj} takes the form

$$\log h_i(t) = \log \lambda_i(t_{traj}) + \boldsymbol{b}(t)^T \boldsymbol{\rho}_t + x_i \boldsymbol{b}(t)^T \boldsymbol{\rho}_x, \tag{7}$$

where $t > t_{traj}$ and $\boldsymbol{b}(\cdot)$ is equal to $\boldsymbol{g}(\cdot)$ but excluding the intercept. One can see that hazard models (3) and (5) are special cases of model (7) where $\boldsymbol{\rho}_t = \beta \cdot \boldsymbol{\gamma}_t$, $\boldsymbol{\rho}_x = \beta \cdot \boldsymbol{\gamma}_x$ and $\lambda_0(t) = \lambda_i(t_{traj})$. Note that the log hazard at time t_{traj} for an individual implicitly takes into account the value of their random effects. We compare the proposed joint cure rate model with and without extrapolation in Sect. 3.2 to show the robustness of the proposed approach.

2.6 Model Estimation and Posterior Inference

Using the observed data likelihood $L(\boldsymbol{\xi}|\mathbf{D})$, the posterior distribution for the fixed effects takes the form $\pi(\boldsymbol{\xi}|\mathbf{D}) \propto L(\boldsymbol{\xi}|\mathbf{D})\pi^{(f)}(\boldsymbol{\xi})$, where $\pi^{(f)}(\boldsymbol{\xi})$ is the fitting prior (Wang and Gelfand 2002). However, as discussed in Sect. 2.1, the observed data likelihood $L(\boldsymbol{\xi}|\mathbf{D})$ cannot be computed in closed form. Nonetheless, it is straightforward to use Markov chain Monte Carlo (MCMC) methods to sample from $\pi(\boldsymbol{\xi}, \boldsymbol{\theta}|\mathbf{D})$ based on the likelihood $L(\boldsymbol{\xi}, \boldsymbol{\theta}|\mathbf{D})$ and use the samples of $\boldsymbol{\xi}$ to approximate $\pi(\boldsymbol{\xi}|\mathbf{D})$. However, because of the substantial computational burden of using MCMC for large scale simulations, we consider a posterior approximation for inference during design simulations. We use standard software (e.g., the NLMIXED procedure) to obtain an asymptotic posterior approximation for $\pi(\boldsymbol{\xi}|\mathbf{D})$ based on the Bayesian central limit theorem (Chen 1985). Specifically, $\pi(\boldsymbol{\xi}|\mathbf{D}) \approx \text{Normal}(\hat{\boldsymbol{\xi}}, \hat{\Sigma}_{\boldsymbol{\xi}})$, where $\hat{\boldsymbol{\xi}}$ and $\hat{\Sigma}_{\boldsymbol{\xi}}$, respectively, are the maximum likelihood estimator (MLE) and approximate asymptotic covariance for the MLE, obtained by maximizing the approximate observed data likelihood obtained by integrating over the $\boldsymbol{\theta}_i$ using Gaussian Quadrature. Application of the delta method (Doob 1935) yields an approximate posterior for $\pi(\Delta(t_0, \omega_0)|\mathbf{D}) \approx \text{Normal}(\hat{\Delta}, \hat{\sigma}^2_{\Delta})$, where $\hat{\Delta}$ and $\hat{\sigma}^2_{\Delta}$ are the maximum likelihood estimator and estimated asymptotic variance of $\Delta(t_0, \omega_0)$, respectively. It follows that $P(\Delta(t_0, \omega_0) < 1|\mathbf{D}) \approx 1 - \Phi\left(\frac{\hat{\Delta}-1}{\hat{\sigma}_{\Delta}}\right)$.

2.7 Bayesian Sample Size Determination

The proposed method is designed to identify the smallest sample size required for a trial, subject to Bayesian type I error rate and power requirements. 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 $\boldsymbol{\xi}$, respectively. The null sampling prior gives non-zero weight to values of $\boldsymbol{\xi}$ such that $\Delta(t_0, \omega_0) \geq 1$ and the alternative sampling prior such that $\Delta(t_0, \omega_0) < 1$. In this paper, we only consider point-mass sampling priors such that $\pi_0^{(s)}(\boldsymbol{\xi}) = 1(\boldsymbol{\xi} = \boldsymbol{\xi}_0)$ and $\pi_1^{(s)}(\boldsymbol{\xi}) = 1(\boldsymbol{\xi} = \boldsymbol{\xi}_1)$, where $1\{A\}$ is an indicator that A is true. When applying the point-mass sampling priors, the Bayesian type I error rate and power closely align with the frequentist versions.

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 $P(\Delta(t_0, \omega_0) < 1|\mathbf{D}) \geq p_0$. For a fixed value of $\boldsymbol{\xi}$, the null hypothesis rejection rate is defined as $r(\boldsymbol{\xi}) = E[1\{P(\Delta(t_0, \omega_0) < 1|\mathbf{D}) \geq p_0\}|\boldsymbol{\xi}]$. The Bayesian type I error rate and power are defined as $\alpha^{(s)} = E[r(\boldsymbol{\xi})|\pi_0^{(s)}]$ and $1 - \beta^{(s)} = E[r(\boldsymbol{\xi})|\pi_1^{(s)}]$, which are weighted averages of $r(\boldsymbol{\xi})$ with weights determined by $\pi_0^{(s)}(\boldsymbol{\xi})$ and $\pi_1^{(s)}(\boldsymbol{\xi})$, respectively.

2.8 Simulation Based Sample Size Determination

We use simulations to identify the required sample size such that the design has sufficiently high Bayesian power. The number of patients enrolled in the trial are chosen

to obtain a specified number of events in a specified interval of time *on average*. Let the sample size and number of events be given by n and v, respectively. Following Xu et al. (2020), we consider an approach that fixes the ratio $r = \frac{n}{v}$ but varies the number of events. If n_1 patients result in obtaining v_1 events in a specific time frame, then to obtain $v_2 \geq v_1$ events in the same time frame, one should increase n_2 proportionally. The trial duration t_{max} must be taken to be large enough so that there is a high probability that the observed data will allow for estimation of the location of the plateau in the survival curves. As such, taking n to be very large, even if cost feasible, is disadvantageous. In our simulations, n was chosen as described above.

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

S1. Let $v_1, \ldots, v_K$ denote the ordered potential 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.

Under point-mass null sampling priors along with a non-informative fitting prior, by taking $p_0 = 1 - \alpha^{(s)}$, the Bayesian type I error rate will be approximately $\alpha^{(s)}$ and therefore no specific effort is needed to control the Bayesian type I error rate at $\alpha^{(s)}$. Nonetheless, one can always compute the exact Bayesian type I error rate through simulation to ensure it is sufficiently close to the desired nominal level. The simulations studies presented in Sect. 3 illustrate that the property $\alpha^{(s)} \approx 1 - p_0$ indeed holds quite well.

We expand step S3 from the algorithm given above. Letting B be the number of simulation studies to be performed, to estimate the Bayesian power $1 - \beta_k^{(s)}$ associated with event total v_k, one does the following:

S3.1. Sample $\boldsymbol{\xi}^{(b)}$ from the alternative sampling prior $\pi_1^{(s)}(\boldsymbol{\xi})$.
S3.2. Simulate the *observed* data $\mathbf{D}^{(b)}$.
S3.3. Estimate the posterior distribution $\pi(\phi(t_0)|\mathbf{D})$ using an approach described in Sect. 2.6 and compute the null hypothesis rejection indicator

$$r^{(b)} = 1\{P(\Delta(t_0, \omega_0) < 1|\mathbf{D}^{(b)}) \geq p_0\}.$$

S3.4. Approximate the Bayesian power:

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

The new trial data $\mathbf{D}^{(b)}$ can be simulated using the following steps. Let J be the total number of time points at which the longitudinal outcome will be measured. Denote the scheduled time points for longitudinal measurements as $m = (m_1, \ldots, m_J)$. Thus, J

is the number of outcome measurements that will be obtained for patients who do not experience an event or drop out of the trial for other reasons. For patient $i = 1, \ldots, n$, we do the following:

1. Simulate the enrollment time r_i using a chosen enrollment distribution.
2. Simulate x_i based on a treatment assignment distribution (or rule) and $\boldsymbol{z}_i$ based on the covariate distribution.
3. Simulate $\boldsymbol{\theta}_i$ from $N(\mathbf{0}, \Sigma_\theta)$ and ϵ_{ij} from $N(0, \sigma^2)$ for $j = 1, \ldots, J$.
4. Compute $y_{ij} = \mu_i(t_{ij}) + \epsilon_{ij}$.
5. Compute $\eta_i = \exp(x_i\psi_x + z_i^T\psi_z)$ (or $\eta_i = \eta_0$) and simulate N_i from Poisson(η_i).
6. Simulate $P_{ij} \sim F(t)$ independently for $j = 1, \ldots, N_i$ and compute $y_i = \min\{P_{ij}, j = 0, \ldots, N_i\}$ with $P_{i0} = t_{max}$.
7. Simulate the time-to-censorship c_i based on the chosen censorship distribution.
8. Set $s_i = \min(y_i, c_i)$ and $\delta_i = I(y_i \leq c_i)$.

For patient $i = 1, \ldots, n$,

1. Remove any patient with $r_i \geq t_{\max}$ (Patients whose simulated enrollment time occurs after the study terminates).
2. Remove any longitudinal outcome y_{ij} occurring after time $r_i + s_i$.

3 Example Application: Bayesian Clinical Design for Breast Cancer

Our design methodology is motivated by a breast cancer trial undertaken by the International Breast Cancer Study Group (IBCSG) (IBCSG 1996). The trial, IBCSG Trial VI, was conducted in pre-menopausal women with node-positive breast cancer to investigate the efficacy of different durations of adjuvant chemotherapy (3 vs 6 cycles of CMF – cyclophosphamide, methotrexate, and fluorouracil) and whether reintroduction of CMF provided added benefit. Treatment strategies were evaluated with respect to the overall survival and relapse-free survival endpoints. During the study, four measures of quality of life (appetite, mood, coping, physical well-being) were scheduled to be collected at baseline and every three months for up to two years (Hürny et al. 1992).

For our example application, we design a similar trial evaluating two treatments (e.g., whether or not chemotherapy was reintroduced) with respect to a primary RFS time-to-event endpoint. Here as in IBCSG (1996), RFS is defined as the time to either cancer-recurrence or death, whichever occurred first. We consider one quality of life measure (e.g., coping score) in our example. Similar to IBCSG Trial VI, we assume quality of life scores are collected every 3 months starting at baseline for up to 2 years, and the scores are approximately normally distributed (after appropriate transformation, e.g., a square root transformation as was done in Zhang et al. (2016)). We fix t_{max} based on the time when the plateau is expected to occur based on prior studies (Psioda and Ibrahim 2018). The IBCSG trial followed patients for up to 14 years and the RFS curve plateaued after approximately 11 years. Hence our simulations were constructed by limiting the number of patients enrolled so that the number with follow-up at least as long as is expected to see the survival curve plateau was non-negligible. In the simulated trials, patients were randomized to the two treatments using a 1:1

allocation scheme. Patient accrual was simulated to be uniform over a 1 year period and censoring (i.e., dropout) was assumed to follow a mixture distribution whereby patients had a 0.05 probability of dropping out of the trial and, conditional on being a dropout, the time to dropout was simulated to be uniform over the duration. In IBCSG Trial VI, the number of positive lymph nodes identified for lymphadenectomy was prognostic of survival and so we included a binary covariate (> 3 vs ≤ 3 nodes) in our hypothetical trial simulations. The covariate was simulated such that approximately 50% of the patients were in the more severe group.

We assumed the longitudinal process for the ith patient is given by $\mu_i(t) = \theta_i + \boldsymbol{g}(t)^T \boldsymbol{\gamma}_t + x_i \boldsymbol{g}(t)^T \boldsymbol{\gamma}_x + \gamma_z z_i$, where z_i is an indicator that the patient had > 3 positive lymph nodes. We considered a four-component piecewise linear trajectory function with knots at 0.25, 0.75, and 1.25 years. For the promotion time hazard, we simulated data using models (3) and (4) because the treatment (i.e, chemotherapy) is expected to eliminate the metastasis competent tumor cells. A five-component piecewise constant function was assumed for the baseline hazard with knots at times 2.00, 2.82, 3.82, and 5.60 years. Knot placement was determined fitting the proposed joint cure rate model to the IBCSG data such that each time interval of the trajectory function and promotion time hazard had approximately the same number of events, respectively.

For the construction of $\Delta(t_0, \omega_0)$, we took $\omega_0 = 0.5$, which is our proposed default choice and $t_0 = 14$ equal to the approximate expected duration of the trial. We considered various sampling priors for the treatment effects and the association parameter. For the association parameter, we considered point-mass priors with $\beta \in \{-0.15, -0.3\}$. For the treatment effect on the probability of cure, we considered point-mass priors with $\psi_x \in \{0.0, -0.2\}$. Because of the randomized nature of the trial, we only considered $\gamma_{x,0} = 0.0$. For the treatment effect on the longitudinal process, we considered point-mass priors on the four slope parameters $\{\gamma_{x,1}, \ldots, \gamma_{x,4}\}$ including $\{0.0, 0.0, 0.0, 0.0\}$ (no effect), $\{0.2, 0.2, 0.2, 0.2\}$ (increasing treatment effect), and $\{0.4, 0.3, 0.2, 0.1\}$ (benefit but decreasing effectiveness over time). The nuisance parameter values were taken to equal the approximate posterior modes based on our analysis of the IBCSG data as shown in Appendix B of the Supplementary Materials. To identify the desired number of events required to achieve Bayesian power equal to 0.8, we considered $v = 150$ to 300 in increments of 25. The sample size was chosen such that a non-negligible fraction of patients were observed through the period where the plateau is expected. A total of 4000 simulated trials were performed to estimate the operating characteristics for each choice of sampling prior and each sample size considered. Table 2 includes the parameter labels used in this paper, as well as their descriptions. The values for parameters in the second half (starting with σ) can be chosen based on estimates obtained by fitting the proposed models to available data if possible (e.g., IBCSG study data in our case).

We evaluated the performance of the proposed joint cure rate model (JCRM) against a semiparametric cure rate model and a joint model without a cure fraction. We considered two types of scenarios: one where the trajectory in the longitudinal process did not change its functional form after the measurement of the longitudinal outcome ended and another where the trajectory changed its functional form after the assessment ended. Section 3.1 evaluates the performance of the JCRM when the trajectory maintains its functional form. Section 3.2 demonstrates the robustness afforded by the

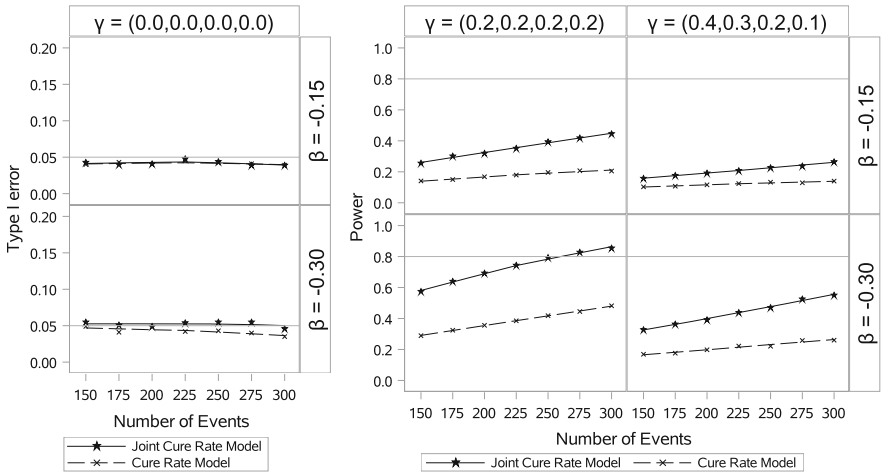

Fig. 2 Estimated type I error and power curves when the treatment effect $\psi_x = 0$

procedure described in Sect. 2.5 to avoid extrapolating the trajectory when in fact it does change its functional form after the longitudinal outcome ceases to be measured.

In Appendix D of the Supplementary Materials, we also evaluated the performance of the proposed JCRM in the scenario where the cure rate is zero (i.e., when the joint model without cure is the correct model). A comparison of the estimated Bayesian type I error and power between the JCRM and the joint model without a cure fraction shows the implementation of JCRM is robust to this type of model misspecification.

3.1 Evaluation of the JCRM When the Trajectory Maintains its Functional Form

In this section, we evaluate the performance of the JCRM against a semiparametric cure rate model and a joint model without a cure fraction. The knot placement for the semiparametric cure rate model and the joint model without a cure fraction is the same as proposed for the JCRM. The estimated type I error rate and power curves based on different approaches are shown in Figs. 2 through 4. Figures 2 and 3 compare the performance of the JCRM to the semiparametric cure rate model. Figure 2 presents the estimated type I error and power curves when there is no treatment effect on the probability of cure (i.e., $\psi_x = 0$). Figure 3 presents the estimated power curves when there is only an effect on the probability of cure (column 1) and are effects on both the probability of cure and the promotion time distribution (columns 2 and 3).

Both approaches have Bayesian type I error controlled at approximately the nominal level (i.e., $\alpha^{(s)} = 0.05$), regardless of the strength of the association parameter (Fig. 2, column 1). When treatment impacts only the probability of cure (Fig. 3, column 1), power estimates based on the cure rate model are higher than those based on the JCRM because the default choice of $\omega_0 = 0.5$ results in an estimand giving a significant amount of weight to a null effect. If the treatment only has an affect on the probability of cure, it is reasonable to fit a cure rate model. However, fitting

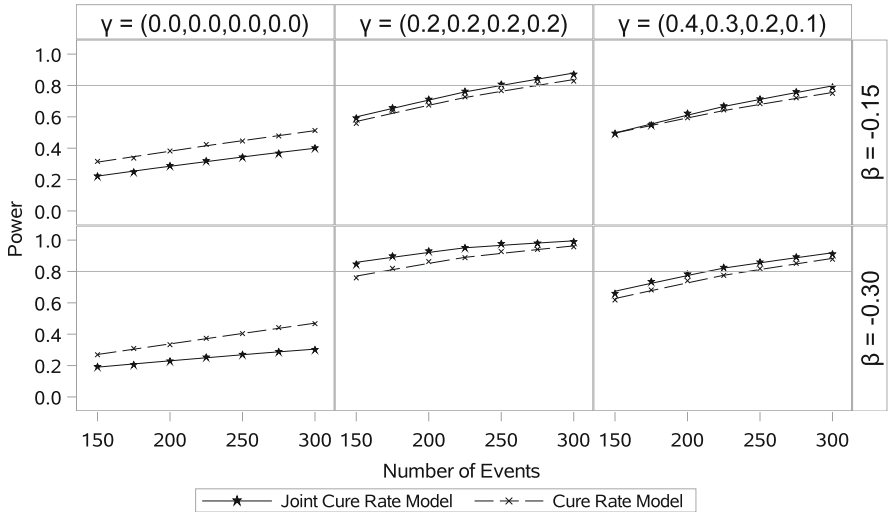

Fig. 3 Estimated power curves when the treatment effect $\psi_x = -0.2$

Table 1 Power estimates based on different values of ω_0

	Joint cure rate model		Cure rate model
	$\omega_0 = 0.5$	$\omega_0 = 1.0$	
$\beta = -0.15$	0.290	0.445	0.382
$\beta = -0.30$	0.229	0.405	0.333

Number of events $v = 200$

the proposed JCRM with $\omega_0 = 1$ leads to similar or even higher power estimates compared to the semiparametric cure rate model as shown in Table 1. Columns 2 and 3 in both Figs. 2 and 3 illustrate that designs based on the JCRM have higher power than those based on the semiparametric cure rate model when there is a moderate or large treatment effect on the promotion time distribution.

When evaluating the performance of the proposed approach, a natural comparison is to consider a joint model for longitudinal and time-to-event data but without a cure fraction. In many cases, the type I error and power estimates based on the joint model without a cure fraction are similar to those based on the JCRM, such as the cases in Figs. 2 and 3. However, this is not true for all scenarios. Figure 4 presents the estimated type I error and power curves based on the proposed JCRM, the joint model without a cure fraction, and the semiparametric cure rate model. The corresponding average fitted population survival curves based on the JCRM and the average fitted survival curves based on the joint model without a cure fraction for the two treatment groups are shown alongside. For scenarios corresponding to the null hypothesis, the hazard for the treated group is set equal to that of the control group. All curves are plotted for the severe node group with patient-specific random effects fixed at 0. We can see that all three approaches provide similar type I error rates with the designs based on the JCRM having higher power than the other two approaches. To further

Table 2 User inputs with descriptions

User inputs	Description
M	Number of linear trajectory components
$k = (k_1, \ldots, k_{M-1})^*$	Knot placement for linear trajectory
K	Number of baseline hazard components
$l = (l_1, \ldots, l_{K-1})$	Knot placement for baseline hazard
J	Number of longitudinal measures
$m = (m_1, \ldots, m_J)$	Scheduled time points for longitudinal measurement
ρ	Dropout probability
t_{traj}	Longitudinal outcome measurement duration
t_{max}	Trial duration
σ	Standard deviation of longitudinal measurement error
Σ_θ	Standard deviation of random intercept
γ_z	Coefficient vector for baseline covariates in longitudinal process
γ_t	Coefficient vector for time intervals in longitudinal process
γ_x	Coefficient vector for time and treatment interactions in longitudinal process
λ	Vector of constant baseline hazards
β	Association parameter
$\psi = (\psi_x, \psi_z)$	Coefficient vector in cure fraction model (treatment group, baseline covariates)

Note that the partition $k_1, \ldots, k_{M-1}$ corresponds to a partition of the entire time axis at least up to time t_{traj}. In the case where assessment of the longitudinal trajectory ends prior to the end of follow-up (e.g., $t_{traj} < t_{max}$) and extrapolation is to be avoided, it will be the case that $t_{traj} = k_j$ for some $1 < j < M-1$. In these cases, the trajectory function $g(t)$ will apply prior to time k_j and the trajectory function follows the form $b(t)$ after. This applies to both the data generation process as well as the fitted model, though there is no requirement that the model for the data generation process is identical to that fit to the data

illustrate instances where the joint model without a cure fraction can perform poorly, in Appendix C of the Supplemental Materials, we compared the JCRM against the joint model without a cure fraction with both models having an exponential baseline hazard instead of piecewise constant baseline hazards. The results presented in that appendix illustrate that, as the joint models become increasingly parametric (e.g., moving from a piecewise exponential to an exponential baseline hazard) the necessity to model the cure fraction becomes increasingly important due to a lack of flexiblity afforded by the joint model without a cure fraction.

3.2 Evaluation of the JCRM When the Trajectory Changes its Functional Form

In this section, we investigate the impact of changes in the functional form of the trajectory after the longitudinal outcome ceases to be measured. We compared the Bayesian type I error rate and power based on fitting a JCRM with extrapolation of the trajectory after the longitudinal measurement ends (at time t_{traj}) to those based on a JCRM without extrapolation using the hazard defined in (7). The trials were

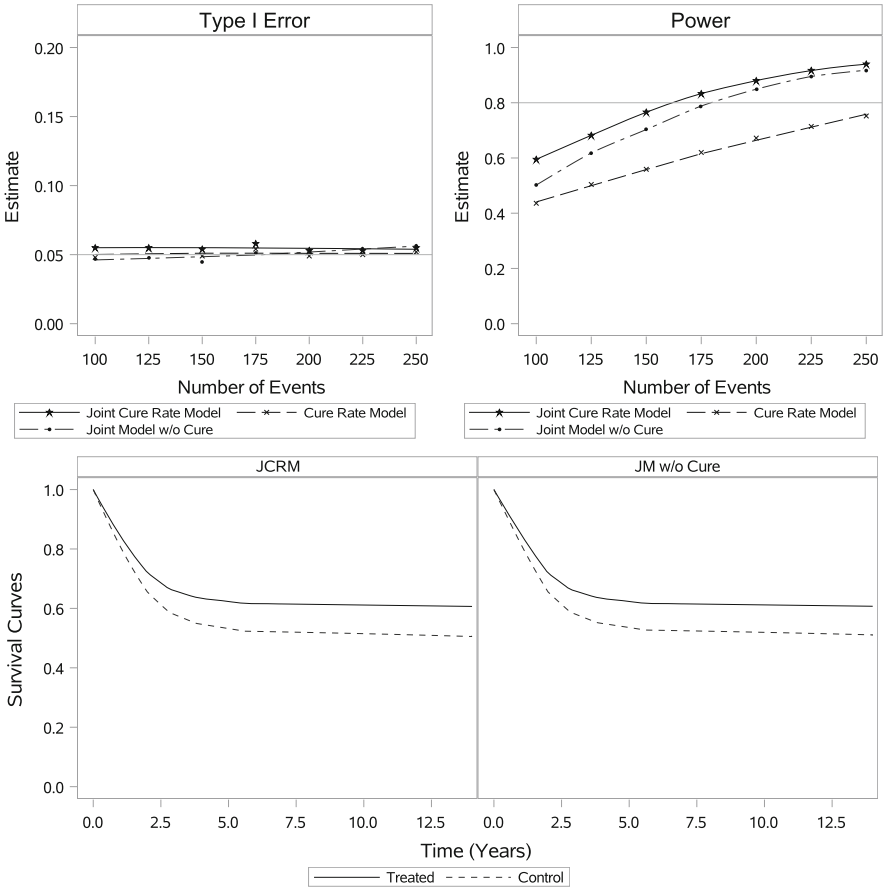

Fig. 4 Estimated type I error and power curves based on the joint cure rate model, the joint model without a cure fraction and a semiparametric cure rate model

simulated such that the trajectory continued to change its functional form until the end of follow up, though collection of longitudinal measures to be used for its estimation ceased at $t_{traj} = 2$ years. For the JCRM with extrapolation, both the longitudinal trajectory and the promotion time hazard functions were constructed based on the same knots proposed in Sect. 3. For the JCRM without extrapolation, the promotion time hazard before t_{traj} was assumed to follow model (3) with the trajectory having the same knots as used for the trajectory in the JCRM with extrapolation. After t_{traj}, the promotion time hazard was modeled as a three-component piecewise linear function with pre-specified knots at times 3.50 and 5.00 years. Figures 5 and 6 present the estimated Bayesian type I error and power curves based on fitting both JCRMs. The corresponding average fitted population survival and population hazard curves for the two treatment groups are shown alongside. For scenarios corresponding to the null hypothesis, the hazard for the treated group is set equal to that of the control group. All curves are plotted for the severe node group with patient-specific random

effects fixed at 0. The parameter values used to simulate the trials for Fig. 5 were taken to equal the approximate posterior modes based on fitting the JCRM without extrapolation to the IBCSG data. The IBCSG trial had the last course of adjuvant chemotherapy given at month 15 and the hazard appears to spike shortly after the intervention period before it decreased. We modified the parameter values estimated from the IBCSG data to exhibit a less extreme spike as well and that scenario is shown in Fig. 6. Figure 5 shows that the type I error rate estimates based on the JCRM with extrapolation were inflated and the power estimates were lower than those based on the JCRM without extrapolation when the event number $v > 200$. From Fig. 6, the type I error and power estimates based on the JCRMs with and without extrapolation are nearly identical, but the JCRM with extrapolation has a population hazard function that is more poorly estimated compared to the estimated population hazard function based on the JCRM without extrapolation, which is approximately unbiased for the "true" curve. In summary, the type I error rate and power are similar for both JCRMs in many cases, though type I error control is generally better without extrapolation. Moreover, the JCRM with extrapolation has notably poorer performance at estimating the underlying population hazard and survival function.

4 Discussion

In the example application, we used point-mass sampling prior distributions based on parameter estimates from an analysis of the IBCSG data. More generally, the Bayesian framework for power and type I error evaluation is applicable for non-degenerate sampling priors on the parameters. For more extensive discussion on the use of non-degenerate sampling priors for computing Bayesian power and type I error rates, we refer the readers to the work of Psioda and Ibrahim (2018, 2019). In a joint modelling setting, even for point-mass sampling priors, choosing the sampling priors can be challenging. The authors would refer the readers to the Supplementary Materials of Xu et al. (2020) for suggestions on how to determine the point-mass null/alternative sampling priors.

The design methodology developed in the paper is for a single, continuous longitudinal outcome. It may be extended to two or more longitudinal outcomes, but the relationships between different longitudinal outcomes and the construction of a treatment's working pathway need to be carefully considered. Extending the proposed framework to allow for multiple types of longitudinal outcomes is a topic for future research for the authors.

In the proposed joint modelling framework, we included the effects of treatment and baseline covariates as "main" effects in either the hazard model or cure fraction model but not in both. Extensive simulation studies (data not shown) confirm that these two effects are only weakly identifiable relative to one another, and thus it is problematic to model both. Approaches for modelling treatment effects in both hazard and cure rate models in the joint model setting is a future topic for the authors.

The baseline hazard in (7) is assumed to be constant over the interval (t_{traj}, t_{max}) but allowed to vary over patients due to the inclusion of random effects. It can be generalized to incorporate a time-varying baseline hazard (e.g., adding a common

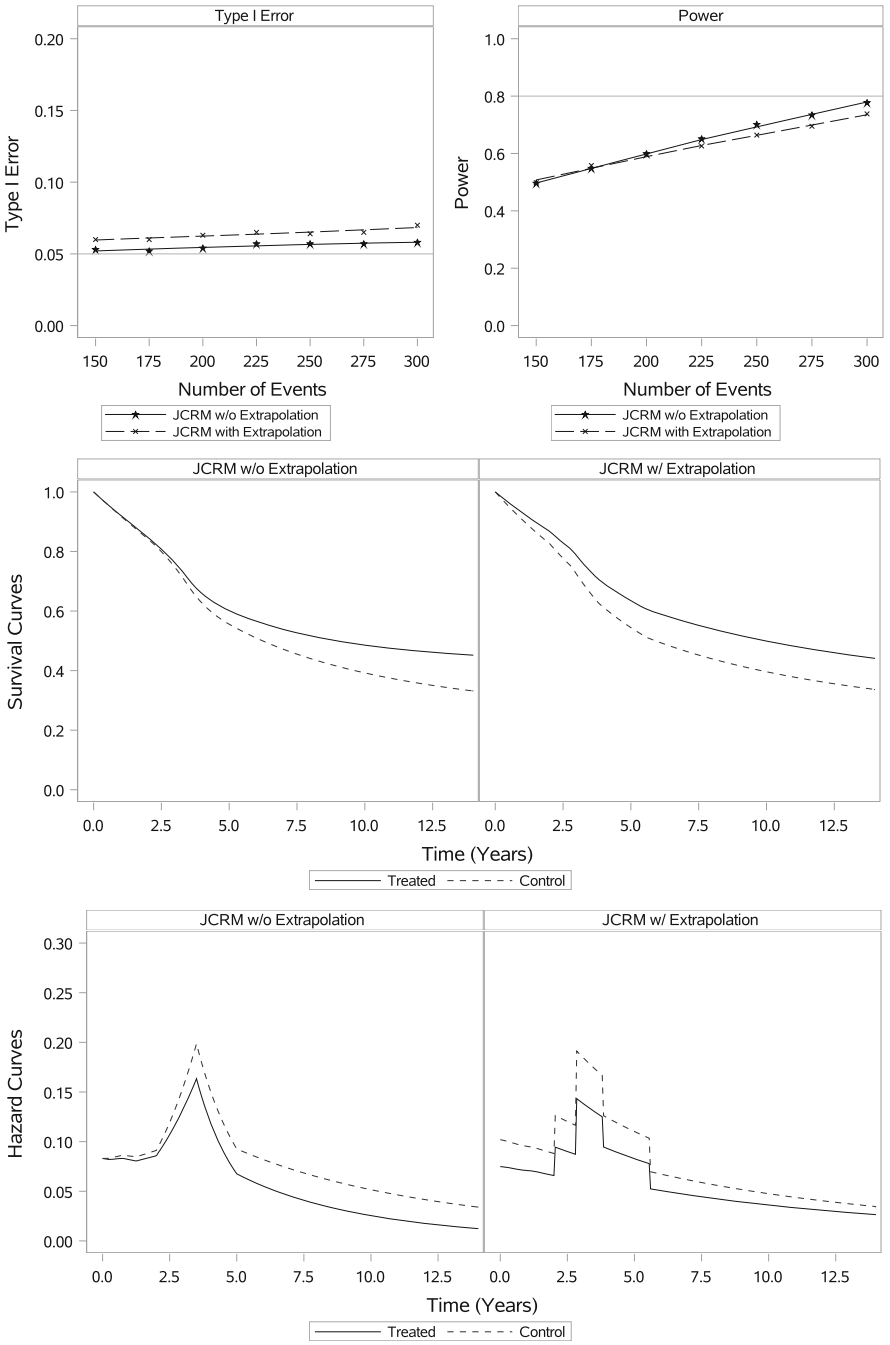

Fig. 5 Estimated power and type I error curves based on the JCRMs with and without extrapolation

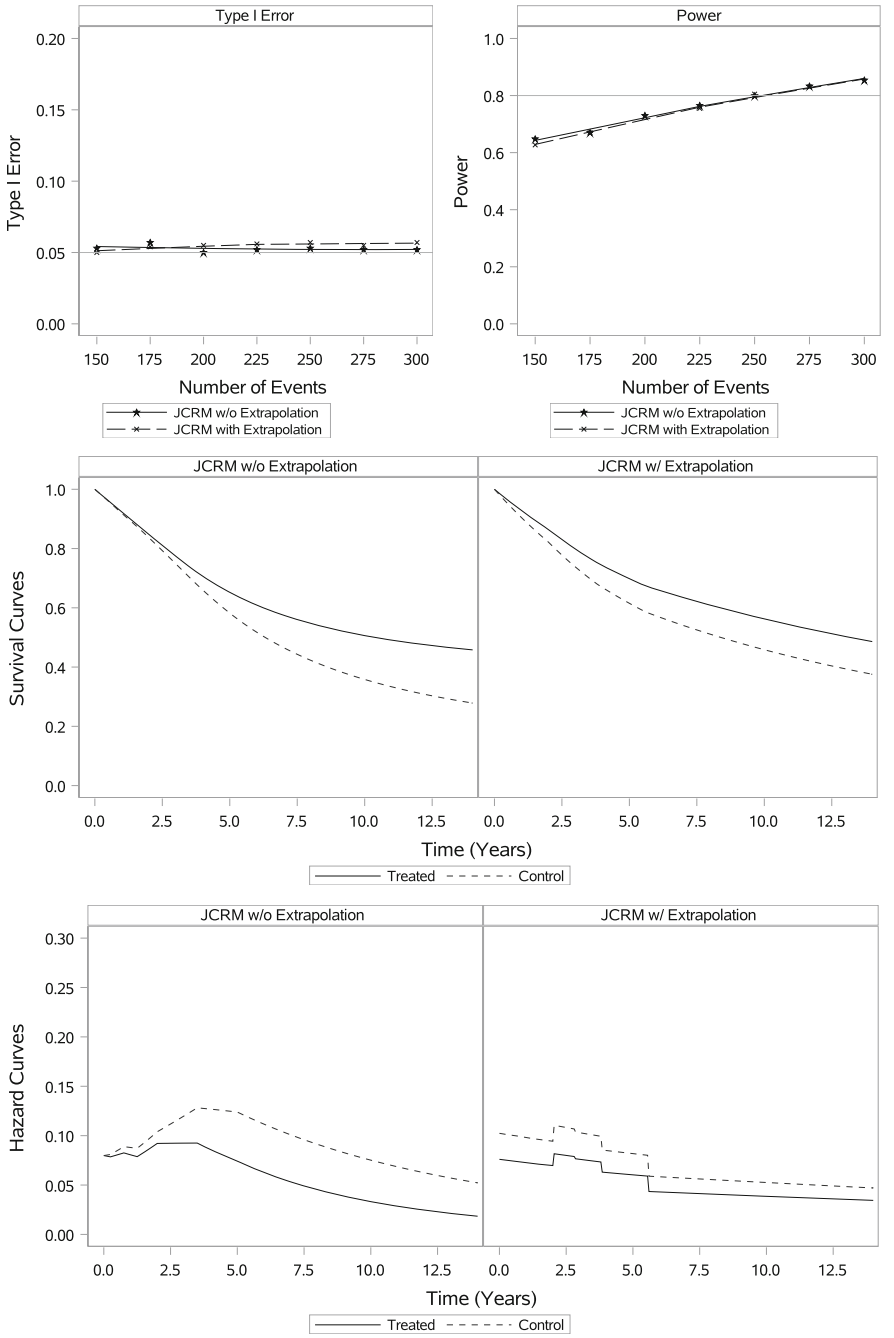

Fig. 6 Estimated power and type I error curves based on the JCRMs with and without extrapolation

piecewise constant baseline hazard beyond t_{traj}). The generalization is easy to implement but not considered herein as our primary objective is to explore the possibility of implementing piecewise linear form of hazard function to avoid extrapolation.

In our development of the hazard model, the random effects are allowed to have time varying impact through the longitudinal process. In particular, in the case of a user-specified piecewise linear trajectory, both a random intercept and interval-specific random slopes may be included in the model though in our examples we only included a random intercept. In the case of random slopes for the piecewise linear trajectory, it would be difficult to carry their effect forward in the extrapolation model (other than allowing them to implicitly determine the subject-specific hazard at time t_{traj}). The effect of the random intercept is implicitly accounted for already.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10985-022-09581-5.

References

Berkson J, Gage RP (1952) Survival curve for cancer patients following treatment. J Am Stat Assoc 47(259):501–515

Brown ER, Ibrahim JG (2003) A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 59(2):221–228

Chen CF (1985) On asymptotic normality of limiting density functions with Bayesian implications. J R Stat Soc Ser B (Methodol) 47(3):540–546

Chen LM, Ibrahim JG, Chu H (2011) Sample size and power determination in joint modeling of longitudinal and survival data. Stat Med 30(18):2295–2309. https://doi.org/10.1002/sim.4263

Chen MH, Harrington DP, Ibrahim JG (2002) Bayesian cure rate models for malignant melanoma: a case-study of Eastern Cooperative Oncology Group trial E1690. J R Stat Soc Ser C (Appl Stat) 51(2):135–150

Chen MH, Ibrahim JG, Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. J Am Stat Assoc 94(447):909–919

Chen MH, Ibrahim JG, Sinha D (2004) A new joint model for longitudinal and survival data with a cure fraction. J Multivar Anal 91(1):18–34. https://doi.org/10.1016/j.jmva.2004.04.005

Chi YY, Ibrahim JG (2006) Joint models for multivariate longitudinal and multivariate survival data. Biometrics 62(2):432–445

Crowther MJ, Abrams KR, Lambert PC (2013) Joint modeling of longitudinal and survival data. Stand Genomic Sci 13(1):165–184. https://doi.org/10.1177/1536867X1301300112

De Gruttola V, Tu XM (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics 50(4):1003–1014. https://doi.org/10.2307/2533439

Doob JL (1935) The limiting distributions of certain statistics. Ann Math Stat 6(3):160–169

Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 38(4):1041–1046. https://doi.org/10.2307/2529885

Hürny C, Bernhard J, Gelber RD, Coates A, Castiglione M, Isley M, Dreher D, Peterson H, Goldhirsch A, Senn HJ (1992) Quality of life measures for patients receiving adjuvant therapy for breast cancer: an international trial. Eur J Cancer 28(1):118–124. https://doi.org/10.1016/0959-8049(92)90399-M

International Breast Cancer Study Group (1996) Duration and reintroduction of adjuvant chemotherapy for node-positive premenopausal breast cancer patients. J Clin Oncol 14(6):1885–1894. https://doi.org/10.1200/JCO.1996.14.6.1885

Ibrahim JG, Chen MH, Sinha D (2001) Bayesian survival analysis. Springer, Berlin

Kim S, Zeng D, Li Y, Spiegelman D (2013) Joint modeling of longitudinal and cure-survival data. J Stat Theory Pract 7(2):324–344. https://doi.org/10.1080/15598608.2013.772036

Kirkwood J, Ibrahim JG, Sondak, VK, Richards JM, Flaherty LE, Ernstoff M, Smith, TJ, Rao UN, Steele M, Blum R (2000) High- and low-dose interferon Alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190, vol 18. https://doi.org/10.1200/JCO.2000.18.12.2444

Psioda MA, Ibrahim JG (2018) Bayesian design of a survival trial with a cured fraction using historical data. Stat Med 37(26):3814–3831
Psioda MA, Ibrahim JG (2018) Bayesian design of a survival trial with a cured fraction using historical data. Stat Med. https://doi.org/10.1002/sim.7846
Psioda MA, Ibrahim JG (2019) Bayesian clinical trial design using historical data that inform the treatment effect. Biostatistics 20(3):400–415. https://doi.org/10.1093/biostatistics/kxy009
Tsodikov A, Loeffler M, Yakovlev A (1998) A cure model with time-changing risk factor: an application to the analysis of secondary leukaemia. A report from the international database on Hodgkin's disease. Stat Med 17(1):27–40
Wang F, Gelfand AE (2002) A simulation based approach to sample size determination under a given model and for separating models. Stat Sci 17(2):193–208. https://doi.org/10.1214/ss/1030550861
Withers H, Peters LJ, Taylor JMG, Owen JB, Morrison WH, Schultheiss TE, Keane T, O'Sullivan B, van Dyk J, Gupta N, Wang CC, Jones CU, Doppke KP, Myint S, Thompson M, Parsons JT, Mendenhall WM, Dische S, Aird EGA, Henk J, Bidmead MAM, Svoboda V, Chon Y, Hanlon AL, Peters TL, Hanks GE (1995) Local control of carcinoma of the tonsil by radiation therapy: an analysis of patterns of fractionation in nine institutions. Int J Radiat Oncol Biol Phys 33(3):549–562. https://doi.org/10.1016/0360-3016(95)00228-Q
Woods LM, Rachet B, Lambert PC, Coleman MP (2009) Cure from breast cancer among two populations of women followed for 23 years after diagnosis. Ann Oncol 20(8):1331–1336
Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53(1):330–339. https://doi.org/10.2307/2533118
Xu J, Psioda MA, Ibrahim JG (2020) Bayesian design of clinical trials using joint models for longitudinal and time-to-event data. Biostatistics. https://doi.org/10.1093/biostatistics/kxaa044
Xu J, Zeger SL (2001) The evaluation of multiple surrogate endpoints. Biometrics 57(1):81–87. https://doi.org/10.1111/j.0006-341X.2001.00081.x
Zhang D, Chen MH, Ibrahim JG, Boye ME, Shen W (2016) JMFit: a SAS macro for joint models of longitudinal and survival data. J Stat Softw. https://doi.org/10.18637/jss.v071.i03