Bayesian Design of Superiority Clinical Trials for Recurrent Events Data with Applications to Bleeding and Transfusion Events in Myelodyplastic Syndrome

Ming-Hui Chen[1], **Joseph G. Ibrahim**[2], **Donglin Zeng**[2], **Kuolung Hu**[3], and **Catherine Jia**[3]

Joseph G. Ibrahim: ibrahim@bios.unc.edu

[1]Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, Connecticut 06269, U.S.A

[2]Department of Biostatistics, University of North Carolina, McGavran Greenberg Hall, CB#7420, Chapel Hill, North Carolina 27599, U.S.A

[3]Global Biostatistical Science, Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320, U.S.A

Summary

In many biomedical studies, patients may experience the same type of recurrent event repeatedly over time, such as bleeding, multiple infections and disease. In this article, we propose a Bayesian design to a pivotal clinical trial in which lower risk myelodysplastic syndromes (MDS) patients are treated with MDS disease modifying therapies. One of the key study objectives is to demonstrate the investigational product (treatment) effect on reduction of platelet transfusion and bleeding events while receiving MDS therapies. In this context, we propose a new Bayesian approach for the design of superiority clinical trials using recurrent events frailty regression models. Historical recurrent events data from an already completed phase 2 trial are incorporated into the Bayesian design via the partial borrowing power prior of Ibrahim et al. (2012, *Biometrics* **68**, 578–586). An efficient Gibbs sampling algorithm, a predictive data generation algorithm, and a simulation-based algorithm are developed for sampling from the fitting posterior distribution, generating the predictive recurrent events data, and computing various design quantities such as the type I error rate and power, respectively. An extensive simulation study is conducted to compare the proposed method to the existing frequentist methods and to investigate various operating characteristics of the proposed design.

Keywords

Clinical trial design; Gibbs sampling; Myelodysplastic syndrome; Power prior; Recurrent events; Type I error rate and power

---

1. Introduction

Recurrent events data are increasingly playing a dominant role in the design of clinical trials in practice. One of the main reasons for this is that time-to-event data are now often collected on several endpoints, or the same endpoint at different points in time. There are many examples of recurrent events data that arise in practice, such as recurrent infections, disease episodes, bleeding, recurrences of tumors, etc. As a result of these types of settings, statistical clinical trials designs are now being formally considered in the recurrent events arena with the hope that such designs will more fully capture the disease process and hence yield greater efficiency and power than designs based on a univariate time-to-event endpoint. Simply put, having more time-to-event measurements for a subject generally yields greater power than having only one time-to-event endpoint for each subject.

There has been some literature on designing clinical trials in the recurrent events context, essentially all of it from a frequentist perspective. The design settings considered in Hughes (1997) and Bernardo and Harrington (2001) are based on a multiplicative intensity model and a marginal proportional hazards model, respectively. Using the test of Cook (1995), Lawless and Nadeau (1995) and Matsui (2005) consider designs using a nonhomogeneous Poisson process model. Their methods are parametric in that, conditional on a frailty, the intensity of a homogeneous Poisson process is needed as an input parameter for the sample size calculations. Song, Kosorok, and Cai (2008) propose a covariate-adjusted logrank test to improve the power of the tests and to adjust for random imbalances of the covariates at baseline using a semiparametric approach. They derive a sample size formula based on the limiting distribution of the robust log-rank statistic. Chen, Ibrahim, and Chu (2014) develop a sample size determination method for the shared frailty model to investigate the treatment effect on multivariate event times. They consider sample size determination for testing the treatment effect on one time-to-event while treating the other event times as nuisance, and compare the power from a multivariate frailty model to that of a univariate parametric and semi-parametric survival model. There has been essentially no literature on developing Bayesian design methods for recurrent events data. Bayesian methods are inherently attractive in this complex setting since they are more computationally flexible, making no large sample assumptions, are more amenable to adaptive design settings with multiple endpoints than frequentist methods, generally have excellent frequentist operating characteristics (e.g., Chen et al., 2011; Ibrahim et al., 2012), and allow for the natural incorporation of historical data and other prior information in the design that can lead to more desirable designs than their frequentist counterparts.

To motivate our methodology, we consider a case study of myelodysplastic syndromes (MDS), where the interest lies in evaluating a treatment effect of an investigational product (IP) for reducing recurrent bleeding or transfusion events among lower risk MDS patients receiving disease modifying therapy. General symptoms associated with MDS include fatigue, dizziness, weakness, bruising, and bleeding. The disease is characterized by progressive impairment of myelodysplastic stem cells that increases the risk of evolution into acute myeloid leukemia (AML) (Malcovati et al., 2005). In this article, we aim to design a pivotal clinical trial in which lower risk MDS patients are treated with MDS disease modifying therapies along with investigational product (IP). The primary study objective is

to demonstrate the treatment effect on reduction of platelet transfusion utilization and bleeding event occurrences during the course of receiving MDS therapies. A superiority trial design is therefore considered since IP is expected to reduce the transfusion incidence. The most common side effect of MDS disease modifying therapies is thrombocytopenia (platelet count $< 100 \times 10^9$/L). Thrombocytopenia prevents the on-schedule of MDS therapies administered and increases the risk of bleeding, and thus could result in morbidity. Platelet transfusion interventions are often required therapeutically (given to patients who are actively bleeding) and prophylactically (to prevent future bleeding) during the therapy cycles. Therefore, the incidence of platelet transfusion intervention and bleeding event occurrence are highly correlated, and as a result, the composite endpoint of bleeding events and platelet transfusion events is currently being explored as a primary endpoint.

In this article, we propose a novel approach to Bayesian clinical trials design for superiority trials with recurrent events data. We postulate a semiparametric cumulative intensity model with covariates that is very flexible and allows for very general shapes of the baseline intensity. By taking the baseline intensity to be piecewise constant, our model can accommodate a wide variety of general settings well beyond proportional intensities. In our Bayesian design development, we consider the sampling and fitting priors of Wang and Gelfand (2002) that lead to flexible and general classes of Bayesian designs allowing for incorporation of historical data. Using the Wang and Gelfand (2000) framework, we are able to define a key quantity that determines the frequentist operating characteristics of the design, namely the type I error rate and power. The MDS trial is perfectly suited for this type of design since our goal is to design a study with a moderate number of subjects, and therefore large sample approximations are not at all desirable. Also, we have historical data available that will facilitate a more efficient design with desirable operating characteristics. We compare our Bayesian design to the naïve sample size calculation under a Poisson assumption as well as that of Song et al. (2008), and show that our design has more desirable operating characteristics than either of these two methods.

The rest of this article is organized as follows. In Section 2, we present the cumulative intensity model and derive the likelihood function. In Section 3, we present the proposed Bayesian methodology for superiority design. We define the sampling and fitting priors along with the key Bayesian criterion for determining the operating characteristics of the design. We also provide details on the predictive data generation algorithm. In Section 4, we present a general model for historical data, propose the partial borrowing power prior as the fitting prior, and discuss the sampling prior that is used for generating the predictive data. In Section 5, we derive several attractive properties of the full conditional distributions, develop an efficient Gibbs sampling algorithm via elegant reparameterizations, and provide a simulation-based algorithm for computing the type I error rate and power. Section 6 presents a detailed design for the MDS clinical trial in the presence of historical data as well as a detailed simulation study examining the operating characteristics of our proposed design. We close the article with some discussion in Section 7.

2. The Gamma Frailty Regression Model

For single-type recurrent events data, we consider the cumulative intensity model:

$$\mathrm{d}\Lambda_i(t) = \omega_i \mathrm{d}\Lambda_0(t)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta}), \quad (1)$$

where $\Lambda_0(t)$ is the baseline cumulative intensity function, z_i denotes the treatment indicator, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ is a p-dimensional vector of baseline covariates associated with subject i, γ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ are the corresponding regression coefficients, and ω_i is the subject dependent frailty. We further assume that $\omega_i \sim \mathrm{Gamma}(1/\tau, 1/\tau)$, which is a gamma distribution with mean 1 and variance τ. The model defined by (1) is thus termed as the recurrent events frailty model. We note that a large value of τ implies high dependence among recurrent event times. Let $\lambda_0(t) = \dfrac{\mathrm{d}\Lambda_0(t)}{\mathrm{d}t}$ denote the baseline intensity function. We assume a piecewise constant intensity function for $\lambda_0(t)$. Specifically, let $s_0 = 0 < s_1 < s_2 < \cdots < s_K = \infty$ denote a partition of $R^+ = (0, \infty)$. We assume a constant intensity λ_k over the kth interval $I_k = (s_{k-1}, s_k]$ for $k = 1, 2, \ldots, K$. That is, $\lambda_0(t) = \lambda_k$ if $t \in I_k$ for $k = 1, \ldots, K$, and we write $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)'$.

Let T denote the duration of the trial and let $N_i(t)$ denote the number of recurrent events by time t for $t \leq T$. We assume that the $N_i(t)$'s are independent across subjects. The observed data consists of the recurrent events $D_n = \{N_i(t \wedge c_i), c_i, z_i, \boldsymbol{x}_i, i = 1, \ldots, n\}$, where the censoring time $c_i \leq T$ is conditionally independent of $N_i(t)$ given z_i and $\boldsymbol{x}_i$ for $i = 1, \ldots, n$. Under the model defined by (1), the observed-data likelihood function is given by

$$L(\boldsymbol{\theta}|D_n) = \prod_{i=1}^{n} \int \left[\prod_{t \leq c_i} \left\{ \omega_i \lambda_0(t)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}^{\mathrm{d}N_i(t)} \right] \times \exp\{-\omega_i\Lambda_0(c_i)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta})\} f(\omega_i|\tau)\mathrm{d}\omega_i, \quad (2)$$

where $f(\omega_i|\tau)$ is the density of a $\mathrm{Gamma}(1/\tau, 1/\tau)$ distribution and $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}, \tau)$. We note that when $\tau \to 0$, $P(w_i = 1|\tau) \to 1$ and in this case, the model in (1) reduces to a non-homogeneous Poisson process model with the intensity function given by

$\mathrm{d}\Lambda_i(t) = \mathrm{d}\Lambda_0(t)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta})$, where $\Lambda_0(t)$ is defined by (1). The observed likelihood function for the Poisson process model is given by

$$L_{\mathrm{Poisson}}(\boldsymbol{\theta}|D_n) = \prod_{i=1}^{n} \prod_{t \leq c_i} \left\{ \lambda_0(t)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}^{\mathrm{d}N_i(t)} \exp\left\{ -\Lambda_0(c_i)\exp(z_i\gamma + \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}.$$

3. Proposed Methodology

3.1. Bayesian Design

The hypotheses for "superiority" testing are given as follows: H_0: $\exp(\gamma) \geq \delta$ versus H_1: $\exp(\gamma) < \delta$, where δ is a prespecified design margin. For a superiority trial, $\delta < 1$ and the trial is successful if H_1 is accepted. Note that $\exp(\gamma)$ is the intensity ratio of the treatment compared to the control arm (placebo).

Following Wang and Gelfand (2002) and Chen et al. (2011), let $\pi^{(s)}(\boldsymbol{\theta})$ denote the sampling prior, which is used to generate the data, and also let $\pi^{(f)}(\boldsymbol{\theta})$ denote the fitting prior, which is used to fit the model once the data are generated. Under the fitting prior, the posterior distribution of $\boldsymbol{\theta}$ given the data D_n takes the form $\pi^{(f)}(\boldsymbol{\theta}|D_n) \propto L(\boldsymbol{\theta}|D_n)\pi^{(f)}(\boldsymbol{\theta})$, where $L(\boldsymbol{\theta}|D_n)$ is defined by (2). We note that $\pi^{(f)}(\boldsymbol{\theta})$ may be improper as long as the resulting posterior

$\pi^{(f)}(\theta|D_n)$ is proper. Further we let $f^{(s)}(D_n)$ denote the marginal distribution of D_n based on the sampling prior. The key design quantity is defined as follows:

$$\beta_s^{(n)}=E_s\left[1\{P(\exp(\gamma)<\delta|D_n,\pi^{(f)})\geq\phi_0\}\right], \quad (3)$$

where the indicator function $1\{A\}$ is 1 if A is true and 0 otherwise, $\phi_0>0$ is a prespecified quantity, the probability $P(\cdot|D_n,\pi^{(f)})$ is computed with respect to the posterior distribution given the data D_n and the fitting prior $\pi^{(f)}(\theta)$, and the expectation $E_s[\cdot]$ is taken with respect to the marginal distribution of D_n under the sampling prior $\pi^{(s)}(\theta)$.

Let Θ_0 and Θ_1 denote the parameter spaces corresponding to H_0 and H_1. We let $\bar{\Theta}_0$ and $\bar{\Theta}_1$ denote the closures of Θ_0 and Θ_1 and specify the sampling prior $\pi_0^{(s)}(\theta)$ with support $\Theta_B = \bar{\Theta}_0\cap\bar{\Theta}_1$ and another sampling prior $\pi_1^{(s)}(\theta)$ with support $\Theta_1^*\subset\Theta_1$. Using (3), for given $0<\alpha_0<1$ and $0<\alpha_1<1$, we compute $n_{\alpha_0}=\min\{n:\beta_{s0}^{(n)}\leq\alpha_0\}$ and $n_{\alpha_1}=\min\{n:\beta_{s1}^{(n)}\geq 1-\alpha_1\}$, where $\beta_{s0}^{(n)}$ and $\beta_{s1}^{(n)}$ given in (3) corresponding to $\pi^{(s)}=\pi_0^{(s)}$ and $\pi^{(s)}=\pi_1^{(s)}$ are the type I error rate and power, respectively. Then, the sample size is given by $n_B = \max\{n_{\alpha_0}, n_{\alpha_1}\}$. For the design of a superiority trial, we choose $\phi_0 \geq 0.95$ to ensure that the entire 95% Bayesian credible interval of $\exp(\gamma)$ is below δ. Common choices of α_0 and α_1 include $\alpha_0 = 0.05$ and $\alpha_1 = 0.20$ so that the sample size n_B guarantees that the type I error rate is less than or equal to 0.05 and the power is at least 0.80.

3.2. Predictive Data Generation Algorithm

We use the sampling prior, $\pi^{(s)}(\theta)$, to generate the predictive data D_n. In other words, we view the distribution $f^{(s)}(D_n)$ as the prior predictive marginal distribution of the data. The prior predictive data generation algorithm is given as follows: **Step 0**. Set T, n, K, $s_0 = 0 < s_1 < \cdots < s_{K-1} < s_K = \infty$, and $0 < p < 1$, which is the probability of treatment assignment; **Step 1**. Generate $\theta \sim \pi^{(s)}(\theta)$; **Step 2**. Generate a vector of covariates, x, from a known distribution $f(x)$ and independently generate the treatment assignment indicator $z \sim$ Bernoulli(p); **Step 3**. Generate $\omega \sim$ Gamma$(1/\tau, 1/\tau)$; **Step 4**. Set $t_0^*=0$, sequentially generate t_j^* from a piecewise exponential regression model with density

$$f(t|\theta, z, x, \omega, t_{j-1}^*) \propto \lambda_0(t)\omega\exp(z\gamma+x'\beta)\exp\{-\Lambda_0(t)\omega\exp(z\gamma+x'\beta)\} \text{ for } t>t_{j-1}^*, \text{ where}$$

$\Lambda_0(t)=\int_0^t\lambda_0(u)du$, and compute the gap time $y_j^*=t_j^* - t_{j-1}^*$ for $j = 1, 2, \ldots$; **Step 5**. Generate $c^* \sim g(c|x)$ independently, where $g(c|x)$ is a pre-specified distribution for the censoring time, and let $c = c^* \wedge T$; **Step 6**. Compute $N(t)=\max_j\{j:\sum_{\ell=1}^j y_\ell^* \leq t\}$ for $t \leq c$ and record all jump times; and **Step 7**. Independently repeat Steps 2–6 n times to obtain D_n.

4. Specification of the Fitting and Sampling Priors

4.1. The Model for the Historical Data

Let T_0 denote the duration of the historical trial and let $N_{0i}(t)$ denote the number of recurrent events by time t for $t \leq T_0$. We assume that the $N_{0i}(t)$'s are independent across subjects. The observed data consists of the recurrent events $D_0 = \{N_{0i}(t \wedge c_{0i}), c_{0i}, z_{0i}, x_{0i}, i = 1, \ldots, n_0\}$,

where c_{0i} T_0 is the noninformative censoring time, z_{0i} denotes the treatment indicator, and x_{0i} is a p-dimensional vector of the baseline covariates associated with subject i. Here, we assume that z_{0i} and z_i are the indicators of the same treatment and x_{0i} and x_i are the vectors of the same baseline covariates. Let $\Lambda_{h0}(t)$ denote the cumulative baseline intensity function based on the historical data and $\lambda_{h0}(t) = \dfrac{d\Lambda_{h0}(t)}{dt}$. We consider a piecewise constant intensity function for $\lambda_{h0}(t)$, which may be different than $\lambda_0(t)$. Specifically, let $s_{00} = 0 < s_{01} < \cdots < s_{0K_0} = \infty$ denote a partition of $R^+ = (0, \infty)$. We assume a constant intensity λ_{0k} over the kth interval $I_{0k} = (s_{0,k-1}, s_{0k}]$ for $k = 1, 2, \ldots, K_0$ as $\lambda_{h0}(t) = \lambda_{0k}$ if $t \in I_{0k}$ for $k = 1, \ldots, K_0$. We write $\lambda_{h0} = (\lambda_{01}, \ldots, \lambda_{0K_0})'$ and further assume that the cumulative intensity function for the recurrent events based on the historical data is given by

$$d\Lambda_{hi}(t) = \omega_{0i} d\Lambda_{h0}(t) \exp(z_{0i}\gamma + x'_{0i}\beta), \quad (4)$$

where ω_{0i} is the subject dependent frailty, which follows a Gamma$(1/\tau_0, 1/\tau_0)$ distribution, and τ_0 may be different than τ.

Comparing (4) to (1), we see that the models for the historical data and the current data share only two common parameters γ and β. This assumption is reasonable when the treatment and covariates in the current data are the same as those in the historical data. Thus, strength from the historical data is borrowed only through the common parameters γ and β. Having different parameters, λ_{h0} and τ_0, and a different partition, $s_{00} = 0 < s_{01} \cdots < s_{0K_0} = \infty$, for $\lambda_{h0}(t)$, for the historical data provides us with greater flexibility in accommodating different baseline intensity functions and different frailty distributions to capture the dependence among recurrent event times in the current and historical data. In addition, we use the partial borrowing power prior of Ibrahim et al. (2012) as the fitting prior to control the influence of the historical data on the current study when γ and β in the historical data are different than those in the current study.

4.2. Partial Borrowing Power Priors

Under the model defined by (4), the observed-data likelihood function based on D_0 is given by

$$L(\gamma, \beta, \lambda_{h0}, \tau_0 | D_0) = \prod_{i=1}^{n_0} f(N_{0i}(t \wedge c_{0i}) | z_{0i}, x_{0i}, \gamma, \beta, \lambda_{h0}, \tau_0), \quad (5)$$

where

$$f(N_{0i}(t \wedge c_{0i}) | z_{0i}, x_{0i}, \gamma, \beta, \lambda_{h0}, \tau_0) = \prod_{t \leq c_{0i}} \left\{ \tau_0 \lambda_{h0}(t) \exp(z_{0i}\gamma + x'_{0i}\beta) \right\}^{dN_{0i}(t)} \Gamma(\frac{1}{\tau_0}$$
$$+ \sum_{t \leq c_{0i}} dN_{0i}(t)) \left\{ \Gamma(\frac{1}{\tau_0}) \right\}^{-1} \left\{ 1 + \tau_0 \Lambda_{h0}(c_{0i}) \exp(z_{0i}\gamma + x'_{0i}\beta) \right\}^{-\frac{1}{\tau_0} - \sum_{t \leq c_{0i}} dN_{0i}(t)}.$$

Following Ibrahim and Chen (2000) and Ibrahim et al. (2012), we propose the following partial borrowing power prior for θ,

$$\pi^{(f)}(\boldsymbol{\theta}|D_0, a_0) \propto \left[\int \{L(\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}_{h0}, \tau_0|D_0)\}^{a_0} \pi_0^{(f)}(\boldsymbol{\lambda}_{h0}, \tau_0) \mathrm{d}\boldsymbol{\lambda}_{h0}\mathrm{d}\tau_0 \right] \times \pi_0^{(f)}(\boldsymbol{\theta}), \quad (6)$$

where $0 \le a_0 \le 1$, $\pi_0^{(f)}(\boldsymbol{\lambda}_{h0}, \tau_0)$ is an initial prior for $(\boldsymbol{\lambda}_{h0}, \tau_0)$, and $\pi_0^{(f)}(\boldsymbol{\theta})$ is an initial prior for $\boldsymbol{\theta}$. The parameter a_0 can be interpreted as a relative precision parameter for D_0. One of the main roles of a_0 is that it controls the heaviness of the tails of the prior for γ and β. As a_0 becomes smaller, the tails of (6) become heavier. If $a_0 = 1$, (6) corresponds to the update of $\pi_0^{(f)}(\boldsymbol{\theta})$ using Bayes theorem while $a_0 = 0$ is equivalent to a prior specification with no incorporation of historical data. Thus, the parameter a_0 controls the influence of the historical data on the current study. In (6), the initial prior $\pi_0^{(f)}(\boldsymbol{\theta})$ is further specified as

$$\pi_0^{(f)}(\boldsymbol{\theta}) = \pi_0^{(f)}(\gamma, \boldsymbol{\beta})\pi_0^{(f)}(\boldsymbol{\lambda})\pi_0^{(f)}(\tau), \quad (7)$$

where $\pi_0^{(f)}(\gamma, \boldsymbol{\beta}) \propto 1$ (an improper uniform prior), $\pi_0^{(f)}(\boldsymbol{\lambda}) \propto \prod_{k=1}^{K} [\lambda_k^{a_{f0k}-1}\exp(-b_{f0k}\lambda_k)]$ with $a_{f0k} \ge 0$ and $b_{f0k} \ge 0$ for $k = 1, \ldots, K$, and $\pi_0^{(f)}(\tau) \propto \tau^{-(d_{f0\tau}+1)}\exp(-b_{f0\tau}/\tau)$ (an inverse Gamma (IG) distribution) with $a_{f0\tau} > 0$ and $b_{f0\tau} > 0$. Similarly, we assume that $\pi_0^{(f)}(\boldsymbol{\lambda}_{h0}, \tau_0) = \pi^{(f)}(\boldsymbol{\lambda}_{h0})\pi_0^{(f)}(\tau_0)$, where $\pi_0^{(f)}(\boldsymbol{\lambda}_{h0}) \propto \prod_{k=1}^{K_0} \left[\lambda_{0k}^{a_{hf0k}-1}\exp(-b_{hf0k}\lambda_{0k}) \right]$ with $a_{hf0k} > 0$ and $b_{hf0k} > 0$ for $k=1, \ldots, K_0$, and $\pi_0^{(f)}(\tau_0) \propto \tau_0^{-(d_{hf0\tau_0}+1)}\exp(-b_{hf0\tau_0}/\tau_0)$ with $a_{hf0\tau} > 0$ and $b_{hf0\tau} > 0$.

4.3. Sampling Priors

For the sampling prior, $\pi_\ell^{(s)}(\boldsymbol{\theta})$, $\ell = 0, 1$, we take

$$\pi_\ell^{(s)}(\boldsymbol{\theta}) \propto \pi^{(s)}(\gamma)\pi^{(s)}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tau). \quad (8)$$

In (8), we first specify a point mass prior for $\pi_\ell^{(s)}(\gamma)$ with $\pi_\ell^{(s)}(\gamma) = \Delta_{\{\gamma=\log(\delta)\}}$ for $\ell = 0$ and $\Delta_{\{\gamma=\gamma_1\}}$ for $\ell = 1$, where $\gamma_1 < \log(\delta)$. In addition, we specify a point mass prior for $\pi^{(s)}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \tau)$.

5. Computational Development

In this section, we develop an efficient Gibbs sampling algorithm to sample from the posterior distribution under the fitting prior and a simulation-based algorithm for computing the type I error and power. Using (2) and (6), the augmented posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\lambda}_0, \tau_0)$ is given by $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}_{h0}, \tau_0|D_n, D_0, a_0) \propto L(\boldsymbol{\theta}|D_n)$ $\{L(\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}_{h0}, \tau_0|D_0)\}^{a_0} \times \pi_0^{(f)}(\boldsymbol{\lambda}_{h0}, \tau_0)\pi_0^{(f)}(\boldsymbol{\theta})$. To develop an efficient Gibbs sampling algorithmm, we consider the reparameterizations $\boldsymbol{\lambda}^* = \tau\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{h0}^* = \tau_0\boldsymbol{\lambda}_{h0}$. The Jacobians of these transformations are $|\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\lambda}^*}| = \tau^{-K}$ and $|\frac{\partial\boldsymbol{\lambda}_{h0}}{\partial\boldsymbol{\lambda}_{h0}^*}| = \tau_0^{-K_0}$. Letting $\boldsymbol{\theta}^* = (\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}^*, \tau, \boldsymbol{\lambda}_{h0}^*, \tau_0)$ the posterior distribution of $\boldsymbol{\theta}^*$ is thus given by

$$\pi(\boldsymbol{\theta}^*|D_n, D_0, a_0) = \pi(\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}^*/\tau, \tau, \boldsymbol{\lambda}_{h0}^*/\tau_0, \tau_0 | D_n, D_0, a_0)\tau^{-K}\tau_0^{-K_0}. \quad (9)$$

Let $[A|B]$ denote the conditional distribution of A given B. The Gibbs sampling algorithm requires us to sample from the following distributions in turn: (i) $[\gamma, \boldsymbol{\beta}|\boldsymbol{\lambda}^*, \tau, \boldsymbol{\lambda}_{h0}^*, \tau_0, D_n, D_0, a_0]$; (ii) $[\boldsymbol{\lambda}^*|\gamma, \boldsymbol{\beta}, \tau, D_n]$; (iii) $\boldsymbol{\lambda}_{h0}^*|\gamma, \boldsymbol{\beta}, \tau_0, D_0, a_0]$; (iv) $[\tau|\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}^*, D_n]$, and (v) $[\tau_0|\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}_{h0}^*, D_0, a_0]$. For (i), we can show that the full conditional density $\pi(\gamma, \boldsymbol{\beta}|\boldsymbol{\lambda}^*, \tau, \boldsymbol{\lambda}_{h0}^*, \tau_0, D_n, D_0, a_0$ is log-concave in γ or each component of $\boldsymbol{\beta}$. For (ii), we first let $\xi_k = \log(\lambda_k^*)$ for $k = 1, \ldots, K$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)'$, and then we can show that the conditional density $\pi(\boldsymbol{\xi}|\gamma, \boldsymbol{\beta}, \tau, D_n)$ is log-concave in each component of $\boldsymbol{\xi}$. For (iii), we use a similar transformation for (ii) to sample $\lambda_{0k}^* = \tau_0\lambda_{0k}$ for $k = 1, \ldots, K_0$. For (iv), we take the transformation $\tau^* = 1/\tau$. Then, the conditional density $\pi(\tau^*|\gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}, D_n)$ is log-concave again. Finally for (v), we use the same transformation for (iv) to sample τ_0. The technical details regarding the log-concavity of these full conditional densities are given in Appendix A. With the log-concavity property, all of the model parameters can be more efficiently sampled from their respective full conditional distributions.

The simulation-based computational algorithm for computing $\beta_s^{(n)}$ in (3) is given as follows: **Step 0**. Specify n, δ, ϕ_0, B (the number of simulations), and M (the Gibbs sample size); **Step 1**. Generate $\boldsymbol{\theta}$ from $\pi^{(s)}(\boldsymbol{\theta})$ given by (8); **Step 2**. Generate D_n using the predictive data generation algorithm; **Step 3**. Run the Gibbs sampler to generate a Gibbs sample $\{\boldsymbol{\theta}^{(m)} = (\gamma^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\lambda}^{(m)}, \tau^{(m)}), m = 1, 2, \ldots, M\}$ of size M from the posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}_0, \tau_0 | D_n, D_0, a_0)$; **Step 4**. Compute $\hat{P}_f = \frac{1}{M}\sum_{m=1}^{M} 1\{\exp(\gamma^{(m)}\} < \delta\}$; **Step 5**. Check whether $\hat{P_f} \geq \phi_0$; **Step 6**. Repeat Steps 1–5 B times; and **Step 7**. Compute the proportion of $\{\hat{P_f} \geq \phi_0\}$ in these B runs, which gives an estimate of $\beta_s^{(n)}$.

6. Bayesian Design of the MDS Clinical Trial

6.1. The Historical MDS Data

A phase 2 trial of a randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of an investigational product (IP) of subjects with low or intermediate risk MDS receiving hypomethylating agents was conducted. MDS subjects receiving hypomethylating agents frequently develop thrombocytopenia (low platelet counts), and this may cause death and disability caused by bleeding. Blood or platelet transfusion are used to prevent or treat subjects' thrombocytopenia. This IP aims to assess the effectiveness of platelet counts increasing during the MDS regime. Overall, 69 subjects were enrolled in the study, 27 (39%) subjects received placebo (standard of care), and 42 (61%) subjects received the IP. Subjects were randomized and stratified by baseline platelet count ($< 50 \times 10^9/L$ vs. $\geq 50 \times 10^9/L$). Subjects were treated by IP weekly in combination with the hypomethylating agent for at least four cycles (around 4 months). Transfusion events were collected daily and bleeding events were collected weekly by diary. In order to properly consider the platelet transfusion impact, any transfusion incidents that occurred within a 3-day interval are considered as just one incident. Similarly, bleeding events happening in the same organ system class on the same date are considered as one incident. Based on this

phase 2 trial data, a summary of the composite endpoint by important baseline characteristics is shown in Table 1. The observed event rates with a 95% confidence interval (CI) based on the Poisson model are presented. From Table 1, patients with baseline platelet count $< 50 \times 10^9/L$ on average experienced 68.7 events per 100 subject weeks in the placebo cohort and 52.0 events for the IP cohort. Further-more, low event rates were observed in the patients with base-line platelet count higher than $50 \times 10^9/L$, which was within the expectation from medical point of view. The higher one's platelet count is, the less likely one would bleed. Throughout the rest of the article, we use this dataset as the historical MDS data with a binary treatment variable ($z_0 = 1$ if IP and $z_0 = 0$ if Placebo) and a baseline covariate ($x_0 = 1$ if platelet count is higher than or equal to $50 \times 10^9/L$ and $x_0 = 0$ if platelet count is less than $50 \times 10^9/L$). We carry out a detailed Bayesian analysis of the historical MDS data in Appendix B. Under the best model according to DIC and LPML in Table A1, the posterior means (Est's), the standard deviations (SD's), and the 95% highest posterior density (HPD) intervals of γ, β, and λ_{h0} are given in Table 2. These estimates of (γ, β, λ_{h0}, τ_0) will be used as the guide values for the design of future trials.

6.2. The Design Setting

From Table 2, we see that many of the posterior estimates of the λ_{0k}'s are similar. Based on the similarity of the posterior estimates, the 18 λ_{0k}'s may be collapsed into eight groups, namely, (i) $\{\lambda_{01}, \lambda_{02}\}$, (ii) $\{\lambda_{03}\}$, (iii) $\{\lambda_{04}\}$, (iv) $\{\lambda_{05}, \lambda_{06}, \lambda_{07}\}$, (v) $\{\lambda_{08}\}$, (vi) $\{\lambda_{09}, \lambda_{0,10}\}$, (vii) $\{\lambda_{0,11}, \ldots, \lambda_{0,17}\}$, and (viii) $\{\lambda_{0,18}\}$. Then, by comparing the average values of the λ_{0k}'s in these 8 groups, we may further collapse these eight groups to form 4 new groups, namely, (a) $\{(i), (ii)\}$, (b) $\{(iii)\}$, (c) $\{(iv), (v), (vi)\}$, and (d) $\{(vii), (viii)\}$. Thus, at the design stage, we collapse the partition with 18 pieces for the historical data to two smaller partitions with $K = 4$ and $K = 8$. We will then compare the designs with $K = 8$ to the designs with $K = 4$. We take the average values of those posterior estimates of the λ_{0k}'s in these groups as the design values for the λ_k's and in addition, according to these groups, we collapse the respective s_{0k}'s to form the s_k's, in which some of those s_k's are rounded to the days nearest to a multiple of 7-days for scheduling convenience. Specifically, for $K = 4$, the design values for the s_i's are $s_1 = 14$, $s_2 = 21$, and $s_3 = 49$; and the corresponding values of λ are $\lambda_1 = 0.14$, $\lambda_2 = 0.31$, $\lambda_3 = 0.13$, and $\lambda_4 = 0.08$. For $K = 8$, the design values for the s_i's are $s_1 = 9$, $s_2 = 14$, $s_3 = 16$, $s_4 = 34$, $s_5 = 38$, $s_6 = 50$, and $s_7 = 115$; and the corresponding values of λ are $\lambda_1 = 0.16$, $\lambda_2 = 0.12$, $\lambda_3 = 0.31$, $\lambda_4 = 0.12$, $\lambda_5 = 0.18$, $\lambda_6 = 0.13$, $\lambda_7 = 0.09$, and $\lambda_8 = 0.03$. For both cases, the design value for β is $\beta = -1.92$ and various design values ranging from $\tau = 0$ to $\tau = 3.18$ are considered. These design values are then used to specify the point mass prior $\pi^{(s)}$ (β, λ, τ) in (8). In the predictive data generation algorithm, we further assume $z \sim$ Bernoulli(0.5) and $x \sim$ Bernoulli(0.44), where the proportion 0.44 was estimated from the historical MDS data. The superiority margin δ and the trial duration are set to be exp(-0.025) = 0.975 and 225 days. For the initial prior in (7), the hyperparameters are specified as $a_{f0k} = b_{f0k} = 0.001$, $k = 1, \ldots, K$, for λ, and $d_{f0\tau} = b_{f0\tau} = 0.001$ for τ so that a noninformative initial prior is specified for (γ, β, λ, τ). Similarly, we take (a_{hf0k}, b_{hf0k}) = (0.001, 0.01), $k = 1, \ldots, K_0$, and ($d_{hf0\tau}$, $b_{hf0\tau}$) = (0.001, 0.01) for the hyperparameters in the initial prior for λ_{h0} and τ_0. Furthermore, $K_0 = 8$ in all computations of the partial borrowing power prior with $a_0 > 0$. To ensure that the type I error rate is controlled under 0.05, we consider $\phi_0 = 0.95$ as well as $\phi_0 = 0.96$. We then investigate the powers and type I error

rates under various sample sizes n, effect sizes γ, and values of τ to account for various degrees of dependency among the recurrent event times in the simulation study. In all computations of the Bayesian type I error rates and powers, $B = 10,000$ simulations and $M = 2,500$ with 200 burn-in iterations within each simulation were used.

6.3. The Simulation Results

Using the rate 1.2 of placebo versus IP for platelet transfusion or bleeding, and based on the naïve sample size calculation under the Poisson assumption, 88 and 116 subjects are required for a placebo-control randomized trial with 80% and 90% of power, respectively, with a treatment duration of 225 days. We note that in the naïve sample size calculation, the homogeneous Poisson process model is assumed and the within-subject recurrent event times are assumed to be independent. These assumptions may essentially lead to an under-estimation of the sample size.

According to Song et al. (2008), the sample size for a two-sided alternative with type I error rate α_0 and power $1 - \alpha_1$ is given by $n = \dfrac{(z_{1-\alpha_0} + z_{1-\alpha_1})^2 (\hat{D}_{1a} + \hat{\sigma}_\omega^2 \hat{D}_2)}{\gamma^2 p_1 p_2 \hat{D}_{1g}^2}$, where γ is the effect size, $p_1 + p_2 = 1$, p_1 and p_2 are the proportions of sample size allocation,

$$\hat{\sigma}_w^2 = \left[\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} N_{ij}(C_{ij})\{N_{ij}(C_{ij}) - 1\}/\hat{D}_2 - 1\right]^+, \hat{D}_{1a}$$

$$= \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\hat{\Lambda}_j(C_{ij}|\boldsymbol{x}_i), \hat{D}_2 = \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\hat{\Lambda}_j^2(C_{ij}|\boldsymbol{x}_i), D_{1g}$$

$$= \left[\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\Lambda}_1(C_{ij}|\boldsymbol{x}_i)\right\} \times \left\{\frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\Lambda}_2(C_{ij}|\boldsymbol{x}_i)\right\}\right]^{1/2}$$ is the right censoring

time, $\Lambda_j(C_{ij}|\boldsymbol{x}_i)$ is the estimated cumulative intensity function, $N_{ij}(C_{ij})$ is the total number of events over the interval $(0, C_{ij}]$ for the ith subject in the jth treatment arm for $i = 1, \ldots, n_j$ and $j = 1, 2$, and $n = n_1 + n_2$. As suggested by Song et al. (2008), we use the historical data to compute $\hat{D}_{1a}, \hat{D}_2, \hat{D}_{1g}, \hat{D}_2$, and $\hat{\sigma}_w^2$. Based on the historical MDS data, we obtain $\hat{D}_{1a} = 5.88$, $\hat{D}_{1g} = 5.98$, $\hat{D}_2 = 205.46$, and $\hat{\sigma}_w^2 = 0$. We note that $\hat{D}_{1a} = 5.88$ matches exactly the average number of events in the historical MDS data, and $\hat{D}_{1a}$ and $\hat{D}_{1g}$ are quite close. We take $p_1 = p_2 = 0.5$ and $\alpha_0 = 0.05$. Then, the frequentist sample sizes are 179 and 240 to achieve the powers of 80% and 90%, respectively, for $\gamma = -0.17$; and 130 and 173 to achieve the powers of 80% and 90%, respectively, for $\gamma = -0.20$. Finally, we would like to mention that the frequentist sample size n is derived under a two-sided alternative hypothesis, which is different than the one-sided alternative hypothesis for superiority testing. We also note that $\hat{\sigma}_w^2 = 0$ implies that independence is assumed on the recurrent event times.

Using the proposed Bayesian design, we first compute the type I error rates and powers under various design values with no incorporation of the historical MDS data ($a_0 = 0$). The results are reported in Table 3. Under a non-homogeneous Poisson process model (i.e., $\tau = 0$), for the effect size $\gamma = -0.2$ and $K = 4$, the powers were 0.89 and 0.87 for $\phi_0 = 0.95$ and $\phi_0 = 0.96$, respectively, for $n = 88$; and 0.95 and 0.94 for $\phi_0 = 0.95$ and $\phi_0 = 0.96$, respectively, for $n = 116$. However, when $K = 8$, $n = 116$, and $\gamma = -0.2$ are needed to achieve 80% power. In this case ($\tau = 0$), the Bayesian sample sizes are comparable to those

calculated using the naïve sample size calculation under the Poisson assumption and are smaller than those using the approach of Song et al. (2008). We further observe from Table 3 that (i) the power increases when the sample size n increases or the effect size (namely, the absolute value of γ) becomes larger; and (ii) the power decreases when τ, ϕ_0, and K increase. Specifically, when $K = 4$, $\tau = 1.5$, and $\gamma = -0.4$, the powers are 0.79, 0.85, and 0.88 for $n = 300$, 350, and 400, respectively, for $\phi_0 = 0.95$; 0.76 for $n = 300$ and $\phi_0 = 0.96$. Also, When $\tau = 1.5$, $\gamma = -0.4$, $n = 300$, and $\phi_0 = 0.95$, the power is 0.78 for $K = 8$. In addition, when $K = 4$, $n = 300$, $\gamma = -0.3$, and $\phi_0 = 0.95$, the powers are 0.90 for $\tau = 0.5$, 0.75 for $\tau = 1.0$, and 0.57 for $\tau = 1.5$. Thus, the design value of τ has the most impact on the power of the Bayesian design. Compared to τ, the differences in the powers are much smaller between $\phi_0 = 0.95$ and $\phi_0 = 0.96$ or between $K = 4$ and $K = 8$. We note that with no incorporation of the historical MDS data ($a_0 = 0$), the type I error rates are controlled at 5% for $\phi_0 = 0.95$ and 4% for $\phi_0 = 0.96$ for all cases. We also note that with no incorporation of historical data, $n = 300$ is sufficient to achieve 80% power when $\tau = 1.0$ and $\gamma = -0.35$, and a larger sample size such as $n = 400$ for $\tau = 2.15$ and $n = 450$ for $\tau = 3.18$ and a larger effect size such as $\gamma = -0.4$ for $\tau = 2.15$ and $\gamma = -0.45$ for $\tau = 3.18$ are needed to achieve 80% power.

Next, we compute the type I error rates and powers under various design values with incorporation of the historical MDS data using the partial borrowing power prior given in (6). The results are shown in Table 4. We see from Table 4 that for given τ, γ, n, and ϕ_0, both the power and the type I error rate increase as a_0 increases when $K = 4$. Compared to the results in Table 3, we see some gain in power when the historical MDS data are incorporated. For instance, when $K = 4$, $\tau = 1.5$, $\gamma = -0.4$, $n = 300$, and $\phi_0 = 0.96$, the power and type I error rate are 0.79 and 0.05 for $a_0 = 0.15$, respectively, while the power and type I error rate are 0.76 and 0.04, respectively, without incorporation of the historical MDS data. When $K = 4$, $\tau = 2.15$, $\gamma = -0.4$, and $n = 400$, the powers and type I error rates are 0.81 and 0.05, respectively, when $\phi_0 = 0.95$ and $a_0 = 0.1$, and 0.79 and 0.05, respectively, when $\phi_0 = 0.96$ and $a_0 = 0.15$. Under the same setting, the powers are 0.79 for $\phi_0 = 0.95$ and 0.75 for $\phi_0 = 0.96$ when $a_0 = 0$. Thus, we see about a 2% increase in power when $\phi_0 = 0.95$ and a 4% increase in power when $\phi_0 = 0.96$. We further see that when $K = 8$, $\tau = 1.5$, $\gamma = -0.4$, and $n = 300$, the powers and type I error rates are 0.80 and 0.05, respectively, for $\phi_0 = 0.95$, and 0.76 and 0.04, respectively, for $\phi_0 = 0.96$ when $a_0 = 0.15$; and 0.78 and 0.05, respectively, for $\phi_0 = 0.95$, and 0.75 and 0.04, respectively, for $\phi_0 = 0.96$ when $a_0 = 0$. When the model becomes more complex, namely, $K = 8$, the historical data play a minimum role in improving powers under the various design settings but the type I errors are well controlled at 5%. Since the sample size ($n_0 = 69$) of the historical MDS data is small, we do not expect that the incorporation of the historical MDS data leads to a substantial gain in power. As demonstrated in Chen et al. (2011) and Ibrahim et al. (2012), more gain in power can be obtained when a large historical dataset is available. The general strategy for selecting a_0 is to find the largest value of a_0 such that the type I error rate is still controlled at about 5%. For example, when $K = 4$, $\tau = 2.15$, $n = 400$, and $\phi_0 = 0.95$, the largest value of a_0 is about 0.1 since when $a_0 = 0.15$, the type I error rate exceeds 5%.

We mention that under the design setting discussed in this section, the average numbers of events across the 10,000 simulated datasets range from 11.4 to 13.5 for $K = 4$ and 8.1 to 9.6

for $K = 8$. The difference in average numbers of events between $K = 4$ and $K = 8$ is one factor leading to higher power under the designs with $K = 4$. Another factor leading to lower power under the designs with $K = 8$ is that more λ_{0k} parameters need to be estimated in the posterior inference once the predictive data D_n are generated. In estimating the baseline intensity for the historical data, the posterior estimate of λ_{04} is much higher than other λ_{0k}'s as shown in Table 2. Both the historical data and the clinical literature (e.g., Silverman et al., 2006) suggest that the hematology values tend to reach the nadir values during the second or third week of the cycle under the study setting. Therefore, we set up λ_{02} for $K = 4$ and λ_{03} for $K = 8$ to be higher than other λ_{0k}'s to characterize the simulated data better. However, we have also conducted additional simulations to examine whether the power and type I error are sensitive to the design value of λ_{02} under the designs with $K = 4$. Specifically, we use $\lambda_{02} = 0.15$ instead of $\lambda_{02} = 0.31$ for the designs with $K = 4$. The resulting powers and type I errors for $\tau = 1.5$ and $n = 300$ are shown in Table 5. We see that the powers and type I errors for $a_0 = 0$ in Table 5 are very similar to those shown in Table 3 and the powers and type I errors for $a_0 = 0.05, 0.1, 0.15, 0.2, 0.5$, and 1.0 are very similar to those in Table 4. Thus, the power and type I error are quite robust to the design value of λ_{02}. In addition, we carry out additional simulations to compute the powers and type I errors under non-point mass sampling priors for the key parameters γ, τ, and β. From Table A2 in Appendix C, we see that the powers under non-point mass sampling priors are lower than those in Table 4.

Although the power, type I error rate and sample sizes are the key components in the design, a detailed examination and discussion of the posterior estimates of the parameters in the recurrent events frailty model are given in Appendix D. From Table A3 in Appendix D, we see that the average of the posterior means of γ is very close to the corresponding design value, the coverage probability (CP) of γ is close to 95%, and the average of the posterior standard deviations (SD) and the simulation standard error (SE) are almost the same. We also see from the same table that the posterior estimates of β, τ, and $\boldsymbol{\lambda}$ all have similar frequentist properties. We have also examined the posterior estimates under other design values of $(\gamma, \beta, \tau, \boldsymbol{\lambda})$ and sample sizes, and similar results are obtained. These results empirically confirm that the simulation settings such as the number (B) of simulations and the Gibbs sample size are adequate and the Gibbs sampling algorithm performs well.

7. Discussion

In this article, we only considered the partial borrowing power prior to incorporate the historical recurrent events data. This prior is attractive since strength from the historical data is borrowed only through certain common parameters while the models for the current and historical data are allowed to have different parameters. Therefore, the partial borrowing power prior provides us with greater flexibility in accommodating heterogeneity between the current and historical data. The proposed methodology can be extended to allow for other types of priors discussed in Ibrahim and Chen (2000) and Hobbs, Sargent, and Carlin (2012). The empirical performance of the proposed methodology under other types of priors needs to be fully and thoroughly evaluated in future work. In addition, the proposed method can be generalized to handle more complex endpoints including more general event dependence structures, multivariate recurrent events with informative censoring, and time-dependent covariates. Specifically, joint frailty models as proposed in Zeng et al. (2014) can

be adopted to model multivariate recurrent events and informative censoring time simultaneously, where event-specific frailty and shared frailty across different types of events can be built in the models to account for shared information from all the data within the same subject. The sampling algorithm will need to be modified accordingly to efficiently sample from the posterior distribution of model parameters; however, the similar partial borrowing power prior can be developed to borrow information from historical trials in designing a future trial. Additional discussion and implementation details of the proposed methodology are given in Appendix E.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Bernardo MVP, Harrington DP. Sample size calculation for the two-sample problem using the multiplicative intensity. Statistics in Medicine. 2001; 20 557–279.

Chen L, Ibrahim JG, Chu H. Sample size determination in shared frailty models for multivariate time-to-event data. Journal of Biopharmaceutical Statistics. 2014; 24:817–833. [PubMed: 24697500]

Chen M-H, Ibrahim JG, Lam P, Yu A, Zhang Y. Bayesian design of non-inferiority trials for medical devices using historical data. Biometrics. 2011; 67:1163–1170. [PubMed: 21361889]

Cook RJ. The design and analysis of randomized trials with recurrent events. Statistics in Medicine. 1995; 14:2081–2098. [PubMed: 8552888]

Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Analysis. 2012; 7:1–36.

Hughes MD. Power considerations for clinical trials using multivariate time-to-event data. Statistics in Medicine. 1997; 16:865–882. [PubMed: 9160485]

Ibrahim JG, Chen M-H. Power prior distributions for regression models. Statistical Science. 2000; 15:46–60.

Ibrahim JG, Chen M-H, Xia HA, Liu T. Bayesian meta-experimental design: evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. Biometrics. 2012; 68:578–586. [PubMed: 21955084]

Lawless JF, Nadeau JC. Nonparametric estimation of cumulative mean functions for recurrent events. Technometrics. 1995; 37:158–168.

Malcovati L, Porta MG, Pascutto C, Invernizzi R, Boni M, Travaglino E, Passamonti F, Arcaini L, Maffioli M, Bernasconi P, Lazzarino M, Cazzola M. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: A basis for clinical decision making. Journal of Clinical Oncology. 2005; 23:7594–7603. [PubMed: 16186598]

Matsui S. Sample size calculations for comparative clinical trials with over-dispersed Poisson process data. Statistics in Medicine. 2005; 24:1339–1356. [PubMed: 15568188]

Silverman LR, McKenzie DR, Peterson BL, Holland JF, Backstrom JT, Beach CL, Larson RA. Further analysis of trials with azacitidine in patients with myelodysplastic syndrome: studies 8421, 8921, and 9221 by the Cancer and Leukemia Group B. Journal of Clinical Oncology. 2006; 24:3895–3903. [PubMed: 16921040]

Song R, Kosorok MR, Cai J. Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data. Biometrics. 2008; 64:741–750. [PubMed: 18162107]

Wang F, Gelfand AE. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science. 2002; 17:193–208.

Zeng D, Ibrahim JG, Chen M-H, Hu K, Jia C. Multivariate recurrent events in the presence of multivariate informative censoring with applications to bleeding and transfusion events in myelodyplastic syndrome. Journal of Biopharmaceutical Statistics. 2014; 24:429–442. [PubMed: 24605978]

Table 1

A summary of the historical MDS data

Treatment	Baseline characteristics	Observed event rate (event/per 100 subject-week)	95% CI
Placebo	Platelet counts $< 50 \times 10^9$/L	68.7	(58.9, 79.7)
	Platelet counts 50×10^9/L	7.3	(4.3, 11.7)
IP	Platelet counts $< 50 \times 10^9$/L	52.0	(45.3, 59.5)
	Platelet counts 50×10^9/L	8.8	(5.9, 12.7)

Table 2

Posterior estimates of the parameters based on the historical MDS data

Parameter	Est	SD	95% HPD	Parameter	Est	SD	95% HPD
γ	−0.17	0.40	(−0.93, 0.61)	β	−1.92	0.40	(−2.71, −1.14)
τ_0	2.15	0.53	(1.20, 3.18)				
$\lambda_{0,01}$	0.15	0.06	(0.06, 0.28)	$\lambda_{0,10}$	0.14	0.06	(0.05, 0.24)
$\lambda_{0,02}$	0.16	0.07	(0.06, 0.29)	$\lambda_{0,11}$	0.10	0.04	(0.03, 0.18)
$\lambda_{0,03}$	0.12	0.05	(0.04, 0.21)	$\lambda_{0,12}$	0.07	0.03	(0.02, 0.13)
$\lambda_{0,04}$	0.31	0.13	(0.10, 0.56)	$\lambda_{0,13}$	0.11	0.05	(0.04, 0.20)
$\lambda_{0,05}$	0.12	0.05	(0.04, 0.21)	$\lambda_{0,14}$	0.10	0.03	(0.03, 0.19)
$\lambda_{0,06}$	0.11	0.05	(0.03, 0.19)	$\lambda_{0,15}$	0.09	0.04	(0.03, 0.17)
$\lambda_{0,07}$	0.13	0.05	(0.04, 0.23)	$\lambda_{0,16}$	0.08	0.03	(0.03, 0.15)
$\lambda_{0,08}$	0.18	0.08	(0.06, 0.33)	$\lambda_{0,17}$	0.06	0.03	(0.02, 0.12)
$\lambda_{0,09}$	0.13	0.06	(0.04, 0.24)	$\lambda_{0,18}$	0.03	0.01	(0.01, 0.05)

Table 3

Powers under various design settings with $a_0 = 0$

K	Design value of τ	n	Effect size (γ)	Power	
				$\varphi_0 = 0.95$	$\varphi_0 = 0.96$
4	0	88	−0.17	0.78	0.74
			−0.2	0.89	0.87
		116	−0.17	0.86	0.83
			−0.2	0.95	0.94
	0.5	250	−0.3	0.85	0.82
		300	−0.3	0.90	0.88
	1.0	300	−0.3	0.75	0.67
		300	−0.35	0.82	0.80
		300	−0.4	0.91	0.89
	1.5	300	−0.3	0.57	0.53
		300	−0.35	0.69	0.66
		300	−0.4	0.79	0.76
		350	−0.3	0.62	0.60
		350	−0.35	0.75	0.72
		350	−0.4	0.85	0.83
		400	−0.3	0.68	0.64
		400	−0.35	0.80	0.76
		400	−0.4	0.88	0.86
	2.15	400	−0.4	0.79	0.75
		400	−0.45	0.87	0.85
	3.18	400	−0.4	0.65	0.62
		400	−0.45	0.75	0.72

K	Design value of τ	n	Effect size (γ)	Power $\Phi_0 = 0.95$	Power $\Phi_0 = 0.96$
		400	−0.50	0.82	0.78
		450	−0.4	0.69	0.66
		450	−0.45	0.79	0.76
		450	−0.50	0.85	0.83
8	0	88	−0.17	0.65	0.61
			−0.2	0.78	0.75
		116	−0.17	0.76	0.73
			−0.2	0.88	0.86
	0.5	250	−0.3	0.82	0.79
		300	−0.3	0.88	0.86
	1.0	300	−0.3	0.69	0.65
		300	−0.35	0.82	0.77
		300	−0.4	0.89	0.87
	1.5	300	−0.3	0.56	0.52
		300	−0.35	0.68	0.64
		300	−0.4	0.78	0.75
		350	−0.3	0.62	0.58
		350	−0.35	0.74	0.70
		350	−0.4	0.84	0.81
		400	−0.3	0.67	0.63
		400	−0.35	0.78	0.75
		400	−0.4	0.88	0.85
	2.15	400	−0.4	0.78	0.74
		400	−0.45	0.86	0.83

K	Design value of τ	n	Effect size (γ)	Power	
				$\varphi_0 = 0.95$	$\varphi_0 = 0.96$
3.18		400	−0.4	0.63	0.60
		400	−0.45	0.73	0.69
		400	−0.50	0.81	0.78
		450	−0.4	0.68	0.64
		450	−0.45	0.77	0.74
		450	−0.50	0.84	0.81

Type I errors are around 5% for $\varphi_0 = 0.95$ and around 4% for $\varphi_0 = 0.96$ for all cases.

Table 4

Powers and Type I errors under various design settings using the partial borrowing power priors

K	Design value of τ	n	Effect size (γ)	a_0	$\varphi_0 = 0.95$		$\varphi_0 = 0.96$	
					Power	Type I error	Power	Type I error
4	0.5	250	−0.3	0.05	0.86	0.05	0.83	0.04
		250	−0.3	0.2	0.87	0.06	0.84	0.05
		250	−0.3	0.5	0.90	0.08	0.89	0.07
		250	−0.3	1.0	0.95	0.16	0.93	0.13
	1.0	300	−0.3	0.05	0.72	0.05	0.68	0.04
			−0.35		0.83		0.80	
		300	−0.3	0.1	0.73	0.06	0.69	0.05
			−0.35		0.85		0.82	
		300	−0.3	0.2	0.76	0.07	0.72	0.06
			−0.35		0.85		0.83	
		300	−0.3	0.5	0.82	0.10	0.79	0.09
			−0.35		0.91		0.89	
		300	−0.3	1.0	0.90	0.18	0.88	0.15
			−0.35		0.95		0.92	
	1.5	300	−0.4	0.05	0.80	0.05	0.78	0.04
		300	−0.4	0.1	0.82	0.06	0.79	0.04
		300	−0.4	0.15	0.82	0.06	0.79	0.05
		300	−0.4	0.2	0.84	0.07	0.81	0.06
		300	−0.4	0.5	0.89	0.11	0.87	0.09
		300	−0.4	1.0	0.96	0.21	0.94	0.18
		350	−0.35	0.1	0.77	0.06	0.74	0.05
		350	−0.35	0.2	0.79	0.07	0.76	0.06
		350	−0.35	0.5	0.86	0.11	0.84	0.09

K	Design value of τ	n	Effect size (γ)	a₀	φ₀ = 0.95 Power	φ₀ = 0.95 Type I error	φ₀ = 0.96 Power	φ₀ = 0.96 Type I error
		350	−0.35	1.0	0.93	0.21	0.91	0.18
	2.15	400	−0.4	0.05	0.80	0.05	0.77	0.04
		400	−0.4	0.1	0.81	0.05	0.78	0.04
		400	−0.4	0.15	0.82	0.07	0.79	0.05
		400	−0.4	0.2	0.83	0.07	0.80	0.06
		400	−0.4	0.5	0.89	0.11	0.87	0.09
		400	−0.4	1.0	0.95	0.22	0.94	0.19
	3.18	400	−0.45	0.1	0.77	0.06	0.73	0.05
		400	−0.45	0.2	0.80	0.07	0.77	0.06
		400	−0.45	0.5	0.88	0.13	0.85	0.10
		400	−0.45	1.0	0.95	0.26	0.94	0.23
8	0.5	250	−0.3	0.2	0.82	0.05	0.79	0.04
		250	−0.3	0.5	0.83	0.05	0.80	0.04
		250	−0.3	1.0	0.84	0.06	0.81	0.05
	1.0	300	−0.3	0.2	0.68	0.05	0.65	0.05
			−0.35		0.81		0.78	
		300	−0.3	0.5	0.69	0.05	0.65	0.04
			−0.35		0.82		0.78	
		300	−0.3	1.0	0.69	0.05	0.65	0.04
			−0.35		0.82		0.78	
	1.5	300	−0.4	0.1	0.79	0.04	0.75	0.03
		300	−0.4	0.15	0.80	0.05	0.76	0.04
		300	−0.4	0.5	0.79	0.05	0.75	0.04
		300	−0.4	1.0	0.78	0.06	0.74	0.05
	2.15	400	−0.4	0.1	0.78	0.05	0.75	0.04
		400	−0.4	0.2	0.78	0.05	0.75	0.04

K	Design value of τ	n	Effect size (γ)	a₀	φ₀ = 0.95		φ₀ = 0.96	
					Power	Type I error	Power	Type I error
		400	−0.4	0.5	0.78	0.05	0.75	0.04
		400	−0.4	1.0	0.78	0.05	0.74	0.04
3.18		400	−0.45	0.2	0.73	0.05	0.69	0.04
		400	−0.5	0.2	0.81		0.78	
		400	−0.45	0.5	0.73	0.05	0.69	0.04
		400	−0.5	0.5	0.81		0.78	
		400	−0.45	1.0	0.73	0.05	0.69	0.04
		400	−0.5	1.0	0.80		0.76	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Powers and Type I errors under various design settings with and without incorporation of historical data for $K = 4$, $\tau = 1.5$, $n = 300$, and $\lambda = (0.14, 0.15, 0.13, 0.08)'$

Effect size (γ)	a_0	$\Phi_0 = 0.95$		$\Phi_0 = 0.96$	
		Power	Type I error	Power	Type I error
−0.3	0	0.57	0.05	0.53	0.04
−0.35	0	0.69		0.66	
−0.4	0	0.80		0.76	
−0.4	0.05	0.81	0.05	0.78	0.04
−0.4	0.1	0.81	0.06	0.79	0.05
−0.4	0.15	0.83	0.06	0.80	0.05
−0.4	0.2	0.84	0.07	0.81	0.06
−0.4	0.5	0.89	0.12	0.87	0.10
−0.4	1.0	0.96	0.21	0.94	0.18