Biometrics 67, 1163–1170
September 2011
DOI: 10.1111/j.1541-0420.2011.01561.x

Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data

Ming-Hui Chen,[1,*] Joseph G. Ibrahim,[2] Peter Lam,[3] Alan Yu,[4] and Yuanye Zhang[1]

[1]Department of Statistics, University of Connecticut, 215 Glenbrook Road, Storrs, Connecticut 06269, U.S.A.
[2]Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599, U.S.A.
[3]Boston Scientific Corporation, 100 Boston Scientific Way M1, Marlborough, Massachusetts 01752, U.S.A.
[4]Independent Consultant, 15620 76th Pl N, Maple Grove, Minnesota 55311, U.S.A.
[*]*email:* mhchen@stat.uconn.edu

Summary. We develop a new Bayesian approach of sample size determination (SSD) for the design of noninferiority clinical trials. We extend the fitting and sampling priors of Wang and Gelfand (2002, *Statistical Science* **17,** 193–208) to Bayesian SSD with a focus on controlling the type I error and power. Historical data are incorporated via a hierarchical modeling approach as well as the power prior approach of Ibrahim and Chen (2000, *Statistical Science* **15,** 46–60). Various properties of the proposed Bayesian SSD methodology are examined and a simulation-based computational algorithm is developed. The proposed methodology is applied to the design of a noninferiority medical device clinical trial with historical data from previous trials.

Key words: Fitting prior; Hierarchical model; Power prior; Sampling prior; Simulation.

1. Introduction

Recently, the Food and Drug Administration (FDA) released "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials" (February 5, 2010, `www.fda.gov/MedicalDevices/DeviceRegulationand Guidance/GuidanceDocuments/ucm071072.htm`). This document provides guidance on statistical aspects of the design and analysis of Bayesian clinical trials for medical devices. It lays out detailed guidance on the determination of sample size in a Bayesian clinical trial. This document also provides guidance on the evaluation of the operating characteristics of a Bayesian clinical trial design. Specifically, the evaluation of a Bayesian clinical trial design should include type I error (probability of erroneously approving an ineffective or unsafe device), type II error (probability of erroneously disapproving a safe and effective device), and power (the converse of type II error: the probability of appropriately approving a safe and effective device).

Sample size determination (SSD) is a crucial aspect of clinical trial design. In this article, we are particularly interested in the design and analysis of noninferiority trials. There is a vast literature on the frequentist methods of SSD in various noninferiority trials, which includes, for example, D'Agostino, Massaro, and Sullivan (2003); Hung et al. (2003); Rothmann et al. (2003); Hung, Wang, and O'Neill (2005); Kieser and Friede (2007); and Fleming (2008). The literature on Bayesian SSD has been growing recently due to recent advances in Bayesian computation and Markov chain Monte Carlo sampling. Joseph, Wolfson, and Du Berger (1995); Lindley (1997); Rubin and Stern (1998); Katsis and Toman (1999); and Inoue, Berry, and Parmigiani (2005) are the Bayesian SSD articles cited in the FDA 2010 Guidance. An early review of Bayesian SSD is given in Adcock (1997). The most recent work includes Rahme and Joseph (1998); Simon (1999); Wang and Gelfand (2002); De Santis (2007); and M'Lan, Joseph, and Wolfson (2006, 2008). The existing literature on Bayesian SSD primarily focuses on simple normal, one or two sample binomial problems, standard normal linear regression, and generalized linear models (GLMs). Although the literature on Bayesian SSD discusses a variety of performance criteria, the widely used ones include the Bayes' factor (Weiss, 1997), the average posterior variance criterion (see, e.g., Wang and Gelfand, 2002), the average coverage criterion, the average length criterion, the worst outcome criterion (e.g., Joseph et al., 1995; Joseph and Bélisle, 1997), and the approach based on the range of equivalence (see, for instance, Spiegelhalter, Abrams, and Myles, 2004) for superiority/noninferiority trials. Other criteria have also been considered in the literature, including Lindley (1997), Pham-Gia (1997), Lam and Lam (1997), and M'Lan et al. (2006, 2008). However, most of the aforementioned Bayesian articles do not directly address design and analysis of noninferiority trials except for Spiegelhalter et al. (2004).

The rest of the article is organized as follows. In Section 2, we present the design of a noninferiority trial with two treatment arms for evaluating the performance of a new generation of medical devices to motivate the methodology developed in this article. The availability of historical data from first generation medical devices is also discussed in detail. In Section 3, we propose a general framework of Bayesian SSD for

designing a noninferiority trial. Section 4 provides a detailed development of the incorporation of historical data via the hierarchical modeling approach as well as the power prior formulation. The posterior distribution is discussed and a simulation-based computational algorithm is developed in Section 5. In Section 6, we apply the proposed methodology to SSD of the noninferiority medical device trial discussed in Section 2. The proposed Bayesian SSD method is compared to frequentist SSD methods. We show that Bayesian SSD yields a substantial reduction in the sample size compared to a frequentist design. We conclude the article with some discussion and extension of the proposed Bayesian SSD method in Section 7.

2. Design of a Noninferiority Trial with Two Treatment Arms for Medical Devices

We consider designing a clinical trial to evaluate the performance of a new generation of drug-eluting stent (DES) ("test device") with a noninferiority comparison to the first generation of DES ("control device"). Thus, the trial has two arms: test device and control device. The primary endpoint is the 12-month target lesion failure (TLF) (binary) composite endpoint, which is an ischemia-driven revascularization of the target lesion, myocardial infarction (Q-wave and non-Q-wave) related to the target vessel, or (cardiac) death related to the target vessel. The secondary endpoint is the 9-month in-segment percentage diameter stenosis (%DS) (continuous), which is the percentage of narrowing in the coronary artery caused by the plaque. Let $\boldsymbol{y}_t^{(n_t)} = (y_{t1}, y_{t2}, \ldots, y_{tn_t})'$ and $\boldsymbol{y}_c^{(n_c)} = (y_{c1}, y_{c2}, \ldots, y_{cn_c})'$ be the data corresponding to the test device and the control device, respectively, collected from this trial. Let $n = n_t + n_c$ denote the total sample size. Also, we write $\boldsymbol{y}^{(n)} = ((\boldsymbol{y}_t^{(n_t)})', (\boldsymbol{y}_c^{(n_c)})')'$. We assume that the ratio of two sample sizes, $r = \frac{n_c}{n_t}$, is fixed. Thus, $n_t = \frac{n}{1+r}$ and $n_c = \frac{rn}{1+r}$. We choose r to be small, for example, $r = 1/4$, so that $n_t > n_c$. The goal of the trial is to show that the test device is noninferior to the control device.

We assume that $\boldsymbol{y}_t^{(n_t)}$ and $\boldsymbol{y}_c^{(n_c)}$ are two independent random samples. For the primary endpoint, we assume that $y_{ti}(y_{ci})$ follows a Bernoulli distribution $\text{Ber}(p_t)$ $(\text{Ber}(p_c))$. Let $\mu_t = \log\left(\frac{p_t}{1-p_t}\right)$ and $\mu_c = \log\left(\frac{p_c}{1-p_c}\right)$. For the secondary endpoint, we assume that $y_{ti} \sim N(\mu_t, \sigma^2)$ and $y_{ci} \sim N(\mu_c, \sigma^2)$ independently. Let $\boldsymbol{\theta} = (\mu_t, \mu_c)$ for the primary endpoint and $\boldsymbol{\theta} = (\mu_t, \mu_c, \sigma^2)$ for the secondary endpoint. Then, the joint distribution of $\boldsymbol{y}^{(n)}$ for the primary endpoint is given by

$$f(\boldsymbol{y}^{(n)} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_t} \frac{\exp(y_{ti}\mu_t)}{1+\exp(\mu_t)} \times \prod_{i=1}^{n_c} \frac{\exp(y_{ci}\mu_c)}{1+\exp(\mu_c)}, \quad (1)$$

and the joint distribution of $\boldsymbol{y}^{(n)}$ for the secondary endpoint is given by

$$f(\boldsymbol{y}^{(n)} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_t} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_{ti}-\mu_t)^2\right\} \times \prod_{i=1}^{n_c} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_{ci}-\mu_c)^2\right\}. \quad (2)$$

Table 1
Historical data

	12-month TLF % TLF (# of failure/n_{0k})	log(9-month %DS) mean ± SD (n_{0k})
Historical Trial 1	8.2% (44/535)	3.0891 ± 0.6315 (242)
Historical Trial 2	10.9% (33/304)	3.1849 ± 0.5811 (263)

The design parameter is the difference between μ_t and μ_c, namely, $\mu_t - \mu_c$, and the hypotheses for noninferiority testing are

$$H_0: \ \mu_t - \mu_c \geq \delta \ \text{ versus } \ H_1: \ \mu_t - \mu_c < \delta, \quad (3)$$

where δ is a prespecified noninferiority margin. The trial is successful if H_1 is accepted.

Historical data are available from two previous trials on the first generation of DES. The first trial conducted in 2002 evaluated the safety and effectiveness of the slow release paclitaxel-eluting stent for treatment of de novo coronary artery lesions. The second trial conducted in 2004 expanded on the first trial, studied more complex de novo lesions, and involved multiple overlapping stents and smaller and larger diameter stents. Our historical data based on lesion size matched criteria are subsets of the data published in Stone et al. (2004, 2005). A summary of the historical data is given in Table 1. In Table 1, SD stands for standard deviation.

In the next two sections, we will develop the general methodology for Bayesian SSD and elicit priors via historical data.

3. The General Methodology

We first develop a new but general method to determine Bayesian sample size for a noninferiority trial. Denote the data associated with a sample size of n by $\boldsymbol{y}^{(n)}$ and let $\boldsymbol{\theta}$ be the vector of all the model parameters. Then, the joint distribution of $\boldsymbol{y}^{(n)}$ and $\boldsymbol{\theta}$ is written as $f(\boldsymbol{y}^{(n)} \mid \boldsymbol{\theta})\ \pi(\boldsymbol{\theta})$, where $\pi(\boldsymbol{\theta})$ denotes the prior distribution. Let $h(\boldsymbol{\theta})$ be a scalar function that measures the "true" size of the treatment effect. Then, let δ denote the noninferiority margin. Similar to Hung et al. (2003), we assume that the hypotheses for noninferiority testing can be formulated as follows:

$$H_0: \ h(\boldsymbol{\theta}) \geq \delta \ \text{ versus } \ H_1: \ h(\boldsymbol{\theta}) < \delta. \quad (4)$$

Consequently, we let Θ_0 and Θ_1 denote the parameter spaces corresponding to H_0 and H_1. For the hypotheses given in (3), $h(\boldsymbol{\theta}) = \mu_t - \mu_c; \Theta_0 = \{\boldsymbol{\theta} = (\mu_t, \mu_c): \ \mu_t - \mu_c \geq \delta\}$ and $\Theta_1 = \{\boldsymbol{\theta}: \mu_t - \mu_c < \delta\}$ for the primary endpoint; and $\Theta_0 = \{\boldsymbol{\theta} = (\mu_t, \mu_c, \sigma^2): \ \mu_t - \mu_c \geq \delta, \sigma^2 > 0\}$ and $\Theta_1 = \{\boldsymbol{\theta}: \mu_t - \mu_c < \delta, \sigma^2 > 0\}$ for the secondary endpoint.

Following Wang and Gelfand (2002), let $\pi^{(s)}(\boldsymbol{\theta})$ denote the sampling prior and also let $\pi^{(f)}(\boldsymbol{\theta})$ denote the fitting prior. The sampling prior, which captures a certain specified portion of the parameter space in achieving a certain level of performance in SSD, is used to generate the data whereas the fitting prior is used to fit the model once the data are obtained. We note that $\pi^{(f)}(\boldsymbol{\theta})$ may be improper as long as the resulting posterior, $\pi^{(f)}(\boldsymbol{\theta} \mid \boldsymbol{y}^{(n)}) \propto f(\boldsymbol{y}^{(n)} \mid \boldsymbol{\theta})\pi^{(f)}(\boldsymbol{\theta})$, is proper. Further we let $f^{(s)}(\boldsymbol{y}^{(n)})$ denote the marginal distribution that

is induced from the sampling prior. Now, we introduce the key quantity

$$\beta_s^{(n)} = E_s[1\{P(h(\boldsymbol{\theta}) < \delta \,|\, \boldsymbol{y}^{(n)}, \pi^{(f)}) \geq \gamma\}], \tag{5}$$

where the indicator function $1\{A\}$ is 1 if A is true and 0 otherwise, $\gamma > 0$ is a prespecified quantity, the probability is computed with respect to the posterior distribution given the data $\boldsymbol{y}^{(n)}$ and the fitting prior $\pi^{(f)}(\boldsymbol{\theta})$, and the expectation is taken with respect to the marginal distribution of $\boldsymbol{y}^{(n)}$ under the sampling prior $\pi^{(s)}(\boldsymbol{\theta})$.

Now, we propose a new Bayesian SSD algorithm as follows. Let $\overline{\Theta}_0$ and $\overline{\Theta}_1$ denote the closures of Θ_0 and Θ_1. Let $\pi_0^{(s)}(\boldsymbol{\theta})$ denote a "sampling prior" with support $\Theta_B = \overline{\Theta}_0 \cap \overline{\Theta}_1$. Also let $\pi_1^{(s)}(\boldsymbol{\theta})$ denote a "sampling prior" with support $\Theta_1^* \subset \Theta_1$. For given $\alpha_0 > 0$ and $\alpha_1 > 0$, we compute

$$n_{\alpha_0} = \min\left\{n : \ \beta_{s0}^{(n)} \leq \alpha_0\right\} \text{ and } n_{\alpha_1} = \min\left\{n : \ \beta_{s1}^{(n)} \geq 1 - \alpha_1\right\}, \tag{6}$$

where $\beta_{s0}^{(n)}$ and $\beta_{s1}^{(n)}$ given in (5) corresponding to $\pi^{(s)} = \pi_0^{(s)}$ and $\pi^{(s)} = \pi_1^{(s)}$ are the Bayesian type I error and power, respectively. Then, the Bayesian sample size is given by $n_B = \max\{n_{\alpha_0}, n_{\alpha_1}\}$. According to the FDA 2010 Guidance, we choose $\gamma \geq 0.95$. Common choices of α_0 and α_1 include $\alpha_0 = 0.05$ and $\alpha_1 = 0.20$ so that the Bayesian 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. In addition, for a given sample size n_B, the operating characteristic curve can be constructed by varying Θ_1^* inside of Θ_1. If $h(\boldsymbol{\theta})$ is a monotonic function of the distance between Θ_1^* and Θ_B, then the further Θ_1^* is away from Θ_B, the higher the power will be.

A simple illustration: independent and identically distributed (i.i.d.) normal case. Suppose $y_1, y_2, \ldots, y_n$ are i.i.d. $N(\theta, \sigma^2)$, where σ^2 is a known variance parameter. Suppose the hypotheses for noninferiority testing are formulated as follows: H_0: $\theta \geq \delta$ versus H_1: $\theta < \delta$. We specify an improper uniform fitting prior for θ, i.e., $\pi^{(f)}(\theta) \propto 1$. In addition, we specify two point mass sampling priors for θ such that $\pi_0^{(s)}(\theta) = 1$ if $\theta = \delta$ and $\pi_1^{(s)}(\theta) = 1$ if $\theta = 0$. After some algebra, we can show that (i) a necessary condition for achieving a type I error rate of α_0 is $1 - \gamma \leq \alpha_0$ and (ii) if $1 - \gamma \leq \alpha_0$, the Bayesian sample size is the smallest integer n_B satisfying $n_B \geq \frac{\sigma^2}{\delta^2}[\Phi^{-1}(1 - \alpha_1) + \Phi^{-1}(\gamma)]^2$, where Φ denotes the $N(0, 1)$ cumulative distribution function. It is interesting to note that for this simple case, $\beta_0^{(n)} \leq \alpha_0$ always holds for all n when $1 - \gamma \leq \alpha_0$. We also note that the Bayesian sample size n_B is identical to the classical sample size formulation for a one-sided alternative hypothesis when $\alpha_0 = 1 - \gamma$.

4. The Incorporation of Historical Data in Bayesian SSD

Historical data are often available *only* for the control medical device. Now suppose that there are K historical datasets for the control device, denoted by $\boldsymbol{y}_{c0k} = (y_{c0k1}, \ldots, y_{c0kn_{0k}})'$ for $k = 1, \ldots, K$. Let $\boldsymbol{y}_{c0} = (\boldsymbol{y}'_{c01}, \ldots, \boldsymbol{y}'_{c0K})'$ denote all K historical datasets. We develop two approaches, namely, the hierarchical prior and the power prior, to incorporate the historical data $\boldsymbol{y}_{c0}$.

4.1 *Hierarchical Priors*

Under the hierarchical Bernoulli/normal model, we assume that $\boldsymbol{y}_{c0k}$ follows the same model given in either (1) or (2). Let $\boldsymbol{\theta}_0 = (\mu_{c01}, \ldots, \mu_{c0K})'$ (or $\boldsymbol{\theta}_0 = (\mu_{c01}, \ldots, \mu_{c0K}, \ \sigma^2)'$) for the primary (or secondary) endpoint. Then, the joint distribution of $\boldsymbol{y}_{c0}$ is given by $f(\boldsymbol{y}_{c0} \,|\, \boldsymbol{\theta}_0) = \prod_{k=1}^K \prod_{i=1}^{n_{0k}} \times \frac{\exp(y_{c0ki}\mu_{c0k})}{1+\exp(\mu_{c0k})}$ for the primary endpoint and $f(\boldsymbol{y}_{c0} \,|\, \boldsymbol{\theta}_0) = \prod_{k=1}^K \times \prod_{i=1}^{n_{0k}} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(y_{c0ki} - \mu_{c0k})^2\}$ for the secondary endpoint. We further assume $\mu_c \sim N(\mu_{c0}, \ \tau^2)$, where $\tau^2 > 0$, and independently $\mu_{c0k} \sim N(\mu_{c0}, \ \tau^2)$ for $k = 1, \ldots, K$.

Let $\boldsymbol{\theta}^* = (\mu_t, \ \mu_c, \ \boldsymbol{\theta}_0, \ \mu_{c0}, \ \tau^2)'$. Then, the hierarchical prior for $\boldsymbol{\theta}^*$ is given by

$$\pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0}) \propto f(y_{c0} \,|\, \boldsymbol{\theta}_0)\phi(\mu_c \,|\, \mu_{c0}, \tau^2) \times \prod_{k=1}^K \phi(\mu_{c0k} \,|\, \mu_{c0}, \tau^2)\pi_0(\mu_t, \sigma^2, \mu_{c0}, \tau^2), \tag{7}$$

where $\phi(\cdot|\mu_{c0}, \ \tau^2)$ denotes the probability density function (pdf) of a $N(\mu_{c0}, \ \tau^2)$ distribution. In (7), $\pi_0(\mu_t, \ \sigma^2, \ \mu_{c0}, \ \tau^2)$ is the initial prior, which is specified as $\pi_0(\mu_t, \sigma^2, \mu_{c0}, \tau^2) \propto \frac{1}{\sigma^2}(\tau^2)^{-(\xi_0+1)} \exp(-\eta_0/\tau^2)$, where $\xi_0 > 0$ and $\eta_0 > 0$ are two prespecified hyperparameters. The joint prior in (7) is improper because an improper uniform prior is assumed for μ_t and the historical data are borrowed for μ_c and σ^2 via the hierarchical model. Finally, the fitting prior is obtained after integrating out $\mu_{c01}, \ldots, \mu_{c0K}$, μ_{c0}, and τ^2 from (7). Specifically, we have

$$\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}_{c0}) \propto \int \pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0}) d\mu_{c01} \ldots d\mu_{c0K} \, d\mu_{c0} d\tau^2. \tag{8}$$

To specify the sampling prior $\pi^{(s)}(\boldsymbol{\theta})$, we assume μ_t, μ_c, and σ^2 are independent and then specify point mass priors for μ_t and μ_c and use the historical data to specify the sampling prior for σ^2. Specifically, we take

$$\pi^{(s)}(\boldsymbol{\theta}) = \pi^{(s)}(\mu_t)\pi^{(s)}(\mu_c) \text{ or } \pi^{(s)}(\boldsymbol{\theta}) = \pi^{(s)}(\mu_t)\pi^{(s)}(\mu_c)\pi^{(s)}(\sigma^2), \tag{9}$$

where $\pi^{(s)}(\sigma^2) \propto \int f(\boldsymbol{y}_{c0} \,|\, \boldsymbol{\theta}_0) \ [\prod_{k=1}^K \ \phi(\mu_{c0k} \,|\, \mu_{c0}, \tau^2) \ \times \ \frac{1}{\sigma^2} \times (\tau^2)^{-(\xi_0^{(s)}+1)} \exp(-\eta_0^{(s)}/\tau^2)] d\mu_{c01} \ldots d\mu_{c0K} \, d\mu_{c0} d\tau^2$, and $\xi_0^{(s)} > 0$ and $\eta_0^{(s)} > 0$ are prespecified hyperparameters, which may be different from (ξ_0, η_0). As discussed in Section 3, the sampling prior must be proper. We can show that under very mild conditions, the sampling prior $\pi^{(s)}(\sigma^2)$ is proper.

We note that under the normal model, the hierarchical prior (7) for $\boldsymbol{\theta}^*$ reduces to

$$\pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0}) \propto (\sigma^2)\tau^{\frac{1}{2}\sum_{k=1}^{K} n_{0k}-1} \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\left[n_{0k}(\mu_{c0k}-\bar{y}_{c0k})^2+(n_{0k}-1)S_{0k}^2\right]\right\} \times(\tau^2)^{-\{\xi_0+(K+1)/2+1\}} \times \exp\left\{-\frac{1}{\tau^2}\left[\eta_0+\frac{1}{2}(\mu_c-\mu_{c0})^2+\frac{1}{2}\sum_{k=1}^{K}(\mu_{c0k}-\mu_{c0})^2\right]\right\}, \tag{10}$$

where $\bar{y}_{c0k} = (1/n_{0k})\sum_{i=1}^{n_{0k}} y_{c0ki}$ and $S_{0k}^2 = [1/(n_{0k}-1)] \times \sum_{i=1}^{n_{0k}}(y_{c0ki}-\bar{y}_{c0k})^2$ for $k = 1, \ldots, K$. Thus, the fitting prior and the sampling prior depend only on the sufficient statistics $\{(\bar{y}_{0k}, S_{0k}^2),\ k = 1, \ldots, K\}$ from the historical data.

4.2 *Power Priors*

We extend the power priors of Ibrahim and Chen (2000) to build the prior distribution for μ_c or (μ_c, σ^2) when multiple historical datasets are available. For the primary endpoint, we consider the following *normalized power prior* for μ_c given multiple historical data $\boldsymbol{y}_{c0}$,

$$\pi(\mu_c \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = \frac{1}{C(\boldsymbol{a}_0)}\prod_{k=1}^{K}\left[\prod_{i=1}^{n_{0k}}\frac{\exp(y_{c0ki}\mu_c)}{1+\exp(\mu_c)}\right]^{a_{0k}}\pi_0(\mu_c), \tag{11}$$

where $\boldsymbol{a}_0 = (a_{01}, \ldots, a_{0K})'$, $0 \le a_{0k} \le 1$ for $k = 1, 2, \ldots, K$, $\pi_0(\mu_c)$ is an initial prior, and $C(\boldsymbol{a}_0) = \int_0^\infty \prod_{k=1}^{K}[\prod_{i=1}^{n_{0k}} \times \frac{\exp(y_{c0ki}\mu_c)}{1+\exp(\mu_c)}]^{a_{0k}}\pi_0(\mu_c)\,d\mu_c$. When $\pi_0(\mu_c) \propto 1$, (11) reduces to

$$\pi(\mu_c \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = \frac{\exp\left\{\mu_c\sum_{k=1}^{K} a_{0k}\, n_{0k}\, \bar{y}_{c0k}\right\}}{B\left(\sum_{k=1}^{K} a_{0k}\, n_{0k}\, \bar{y}_{c0k}, \sum_{k=1}^{K} a_{0k}\, n_{0k}\,(1-\bar{y}_{c0k})\right)[1+\exp(\mu_c)]^{n_0(\boldsymbol{a}_0)}},$$

where $B(.,\ .)$ denotes the complete beta function, $n_0(\boldsymbol{a}_0) = \sum_{k=1}^{K} a_{0k}\, n_{0k}$, and $\bar{y}_{c0k}$ is defined in (10). For the secondary endpoint, the normalized power prior for μ_c and σ^2 is given by

$$\pi(\mu_c, \sigma^2 \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = \frac{1}{C(\boldsymbol{a}_0)}\prod_{k=1}^{K}\left[\prod_{i=1}^{n_{0k}}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y_{c0ki}-\mu_c)^2\right\}\right]^{a_{0k}}\pi_0(\mu_c, \sigma^2), \tag{12}$$

where $\pi_0(\mu_c,\ \sigma^2)$ is an initial prior and $C(\boldsymbol{a}_0)$ is the normalizing constant, which is similar to the one in (11). Let $\bar{y}_{c0}(\boldsymbol{a}_0) = \frac{\sum_{k=1}^{K} a_{0k}\, n_{0k}\, \bar{y}_{c0k}}{n_0(\boldsymbol{a}_0)}$ and $S_0^2(\boldsymbol{a}_0) = \sum_{k=1}^{K} a_{0k} n_{0k}(\bar{y}_{c0k} - \bar{y}_{c0}(\boldsymbol{a}_0))^2 + \sum_{k=1}^{K} a_{0k}(n_{0k}-1)S_{0k}^2$, where $\bar{y}_{c0k}$ and S_{0k}^2 are defined in (10). When $\pi_0(\mu_c,\ \sigma^2) \propto 1/\sigma^2$, (12) reduces to $\pi(\mu_c, \sigma^2 \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = (\frac{n_0(\boldsymbol{a}_0)}{2\pi\sigma^2})^{1/2}\exp\{-\frac{n_0(\boldsymbol{a}_0)}{2\sigma^2}[\mu_c - \bar{y}_{c0}(\boldsymbol{a}_0)]^2\} \times \{[S_0^2(\boldsymbol{a}_0)/2]^{[n_0(\boldsymbol{a}_0)+1]/2}/\Gamma([n_0(\boldsymbol{a}_0)+1]/2)\}(\sigma^2)^{-\frac{1}{2}[n_0(\boldsymbol{a}_0)+1]} \times \exp\{-S_0^2(\boldsymbol{a}_0)/(2\sigma^2)\}$. To complete the specification of the power prior, we assume that the a_{0k}'s are independent and distributed as $a_{0k} \sim \text{beta}(b_{01},\ b_{02})$, where $b_{01} > 0$ and $b_{02} > 0$ are prespecified hyperparameters. We mention that the normalized power prior is also considered by Duan, Ye, and Smith (2006); Neuenschwander, Branson, and Spiegelhalter (2009); and Hobbs et al. (2009).

Using (11) or (12), the fitting prior of $\boldsymbol{\theta}$ is of the form

$$\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}_{c0}) \propto \int\left[\pi(\tilde{\boldsymbol{\theta}} \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)\prod_{k=1}^{K} a_{0k}^{b_{01}-1}(1-a_{0k})^{b_{02}-1}\right]d\boldsymbol{a}_0\pi_0(\mu_c), \tag{13}$$

where $\pi(\tilde{\boldsymbol{\theta}} \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = \pi(\mu_c \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)$ defined in (11) and $\tilde{\boldsymbol{\theta}} = \mu_c$ for the primary endpoint, and $\pi(\tilde{\boldsymbol{\theta}} \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0) = \pi(\mu_c, \sigma^2 \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)$ defined in (12) and $\tilde{\boldsymbol{\theta}} = (\mu_c, \sigma^2)$ for the secondary endpoint. Similar to the hierarchical prior, the sampling prior $\pi^{(s)}(\boldsymbol{\theta})$ under the normalized power prior is specified as follows: $\pi^{(s)}(\boldsymbol{\theta}) = \pi^{(s)}(\mu_t)\pi^{(s)}(\mu_c)$ or $\pi^{(s)}(\boldsymbol{\theta}) = \pi^{(s)}(\mu_t)\ \pi^{(s)}(\mu_c)\ \pi^{(s)}(\sigma^2)$, where $\pi^{(s)}(\mu_t)$ and $\pi^{(s)}(\mu_c)$ are two prespecified proper priors,

$$\pi^{(s)}(\sigma^2) \propto \int \pi(\tilde{\boldsymbol{\theta}} \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_{0s})\pi_0^{(s)}(\sigma^2)\,d\mu_c, \tag{14}$$

$\boldsymbol{a}_{0s}$ is prespecified and $\pi_0^{(s)}(\sigma^2)$ may be an improper initial prior such as $\pi_0^{(s)}(\sigma^2) \propto 1/\sigma^2$.

In (11) or (12), the parameter a_{0k} controls the influence of the kth historical dataset on $\pi(\tilde{\boldsymbol{\theta}} \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)$. The parameter a_{0k} can be interpreted as a relative precision parameter for the kth historical dataset. One of the main roles of $\boldsymbol{a}_0$ is that it controls the heaviness of the tails of the prior for μ_c (or (μ_c, σ^2)). As all of the a_{0k}'s become smaller, the tails of (11) or (12) become heavier. When $a_{0k} = 1$ for all k with probability 1, (13) corresponds to the update of $\pi_0(\boldsymbol{\theta})$ using Bayes' theorem based on the historical data. When $\boldsymbol{a}_0 = \boldsymbol{0}$ with probability 1, then the power prior does not depend on the historical data. That is, $\boldsymbol{a}_0 = \boldsymbol{0}$ is equivalent to a prior specification with no incorporation of historical data. Thus, the a_{0k}'s control the influence of the multiple historical datasets on the current study. Such control is important in cases where there is heterogeneity among the historical studies, or heterogeneity between the historical and current studies, or when the sample sizes of the historical and current studies are quite different.

We note that the use of historical data via the power priors for Bayesian SSD is also considered by De Santis (2007). We also note that $\boldsymbol{a}_0$ may be considered to be fixed instead of random. For ease of exposition, we consider the primary endpoint. When $\boldsymbol{a}_0$ is fixed, the fitting prior of $\boldsymbol{\theta} = (\mu_t,\ \mu_c)'$ is of the form $\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}_{c0}) \propto \pi(\mu_c \,|\, \boldsymbol{y}_{c0},\ \boldsymbol{a}_0)\ \pi_0(\mu_t)$, where $\pi(\mu_c \,|\, \boldsymbol{y}_{c0},\ \boldsymbol{a}_0)$ is given by (11), $\pi_0(\mu_t)$ is an initial prior for μ_t, and $\boldsymbol{a}_0$ is fixed. When $\boldsymbol{a}_0$ is fixed, we know exactly how much historical data are incorporated in the new trial, and in addition, there is a theoretical connection between the power prior formulation and the hierarchical prior specification as established in Chen and Ibrahim (2006). De Santis (2006) also provides

some useful comments on the fixed-$\boldsymbol{a}_0$ case as well as on power priors for the exponential family. On the other hand, when $\boldsymbol{a}_0$ is random, the amount of incorporation of historical data is determined by the data and hence not prespecified by the data analyst.

5. Posteriors and Computations

For ease of exposition, we only consider the primary endpoint. Instead of directly sampling from $\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0}) \propto f(\boldsymbol{y}^{(n)} \,|\, \boldsymbol{\theta})\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}_{c0})$, where $f(\boldsymbol{y}^{(n)} \,|\, \boldsymbol{\theta})$ is given by (1) and $\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}_{c0})$ is defined in (8) or (13), we consider the augmented fitting posterior distribution parameters $\boldsymbol{\theta}^*$, where $\boldsymbol{\theta}^* = (\mu_t, \mu_c, \mu_{c01}, \ldots, \mu_{c0K}, \mu_{c0}, \tau^2)'$ for the hierarchical prior and $\boldsymbol{\theta}^* = (\mu_t, \mu_c, \boldsymbol{a}_0)$ for the normalized power prior. Then, the augmented fitting posterior distribution of $\boldsymbol{\theta}^*$ is given by $\pi^{(f)}(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0}) \propto f(\boldsymbol{y}^{(n)} \,|\, \boldsymbol{\theta})\ \pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0})$, where $\pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0})$ is defined in (7) under the hierarchical prior, and $\pi(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}_{c0}) \propto \pi(\mu_c \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)[\prod_{k=1}^{K} a_{0k}^{b_{01}-1}(1-a_{0k})^{b_{02}-1}]\pi_0(\mu_t)$ with $\pi(\mu_c \,|\, \boldsymbol{y}_{c0}, \boldsymbol{a}_0)$ defined in (11) under the normalized power prior. Although the posterior distribution $\pi^{(f)}(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0})$ is analytically intractable, sampling from this distribution via the Gibbs sampler is quite straightforward, because the conditional posterior distributions of the components of $\boldsymbol{\theta}^*$ (except for $\boldsymbol{a}_0$) are either known distributions or log concave. For $\boldsymbol{a}_0$, we use the localized Metropolis algorithm discussed in Chen, Shao, and Ibrahim (2000) to sample from its conditional posterior distribution.

Let $\{\boldsymbol{\theta}^{*(m)}, \quad m = 1, 2, \ldots, M\}$ denote a Gibbs sample from the augmented fitting posterior distribution $\pi^{(f)}(\boldsymbol{\theta}^* \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0})$. As $\boldsymbol{\theta}$ is a subvector of $\boldsymbol{\theta}^*$, let $\boldsymbol{\theta}^{(m)}$ denote the corresponding components of $\boldsymbol{\theta}^{*(m)}$ from the mth Gibbs iteration. Then, it is easy to show that $\{\boldsymbol{\theta}^{(m)}, m = 1, 2, \ldots, M\}$ is a Gibbs sample from the fitting posterior distribution $\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0})$. Using this Gibbs sample, a Monte Carlo estimate of $P(h(\boldsymbol{\theta}) < \delta \,|\, \boldsymbol{y}^{(n)}, \pi^{(f)})$ is given by

$$\hat{P}_f = \frac{1}{M}\sum_{m=1}^{M} 1\{h(\boldsymbol{\theta}^{(m)}) < \delta\}. \tag{15}$$

To compute $\beta_s^{(n)}$ in (5), we propose the following computational algorithm: Step 0: Specify n_t, n_c, δ, γ, and N; Step 1: Generate $\boldsymbol{\theta} \sim \pi^{(s)}(\boldsymbol{\theta})$; Step 2: Generate $\boldsymbol{y}^{(n)} \sim f(\boldsymbol{y}^{(n)} \,|\, \boldsymbol{\theta})$; Step 3: Run the Gibbs sampler to generate a Gibbs sample $\{\boldsymbol{\theta}^{(m)}, m = 1, 2, \ldots, M\}$ of size M from the fitting posterior distribution $\pi^{(f)}(\boldsymbol{\theta} \,|\, \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0})$; Step 4: Compute $\hat{P}_f$ via (14); Step 5: Check whether $\hat{P}_f \geq \gamma$; Step 6: Repeat steps 1–5 N times; and Step 7: Compute the proportion of $\{\hat{P}_f \geq \gamma\}$ in these N runs, which gives an estimate of $\beta_s^{(n)}$.

6. Applications to Medical Device Trials

We apply the proposed methodology in designing the noninferiority clinical trial for medical devices discussed in Section 2. We use the historical datasets given in Table 1 to construct our priors in Bayesian SSD. We set $\gamma = 0.95$, which implies a target type I error of 0.05. We notice that the same γ value was also used in Allocco et al. (2010). In all of the computations below, $N = 10{,}000$ and $M = 20{,}000$ were used.

Bayesian SSD for TLF. For the primary endpoint, the margin was set to be $\delta = \text{logit}(4.1\%) = \log\{\frac{0.041}{1-0.041}\}$. We took $(\xi_0, \eta_0) = (0.01, 0.01)$ or $(\xi_0, \eta_0) = (0.001, 0.001)$ for the initial prior of γ in the fitting prior (8), $\pi_0(\mu_c) \propto 1$ and $b_{01} = b_{02} = 1$ for the initial priors of μ_c and a_{0k} in (13). We computed the powers at $\mu_t = \mu_c$ and the type I error at $\frac{\exp(\mu_t)}{1+\exp(\mu_t)} = \frac{\exp(\mu_c)}{1+\exp(\mu_c)} + \frac{\exp(\delta)}{1+\exp(\delta)}$. In other words, we convert μ_t and μ_c back to p_t and p_c in the Bernoulli case. In the sampling prior (9), we assumed a point mass prior at $\mu_c = \text{logit}(9.2\%)$ for $\pi^{(s)}(\mu_c)$, where 9.2% was the pooled proportion for the two historical control datasets, and a point mass prior at $\mu_t = \mu_c$ or $\mu_t = \text{logit}[\frac{\exp(\mu_c)}{1+\exp(\mu_c)} + \frac{\exp(\delta)}{1+\exp(\delta)}]$ for $\pi^{(s)}(\mu_t)$. We first computed the powers and the type I errors for various sample sizes based on the proposed Bayesian SSD without the incorporation of historical data. Table 2 shows the results. Table 2 also presents the powers of the two frequentist methods, namely, the z-test with unpooled variances and the score test (Farrington and Manning, 1990) for noninferiority trials. For both frequentist methods, the target type I error was 0.05. In all calculations, the margin $\delta = 0.041$, $p_c = 9.2\%$, and a 3:1 sample size ratio were used. `PASS 2008` (Hintze, 2008) was used for computing the powers for the two frequentist SSD methods. We see from Table 2 that the proposed Bayesian SSD without incorporation of historical data gives very similar powers compared to the score test for the frequentist SSD, whereas the type I errors of the Bayesian SSD are controlled at or below 5%. Both the score test and Bayesian SSD yield slightly higher powers than the z-test. To achieve 80% power, the z-test requires a total sample size of 1636 with $n_t = 1227$ and $n_c = 409$.

Table 2 also shows the powers and the type I errors of the Bayesian SSD procedure with hierarchical priors and power priors with fixed and random $\boldsymbol{a}_0$. The hierarchical prior with $(\xi_0, \eta_0) = (0.001, 0.001)$ leads to higher powers than the one with $(\xi_0, \eta_0) = (0.01, 0.01)$. In addition, the powers based on the power prior with $\boldsymbol{a}_0$ random are comparable to those based on the hierarchical prior with $(\xi_0, \eta_0) = (0.001, 0.001)$ and the power prior with $\boldsymbol{a}_0$ fixed at $\boldsymbol{a}_0 = (0.3, 0.3)$. These results imply that the power prior with random $\boldsymbol{a}_0$ and the hierarchical prior with $(\xi_0, \eta_0) = (0.001, 0.001)$ borrow approximately 30% of the historical data. With incorporation of the historical data, a sample size of $(n_t, n_c) = (810, 270)$ achieves 80% power. However, based on the frequentist SSD or the Bayesian SSD without incorporation of historical data, a sample size of 1480 with $n_t = 1110$ and $n_c = 370$ is required to achieve 80% power. Thus, the Bayesian SSD with incorporation of historical data leads to a substantial reduction in the sample size.

Bayesian SSD for %DS. For the secondary endpoint, the margin was set to be $\delta = 0.20$. We compute the power at $\mu_t = \mu_c$ and the type I error at $\mu_t = \mu_c + \delta$. In the sampling prior (9), we assume a point mass prior at $\mu_c = 3.15$ for $\pi^{(s)}(\mu_c)$ and a point mass prior at $\mu_t = \mu_c$ or $\mu_t = \mu_c + \delta$ for $\pi^{(s)}(\mu_t)$. `PASS 2008` (Hintze, 2008) was used to compute the powers of the frequentist SSD based on the pooled SD = 0.607. In the Bayesian SSD procedure that does not use any historical data, we used the same pooled SD for σ in generating the data. For the hierarchical prior, we took $(\xi_0, \eta_0) = (0.01, 0.01)$ or $(\xi_0, \eta_0) = (0.001, 0.001)$ for the initial prior of τ in the fitting prior (8) and $\xi_0 = 0.01$ and $\eta_0 = 0.01$ for the initial prior of τ

Table 2
Powers and type I errors for 12*-month TLF*

Total sample size		1000	1080	1200	1280	1480
n_t		750	810	900	960	1110
n_c		250	270	300	320	370
Frequentist SSD						
z-test (unpooled)	Power	0.617	0.646	0.685	0.710	0.764
Score test	Power	0.672	0.699	0.736	0.758	0.807
Bayesian SSD						
No borrowing	Power	0.648	0.676	0.718	0.738	0.800
$\boldsymbol{a}_0 = (0, 0)$	Type I error	0.049	0.048	0.048	0.050	0.044
Hierarchical prior	Power	0.796	0.820	0.841	0.863	0.894
$(\xi_0, \eta_0) = (0.01, 0.01)$	Type I Error	0.044	0.045	0.044	0.049	0.048
Hierarchical prior	Power	0.839	0.860	0.882	0.900	0.922
$(\xi_0, \eta_0) = (0.001, 0.001)$	Type I error	0.038	0.042	0.039	0.040	0.041
Power prior	Power	0.840	0.856	0.884	0.892	0.923
Fixed $\boldsymbol{a}_0 = (0.3, 0.3)$	Type I error	0.030	0.027	0.028	0.030	0.032
Power prior	Power	0.843	0.878	0.897	0.902	0.914
Random $\boldsymbol{a}_0$	Type I error	0.038	0.031	0.029	0.036	0.039

Table 3
Powers and type I errors for 9*-month %DS*

Total sample size		200	240	260	280	308
n_t		150	180	195	210	231
n_c		50	60	65	70	77
Frequentist SSD	Power	0.639	0.709	0.739	0.767	0.801
Bayesian SSD						
No borrowing	Power	0.644	0.699	0.747	0.769	0.805
$\boldsymbol{a}_0 = (0, 0)$	Type I error	0.051	0.049	0.051	0.050	0.048
Hierarchical prior	Power	0.710	0.773	0.800	0.820	0.847
$(\xi_0, \eta_0) = (0.01, 0.01)$	Type I error	0.037	0.038	0.040	0.038	0.039
Hierarchical prior	Power	0.791	0.837	0.871	0.877	0.899
$(\xi_0, \eta_0) = (0.001, 0.001)$	Type I error	0.023	0.024	0.025	0.028	0.027
Power prior	Power	0.812	0.864	0.880	0.899	0.918
Fixed $\boldsymbol{a}_0 = (0.08, 0.08)$	Type I error	0.022	0.023	0.026	0.027	0.028
Power prior	Power	0.805	0.857	0.878	0.893	0.913
Random $\boldsymbol{a}_0$	Type I error	0.013	0.014	0.017	0.015	0.015

in the sampling prior (9). For the power priors, we used (14) with $\boldsymbol{a}_{0s} = (0.05, 0.05)$ for the sampling prior $\pi^{(s)}(\sigma^2)$. Using the same sampling prior, we also computed the powers and type I errors with a fixed $\boldsymbol{a}_0 = (0.08, 0.08)$ in the fitting prior. The results are shown in Table 3.

Similar to TLF, the Bayesian SSD procedure with no incorporation of historical data yields similar powers, with the type I errors controlled at the 5% level, and the hierarchical prior with $(\xi_0, \eta_0) = (0.001, 0.001)$ yields higher powers than the one with $(\xi_0, \eta_0) = (0.01, 0.01)$. From Table 3, we also see that the power prior with random $\boldsymbol{a}_0$ leads to slightly higher powers than the hierarchical prior with $(\xi_0, \eta_0) = (0.001, 0.001)$, and the powers based on the power prior with random $\boldsymbol{a}_0$ are comparable to the power prior with a fixed $\boldsymbol{a}_0 = (0.08, 0.08)$. These results imply that the hierarchical prior borrows less than 8% of historical data, whereas the power prior with random $\boldsymbol{a}_0$ borrows about 8% of the historical data. Similar to TLF, the Bayesian SSD with incorporation of historical data again leads to a substantial reduction in the sample size compared to the frequentist design.

7. Discussion

In this article, we have developed a general methodology of Bayesian SSD, which is particularly suitable for designing a noninferiority clinical trial. We have discussed two types of priors, namely, the hierarchical prior and the normalized power prior, to incorporate historical data. We have shown

Table 4
Powers and type I errors under three sampling priors for 12*-month TLF with* $(n_t, n_c) = (900, 300)$

		Point mass sampling prior at μ_c = logit (p^*_c)					
		$p^*_c = 8.0\%$		$p^*_c = 9.2\%$		$p^*_c = 10.0\%$	
Fitting prior	γ	Power	Type I error	Power	Type I error	Power	Type I error
Hierarchical prior							
(ξ_0, η_0)	0.95	0.894	0.068	0.841	0.044	0.788	0.032
=(0.01, 0.01)	0.96	0.880	0.058	0.816	0.037	0.757	0.027
	0.97	0.854	0.046	0.782	0.027	0.714	0.020
Power prior with $a_{0k} \sim$ beta (b_{01}, b_{02}) in (13)							
$(b_{01}, b_{02}) = (1, 1)$	0.95	0.945	0.070	0.882	0.039	0.799	0.034
$(b_{01}, b_{02}) = (1, 5)$	0.95	0.916	0.061	0.832	0.033	0.760	0.026
$(b_{01}, b_{02}) = (1, 10)$	0.95	0.868	0.053	0.791	0.038	0.728	0.032
$(b_{01}, b_{02}) = (1, 1)$	0.96	0.935	0.055	0.880	0.022	0.765	0.026
	0.97	0.917	0.041	0.848	0.015	0.719	0.009
$(b_{01}, b_{02}) = (1, 5)$	0.96	0.899	0.047	0.803	0.027	0.722	0.021

that Bayesian SSD leads to a substantial reduction in the sample size compared to frequentist SSD. One unique feature of the proposed Bayesian SSD methodology is that we use the historical data only from the control device but not from the test device. This feature is desirable, because for the test device, historical data are often not available. Although we primarily focus on the Bernoulli and normal models in this article, our methodology is applicable to other models in the exponential family. In addition, the proposed methodology can also be extended to GLMs. The computational algorithm given in Section 5 for these two extensions is basically the same. However, there may be two potential complications. First, a closed-form expression of the normalized power prior under GLMs may not be available. Therefore, an efficient Markov chain Monte Carlo sampling algorithm needs to be developed to sample from the fitting posterior distribution in Step 3 of the computational algorithm in Section 5. Second, the determination of the noninferiority margin may be more difficult for some GLMs than the situation without covariates. For example, for binomial regression models, the noninferiority margin based on the difference in two proportions may not be easily converted to the margin on the regression coefficient corresponding to the treatment effect. However, this may not be an issue for other GLMs such as the normal linear regression model.

The proposed Bayesian SSD works best if the historical data from the control device are compatible to the data from the current trial. However, the target type I error and power may not be well maintained when the data from the historical and current trials are not compatible. For noninferiority trials, we have empirically observed that (i) the type I errors are controlled but the powers are lower when the true proportions or means in the control devices from the current trial are greater than those in the historical data; and (ii) the type I errors tend to be larger, but the powers tend to be higher when the true proportions or the true means for the control devices in the current trial are less than those in the historical data. For illustrative purposes, we consider $n = 1200$ with $n_t = 900$ and $n_c = 300$ for the primary endpoint TLF and $n = 280$ with $n_t = 210$ and $n_c = 70$ for the secondary endpoint %DS. For %DS, if a point mass sampling prior at $\mu_c = 3.10$ is assumed and $\gamma = 0.95$, the powers and type I errors are 0.836 and 0.047 for the hierarchical priors with $(\xi_0, \eta_0) = (0.01, 0.01)$ and 0.936 and 0.049 for the power prior with random $\boldsymbol{a}_0$; and if a point mass sampling prior at $\mu_c = 3.20$ is assumed, the powers and type I errors are 0.792 and 0.030 for the hierarchical priors with $(\xi_0, \eta_0) = (0.01, 0.01)$ and 0.815 and 0.004 for the power prior with random $\boldsymbol{a}_0$. In all cases, the type I errors are still controlled at 0.05. However, for TLF, the type I error is not controlled as shown in Table 4. Specifically, if a point mass sampling prior at μ_c = logit (8.0%) is assumed, the type I errors are 0.068 for the hierarchical priors with $(\xi_0, \eta_0) = (0.01, 0.01)$ and 0.07 for the power prior with $a_{0k} \sim$ beta (1, 1) and $\gamma = 0.95$. There are two approaches for resolving this type I error problem. One approach is to change the initial prior beta (b_{01}, b_{02}) for a_{0k} in (13) to downweight the historical control data as suggested by an anonymous associate editor. Another approach is to increase the value of γ, which is recommended in the FDA 2010 Guidance. As shown in Table 4 for TLF, if a point mass sampling prior at μ_c = logit (8.0%) is assumed, the type I error decreases in b_{02} when b_{01} is fixed at 1. When $(b_{01}, b_{02}) = (1, 10)$, which gives an initial prior weight of 10% to the historical control data, the type I error is 0.053. Also, we see from Table 4 that for a fixed initial prior beta (b_{01}, b_{02}), the type I error decreases in γ. In particular, when $(b_{01}, b_{02}) = (1, 1)$ and $\gamma = 0.97$, the type I error is 0.041 if a point mass sampling prior at μ_c = logit (8.0%) is assumed. A combination of these two approaches is also quite effective in controlling the type I error while maintaining good power as shown in Table 4. Further methodological approaches for controlling the type I error are currently under investigation.

Finally, we briefly discuss how to determine whether the trial is successful after it is completed. The computational algorithm developed in Section 5 can still be used for this purpose. Specifically, the following algorithm can be used to determine the outcome of the trial: Step 0: Use the same γ and the same fitting prior specified at the design stage; Step 1:

Obtain the data $\boldsymbol{y}^{(n)}$ at the completion of the trial; Step 2: Run the Gibbs sampler to generate a Gibbs sample $\{\boldsymbol{\theta}^{(m)}, \quad m = 1, 2, \ldots, M\}$ of size M from the fitting posterior distribution $\pi^{(f)}(\boldsymbol{\theta} \mid \boldsymbol{y}^{(n)}, \boldsymbol{y}_{c0})$; Step 3: Compute $\hat{P}_f$ via (14); and Step 4: Declare a success of the trial if $\hat{P}_f \geq \gamma$.

Acknowledgements

Dr M-HC is a statistical consultant for Boston Scientific Corporation. This research was partially supported by Boston Scientific Corporation. The conclusions in this article are entirely those of the authors and do not necessarily represent the views of Boston Scientific Corporation. No conflict of interest exists among the authors. In addition, Drs M-HC and JGI's research was partially supported by NIH grants GM 70335 and CA 74015.

References

Adcock, C. J. (1997). Sample size determination: A review. *The Statistician* **46,** 261–283.

Allocco, D. J., Cannon, L. A., Britt, A., Heil, J. E., Nersesov, A., Wehrenberg, S., Dawkins, K. D., and Kereiakes, D. J. (2010). A prospective evaluation of the safety and efficacy of the TAXUS element paclitaxel-eluting coronary stent system for the treatment of de novo coronary artery lesions: Design and statistical methods of the PERSEUS clinical program. *Trials* **11**:1. `http://www.trialsjournal.com/content/11/1/1`.

Chen, M.-H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis* **1,** 551–574.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

D'Agostino Sr., R. B., Massaro, J. M., and Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues—the encounters of academic consultants in statistics. *Statistics in Medicine* **22,** 169–186.

De Santis, F. (2006). Power priors and their use in clinical trials. *The American Statistician* **60,** 122–129.

De Santis, F. (2007). Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society, Series A* **170,** 95–113.

Duan, Y., Ye, K., and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17,** 95–106.

Farrington, C. P. and Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* **9,** 1447–1454.

Fleming, T. R. (2008). Current issues in non-inferiority trials. *Statistics in Medicine* **27,** 317–332.

Hintze, J. (2008). *PASS 2008. NCSS, LLC*. Kaysville, Utah. Available at `www.ncss.com`.

Hobbs, B. P., Carlin, B. P., Mandrekar, S., and Sargent, D. (2009). *Hierarchical commensurate prior models for adaptive incorporation of historical information in clinical trials*. Technical Report 2009–017. Division of Biostatistics, University of Minnesota.

Hung, H. M. J., Wang, S.-J., Tsong, Y., Lawrence, J., and O'Neill, R. T. (2003). Some fundamental issues with non-inferiority testing in active controlled trials. *Statistics in Medicine* **22,** 213–225.

Hung, H. M. J., Wang, S.-J., and O'Neill, R. T. (2005). A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biometrical Journal* **47,** 28–36.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science* **15,** 46–60.

Inoue, L. Y. T., Berry, D. A., and Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *The American Statistician* **59,** 79–87.

Joseph, L. and Bélisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *The Statistician* **46,** 209–226.

Joseph, L., Wolfson, D. B., and Du Berger, R. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *The Statistician: Journal of the Institute of Statisticians* **44,** 143–154.

Katsis, A. and Toman, B. (1999). Bayesian sample size calculations for binomial experiments. *Journal of Statistical Planning and Inference* **81,** 349–362.

Kieser, M. and Friede, T. (2007). Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in Medicine* **26,** 253–273.

Lam, Y. and Lam, C. V. (1997). Bayesian double-sampling plans with normal distributions. *The Statistician* **46,** 193–207.

Lindley, D. V. (1997). The choice of sample size. *The Statistician* **46,** 129–138.

M'Lan, C. E., Joseph, L., and Wolfson, D. B. (2006). Bayesian sample size determination for case-control studies. *Journal of the American Statistical Association* **101,** 760–772.

M'Lan, C. E., Joseph, L., and Wolfson, D. B. (2008). Bayesian sample size determination for binomial proportions. *Bayesian Analysis* **3,** 269–296.

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28,** 3562–3566.

Pham-Gia, T. (1997). On Bayesian analysis, Bayesian decision theory and the sample size problem. *The Statistician* **46,** 139–144.

Rahme, E. and Joseph, L. (1998). Exact sample size determination for binomial experiments. *Journal of Statistical Planning and Inference* **66,** 83–93.

Rothmann, M., Li, N., Chen, G., Chi, G. Y. H., Temple, R., and Tsou, H.-H. (2003). Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine* **22,** 239–264.

Rubin, D. B. and Stern, H. S. (1998). Sample size determination using posterior predictive distributions. *Sankhyā, Series B* **60,** 161–175.

Simon, R. (1999). Bayesian design and analysis of active control clinical trials. *Biometrics* **55,** 484–487.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: Wiley.

Stone, G. W., Ellis, S. G., Cox, D. A., Hermiller, J., O'Shaughnessy, C. D., Mann, J. T., Turco, M., Caputo, R., Bergin, P., Greenberg, J. D., Popma, J. J., and Russell, M. E., for the TAXUS-IV Investigators. (2004). A polymer-based, paclitaxel-eluting stent in patients with coronary artery disease. *The New England Journal of Medicine* **350,** 221–231.

Stone, G. W., Ellis, S. G., Cannon, L., Mann, J. T., Greenberg, J. D., O'Shaughnessy, C. D., DeMaio, S., Hall, P., Popma, J. J., Koglin, J., and Russell, M. E., for the TAXUS V Investigators. (2005). Comparison of a polymer-based paclitaxel-eluting stent with a bare metal stent in patients with complex coronary artery disease: A randomized controlled trial. *Journal of the American Medical Association* **294,** 1215–1223.

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

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *The Statistician* **46,** 185–191.

Received March 2010. *Revised November* 2010.
Accepted November 2010.