# Chapter 2
# Bayesian Design of Noninferiority Clinical Trials with Co-primary Endpoints and Multiple Dose Comparison

**Wenqing Li, Ming-Hui Chen, Huaming Tan and Dipak K. Dey**

**Abstract** We develop a Bayesian approach for the design of noninferiority clinical trials with co-primary endpoints and multiple dose comparison. The Bayesian approach has the potential of power increase and hence sample size reduction due to the incorporation of the historical data and the correlation structure among multiple co-primary endpoints while it still maintains the family-wise type I error control without additional multiplicity adjustment. In this chapter, we compare the Bayesian method to the conventional frequentist method with or without Bonferroni multiplicity adjustment resulting from the multiple dose comparison. The proposed method is also applied to the design of a clinical trial, in which the study drug at a low dose level and at a high dose level is compared with the active control in terms of the bivariate co-primary endpoints.

## 2.1 Introduction

A noninferiority clinical trial is often designed to demonstrate that a test treatment is not worse than an active control or the current standard of care (SOC). Phase III confirmatory clinical trials are recently seen to be conducted via noninferiority trials in comparison with an active comparator for various reasons, including ethic compliance, comparison effectiveness, benefit and risk assessment. Details

W. Li (✉)
Global Biostatistical Science, Amgen Inc., 1 Amgen Center Dr., Thousand Oaks, CA 91320, USA
e-mail: Wenqingl@amgen.com

M. Chen · D. K. Dey
Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269, USA
e-mail: ming-hui.chen@uconn.edu

D. K. Dey
e-mail: dipak.dey@uconn.edu

H. Tan
Clinical Statistics, Global Innovative Pharma Business, Pfizer Inc., 445 Eastern Point Road, Groton, CT 06340, USA
e-mail: huaming.tan@pfizer.com



Z. Chen et al. (eds.), *Applied Statistics in Biomedicine and Clinical Trials Design,*
ICSA Book Series in Statistics, DOI 10.1007/978-3-319-12694-4_2

of medical reasons and the inherent issues of the conduction of a non-inferiority trial have been discussed extensively (CPMP 2000). A number of health authorities guidelines, including the draft guidance from the US Food and Drug Administration (FDA), have been released to guide the pharmaceutical industry to conduct non-inferiority trials (CPMP Working Party on Efficacy of Medicinal Products Note for Guidance III/3630/92-EN 1995, CHMP 2005, FDA Guidance for industry 2010, ICH Harmonised tripartite guideline 1998, ICH Harmonized tripartite guideline 2000).

There is a substantial literature on both the frequentist design and the Bayesian design for a simple noninferiority clinical trial to compare a test treatment with a control in terms of one primary endpoint, including Liu and Chang (2011) and Chen et al. (2011). Often there is only one primary endpoint involved in the hypothesis testing in a clinical trial. But sometimes multiple endpoints are simultaneously tested in a trial even though the formulation of hypotheses may be different depending on the study objectives, the study design, and the nature of multiplicity. Several corresponding statistical methods have been proposed. Sugimoto et al. (2012) present a convenient formula for sample size calculation in clinical trials with multiple co-primary continuous endpoints. Laska et al. (1992) extend the well-known optimality of the *min* test in the univariate case to the multivariate case and apply to superiority hypothesis testing on multiple endpoints. Kong et al. (2004) adopt the *min* test to non-inferiority hypothesis testing for multiple endpoints following a multivariate normal distribution.

The clinical trial with multiple co-primary endpoints is a special case of the one with multiple endpoints, in which all endpoints are equally important clinically. The conventional frequentist approach for a clinical trial with multiple co-primary endpoints is via the intersection–union testing (IUT; Eaton and Muirhead 2007). Recently, new statistical approaches have been developed to achieve a higher power while the family-wise type I error rate is still controlled. For example, Chuang-Stein et al. (2007) propose an approach based on the notion of controlling the average type I error rate over a restricted null space rather than over the conventional full null space. The other Bayesian approaches include Gonen et al. (2003) and Scott and Berger (2006). While most clinical trials compare two treatments, some trials compare three or more medications, multiple doses of medications, or medical devices against each other or against the standard treatment, which often leads to the multiplicity issue. If the global hypothesis involves multiple comparisons, an appropriate multiplicity adjustment method is required in order to control the family-wise type I error rate. Dmitrienko et al. (2010) give a comprehensive review on the multiple testing procedures widely used in clinical studies, including procedures based on univariate $p$ values (e.g., Bonferroni, Holm, fixed-sequence, Simes, Hommel, and Hochberg procedures), parametric procedures, and resampling-based procedures. A noninferiority clinical trial involving multiple dose levels for a study drug is often designed to demonstrate the noninferiority of the study drug under at least one dose level; hence, it is a typical multiple testing problem and an appropriate multiplicity adjustment method is required in a frequentist design. By now, there is a rich literature on the frequentist design of a noninferiority trial with multiple tests, including Ng (2003), Hung and Wang (2004), Tsong and Zhang (2007), and Röhmel and Pigeot (2010).

In this chapter, we develop a Bayesian approach for noninferiority clinical trials with co-primary endpoints and multiple dose comparison by incorporating historical data. One of the advantages of the Bayesian approach is that it has the potential of increasing power and reducing sample size due to the incorporation of historical data and the correlation structure among the multiple co-primary endpoints. Another advantage of the proposed Bayesian approach is to control the family-wise type I error automatically without an additional multiplicity adjustment. The Bayesian method is also demonstrated and compared with the conventional frequentist method with or without Bonferroni multiplicity adjustment via the design of a clinical trial.

The rest of the chapter is organized as follows. A motivation example of a non-inferiority clinical trial with co-primary endpoints and multiple dose comparison is described in Sect. 2.2. In Sect. 2.3, firstly the statistical settings of the noninferiority clinical trial with co-primary endpoints and multiple dose comparison are introduced. Then the conventional frequentist approach is briefly reviewed, and the Bayesian method using the commonly used conjugate prior and the power prior with fixed power parameter(s) to incorporate historical data for the control group is proposed and described. In Sect. 2.4, the proposed Bayesian method is applied to the non-inferiority clinical trial described in Sect. 2.2 in comparison with the conventional frequentist method. Finally, the chapter ends with the conclusion and discussion in Sect. 2.5.

## 2.2 Design of a Noninferiority Clinical Trial with Two Co-primary Endpoints and Multiple Dose Comparison

An experiment agent is currently in mid to the late-stage development as a treatment of signs and symptoms of benign prostatic hyperplasia or hypertrophy (BPH). BPH is a chronic and progressive condition that adversely affects health-related quality of life (HRQoL) by interfering with normal daily activities and sleep patterns. The International Prostate Symptom Score (IPSS) ranging from 0 to 35, also known as the American Urologic Association Symptom Score (AUA-SS), is collected in a questionnaire. The change of IPSS from the baseline score (denoted by $\Delta$IPSS thereafter) is one of the primary endpoints for all drug trials in the treatment of BPH. Although it is not mandatory, the change from baseline maximum urinary flow rate (denoted by $\Delta$Qmax thereafter) is recommended as another co-primary endpoint by European regulatory authority. In addition, the smaller $\Delta$IPSS is and the bigger $\Delta$Qmax is, the better the treatment effect of the test drug is.

A non-inferiority clinical trial design demonstrating that at least one dose regime of 15 mg QD or 30 mg QD of the experiment compound is non-inferior to the active comparator, Tamsulosin, the SOC for BPH, is explored to support further development of the experiment compound.

Based on the above consideration, a hypothetical study will be a multicenter, double-blind, three-arm parallel trial. The patients will be on placebo for 4 weeks before they are randomized to one of the three arms: 15 mg QD and 30 mg QD of

**Table 2.1** Summary statistics ($n$, mean ± standard deviation and correlation) of $\Delta$IPSS and $\Delta$Qmax for Tamsulosin

| Study | $n$ | $\Delta$IPSS | $\Delta$Qmax (ml/s) | Correlation coefficient between $\Delta$IPSS and $\Delta$Qmax |
|---|---|---|---|---|
| 1 | 244 | −5.1±6.4 | 1.52±3.59 | N/A |
| 2 | 34 | −7.03±5.84 | 1.68±4.08 | −0.29 |

*IPSS* International Prostate Symptom Score

the experiment compound, and Tamsulosin 0.4 mg QD dose group, for 12 weeks. After the 12-week double-blind treatment period, the patients who are randomized to the experiment compound will remain on the same treatment, and the patients who are on Tamsulosin will be re-randomized at the end of 12-week treatment to one of the dose groups of the experiment compound for another 40 weeks to assess the safety and tolerability of the compound. The two co-primary endpoints are $\Delta$IPSS and $\Delta$Qmax at the end of the 12-week double-blind treatment period.

Historical data are available from the two previous studies on Tamsulosin capsule 0.4 mg QD regime. The first study was a multicenter, randomized, double-blind, placebo-controlled, parallel group, phase III trial to evaluate the efficacy and safety of Tamsulosin for the treatment of patients with symptoms of moderate to severe BPH (Narayan and Ashutosh Tewari 1998). This study was conducted by Boehringer Ingelheim Pharmaceuticals, Inc. in 1993. The second study was a multicenter, randomized, double-blind, placebo-controlled, parallel group, phase II trial to evaluate the efficacy and safety of an experiment compound in the treatment of patients with lower urinary tract symptoms (LUTS), in which Tamsulosin was an active comparator (Tamimi et al. 2010). This study was conducted by Pfizer Inc. in 2007. Summary statistics for $\Delta$IPSS and $\Delta$Qmax for the active comparator of Tamsulosin from these two studies are shown in Table 2.1. The pooled standard deviations (SD) for $\Delta$IPSS and $\Delta$Qmax are 6.34 and 3.70 ml/s, respectively. In addition, clinically meaningful non-inferiority margins for $\Delta$IPSS and $\Delta$Qmax are chosen to be 1 and −0.6 ml/s, respectively, to design the noninferiority trial. The historical data in Table 2.1 is incorporated in the Bayesian design developed in the subsequent sections.

## 2.3 Methodology

### *2.3.1 Assumption and Notation*

We assume that there are three treatments in a clinical trial, including the study drug at a high dose level, the study drug at a low dose level, and the (active) control treatment, denoted by the $h$, $l$, and $c$ (treatment) groups, respectively. The objective of the study is to show non-inferiority of the study drug at a (high or low) dose level compared to the control group.

Let $\{y_{gi}, i = 1, 2, \ldots, n_g\}$ be a $J$-dimensional random sample of size $n_g$ collected for the $g$th group. Furthermore, given $\mu_g$ and $\Sigma$, we assume that $y_{gi}$ follows a multivariate normal $N_J(\mu_g, \Sigma)$ distribution, where $\mu_g$ is the mean vector for the $g$th group, and $\Sigma$ is the common variance covariance matrix for all the groups with the dimension of $J \times J$. Let $\theta = (\mu_h, \mu_l, \mu_c, \Sigma)$ denote the collection of parameters.

Without loss of generosity, we assume there are two co-primary endpoints, i.e., $J = 2$, where a smaller value of the first co-primary endpoint is better and a larger value of the second co-primary endpoint is better. Assume $\mu_g = (\mu_{g1}, \mu_{g2})'$, where $\mu_{g1}$ and $\mu_{g2}$ are the true means for the two co-primary endpoints for the $g$th group, respectively. The noninferiority hypotheses comparing the $g$th study drug group, $g = h, l$, with the control group can be formulated as $H_{0g}$: $\mu_{g1} - \mu_{c1} \geq \delta_{g1}$ or $\mu_{g2} - \mu_{c2} \leq \delta_{g2}$ versus $H_{1g}$: $\mu_{g1} - \mu_{c1} < \delta_{g1}$ and $\mu_{g2} - \mu_{c2} > \delta_{g2}$, where $\delta_{g1}$ and $\delta_{g2}$ are the noninferiority margins of the co-primary endpoints. Let $\delta_g = (\delta_{g1}, \delta_{g2})'$ for $g = h, l$. The noninferiority margins should be the same for both high and low doses in the comparison and, hence, we assume that $\delta_h = \delta_l = \delta$, where $\delta = (\delta_1, \delta_2)'$, in the subsequent sections. The objective of the study is to demonstrate non-inferiority of the study drug at a dose level after the noninferiority margin is chosen based on both clinical and statistical considerations.

We assume that $\{y_{gi}, i = 1, \ldots, n_g\}$, $g = h, l, c$, are independent across the groups. The likelihood function of the data can be written as

$$
\begin{aligned}
f(\theta|D) &\propto \Pi_{g=h,l,c} |\Sigma|^{-\frac{n_g}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n_g}(y_{gi} - \mu_g)'\Sigma^{-1}(y_{gi} - \mu_g)\right) \\
&= \Pi_{g=h,l,c} |\Sigma|^{-\frac{n_g}{2}} \exp\Big(-\frac{1}{2} tr\left(\sum_{i=1}^{n_g}(y_{gi} - \bar{y}_g)\right)(y_{gi} - \bar{y}_g)'\Sigma^{-1} \\
&\quad + n_g(\mu_g - \bar{y}_g)(\mu_g - \bar{y}_g)'\Sigma^{-1})\Big),
\end{aligned}
$$

where $\bar{y}_g = (\sum_{i=1}^{n_g} y_{gi})/n_g$, and the data $D = \{(y_{gi}, y_{lj}, y_{ck}),\ i = 1, \ldots, n_h,\ j = 1, \ldots, n_l, k = 1, \ldots, n_c\}$.

### 2.3.2 *Preliminary: The Frequentist Design*

Under the multivariate normal assumption, $(\bar{y}_h, \bar{y}_l, \bar{y}_c, S)$ are the sufficient statistics, where $S$ denotes the pooled matrix of sums of squares and cross products: $S = \sum_g (n_g - 1)S_g$, $g = h, l, c$, and $S_g = \left(\sum_{i=1}^{n_g}(y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)'\right)/(n_g - 1)$.

Let $\mathbf{W}_h = (n_h n_c/(n_h + n_c))^{1/2}(\bar{y}_h - \bar{y}_c - \delta)$, and $\mathbf{W}_l = (n_l n_c/(n_l + n_c))^{1/2}(\bar{y}_l - \bar{y}_c - \delta)$. Then $\mathbf{W}_h$ and $S$ are independent, and $\mathbf{W}_l$ and $S$ are independent, with $\mathbf{W}_h \sim N((n_h n_c/(n_h+n_c))^{1/2}(\mu_h - \mu_c - \delta), \Sigma)$, $\mathbf{W}_l \sim N((n_l n_c/(n_l+n_c))^{1/2}(\mu_l - \mu_c - \delta), \Sigma)$, and $S \sim$ Wishart$(n, \Sigma)$, where $n = n_h + n_l + n_c - 3$.

In the conventional frequentist design, an appropriate multiplicity adjustment method due to the multiple dose comparisons must be adopted in order to control the

family-wise type I error rate. Let $\alpha$ denote the desired overall one-sided significance level, and assume Bonferroni multiplicity adjustment is used to assign significance levels to comparisons of the study drug at the high and the low dose levels to the control group. Suppose that we consider the comparison of the study drug at either the high or the low dose level with the control group first. The standard test involves testing the two endpoints separately at the same significance level of $\alpha/2$ using the one-sided $t$ tests, and rejecting the null hypothesis $H_{0g}$ (in favor of the alternative hypothesis $H_{1g}$) if and only if the two separate $t$ test statistics are significant. Specifically, let $\mathbf{W}_g = (W_{g1}, W_{g2})'$, $g = h, l$, and define $T_{g1} = W_{g1}(s_{11}/n)^{-1/2}$ and $T_{g2} = W_{g2}(s_{22}/n)^{-1/2}$, where $s_{11}$ and $s_{22}$ are the diagonal elements of $S$. $T_{gd}$ has a standard $t$ distribution with degrees of freedom of $n$ when the $d$th element of $\mu_g - \mu_c - \delta$ is zero ($d = 1, 2$). The standard test rejects $H_{0g}$ if and only if $\max_d T_{gd} \leq c_{\alpha/2}$, where $c_{\alpha/2}$ is the $(\alpha/2)$th quantile of the $t$ distribution with degrees of freedom of $n$. Eaton and Muirhead (Eaton and Muirhead 2007) show that the standard test is an IUT and the size of the test is $\alpha/2$. Moreover, the standard test may be conservative because the two statistics $T_{gd}$, $d = 1, 2$, are assumed independent. The type I error rate approaches to $\alpha/2$ as the correlation coefficient of the two end points approaches to one.

Note that the assumption of the equal variance covariance for all groups could be relaxed if necessary. For example, Welch's $t$ test (Welch 1947) is an adaptation of the Student's $t$ test when the two samples have possibly unequal variances. Specifically, the test statistic is given by $T = (\bar{X}_1 - \bar{X}_2)(S_1^2/n_1 + S_2^2/n_2)^{-1/2}$, where $\bar{X}_i$, $S_i^2$, and $n_i$, $i = 1, 2$, are the $i$th sample mean, sample variance, and sample size, respectively. The degrees of freedom $\nu$ associated with the test can be approximated by

$$\nu = (S_1^2/n_1 + S_2^2/n_2)^2/[S_1^4/\{n_1^2(n_1 - 1)\} + S_2^4/\{n_2^2(n_2 - 1)\}].$$

### 2.3.3 The Proposed Bayesian Design

Following Chen et al. (2011), let $\pi^{(f)}(\theta)$ be the fitting prior and also let $\pi^{(s)}(\theta)$ be the sampling prior. The fitting prior is used to perform Bayesian analysis once data are collected. The sampling prior is the distribution for the parameters which we believe the future data would follow, and it is used to generate psuedo-data for the design evaluation, i.e., the type I error and power assessment. We assume the hypotheses: $H_{0g}$: $\eta_g(\theta) \geq \eta^*(\delta)$ versus $H_{1g}$: $\eta_g(\theta) < \eta^*(\delta)$, where $\eta^* = (\delta_1, -\delta_2)'$, $\eta_g(\theta) = (\mu_{g1} - \mu_{c1}, -(\mu_{g2} - \mu_{c2}))'$, the subscript $g = h$ is for the comparison between the high dose group of the study drug and the control group, and $g = l$ is for the comparison between the low dose group of the study drug and the control group.

Define the key quantity:

$$\beta_s = E_s\left[\mathbf{1}\{\cup_{g=h,l}\{P(\eta_g(\theta) < \eta^*(\delta)|D) \geq \gamma_g\}\}\right], \tag{2.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, and $\gamma_g$ is a prespecified credible level in $(0, 1)$, e.g., 0.95. It is reasonable to assume that $\gamma_h = \gamma_l = \gamma$ for our scenario since

there is no differentiation for the high and low dose group comparisons with the control group in terms of $\gamma_g$. Therefore, we use the same credible level $\gamma$ in $\beta_s$ in the subsequent sections. The probability in (2.1) is calculated with respected to the posterior distribution of $\theta$, given the data $D$ and the fitting prior $\pi^{(f)}(\theta)$, and the expectation is taken with respect to the marginal distribution of the data under the sampling prior $\pi^{(s)}(\theta)$.

We use the same Bayesian sample size determination algorithm from (Chen et al. 2011). Let $\Theta_0$ and $\Theta_1$ denote the parameter spaces corresponding to the null and alternative hypothesis, respectively, and let $\bar{\Theta}_0$ and $\bar{\Theta}_1$ be the closure of $\Theta_0$ and $\Theta_1$. Further, let $\pi_0^{(s)}(\theta)$ be the sampling prior with support $\Theta_B = \bar{\Theta}_0 \cap \bar{\Theta}_1$ and let $\pi_1^{(s)}(\theta)$ be the sampling prior with support $\Theta_1^* \subset \Theta_1$. For given $\alpha_0 > 0$ and $\alpha_1 > 0$, we compute

$$n_{\alpha_0} = \min\{n : \beta_{s0} \leq \alpha_0\}; \quad n_{\alpha_1} = \min\{n : \beta_{s1} \geq 1 - \alpha_1\}, \tag{2.2}$$

where $\beta_{s0}$ and $\beta_{s1}$ in (2.2) are the $\beta_s$'s in (2.1) by letting $\pi^{(s)}(\theta)$ be $\pi_0^{(s)}(\theta)$ and $\pi_1^{(s)}(\theta)$, respectively, and they are the Bayesian type I error and power, respectively. The Bayesian sample size is given by $n_B = \max\{n_{\alpha_0}, n_{\alpha_1}\}$. One possible choice of $\gamma$ is 0.975, which is comparable to a significant level of 0.05/2 used for the individual hypothesis test under the frequentist design with multiplicity adjustment. Common choices of $\alpha_0$ and $\alpha_1$ include $\alpha_0 = 0.05$ and $\alpha_1 = 0.2$ so that in a Bayesian design with sample size $n_B$, the family-wise type I error rate is less than or equal to 0.05 and the power is at least 0.8. The choice of $\Theta_1^*$ is often related to the design margins $\delta_{\mathrm{d}}$'s. For example, for a continuous endpoint, a typical choice of $\mu_{gd}$ in $\Theta_1^*$ for the noninferiority hypothesis testing is $\mu_{cd}$.

Historical data can be incorporated via the different forms of the fitting prior, including the power prior with a fixed or random or mixture power parameter, the hierarchical prior, and the hierarchical commensurate and power prior. In this chapter, for simplicity, we consider the commonly used conjugate prior and the power prior with a fixed power parameter. Often the historical data are only available for the control group; hence, a noninformative fitting prior is assumed for the study drug.

### 2.3.4 The Conjugate Prior

The conjugate priors for the unknown parameters $\theta = (\mu_h, \mu_l, \mu_c, \Sigma)$ are given as

$$\Sigma \sim \text{Inv–Wishart}_{\nu_0}(\Lambda_0),$$
$$\mu_g | \Sigma \sim N(\mu_{g0}, \Sigma/\kappa_{g0}),$$

where $\nu_0$, $\Lambda_0$, $\mu_{g0}$, and $\kappa_{g0}$ are known constants. The posterior distributions for $(\mu_g, \Sigma)$, $g = h, l, c$, are in the same families as the prior distributions but with updated parameters. Specifically, the marginal posterior distribution for $\mu_g$ is a multivariate

$t$ distribution:

$$\mu_g|D \sim t_{\nu_n - J+1}\left(\mu_{gn}, \frac{\Lambda_n}{\kappa_{gn}(\nu_n - J + 1)}\right),$$

the marginal distribution of $\Sigma$ is an Inverse–Wishart distribution:

$$\Sigma|D \sim \text{Inv–Wishart}_{\nu_n}(\Lambda_n),$$

and the conditional distribution of $\mu_g$ given $\Sigma$ is a multivariate normal:

$$\mu_g|\Sigma, D \sim N(\mu_{gn}, \Sigma/\kappa_{gn}),$$

where $J = 2$, $\mu_{gn} = \frac{\kappa_{g0}}{\kappa_{g0}+n_g}\mu_{g0} + \frac{n_g}{\kappa_{g0}+n_g}\bar{y}_g$, $\kappa_{gn} = \kappa_{g0} + n_g$, $\nu_n = \nu_0 + n_h + n_l + n_c$, $\Lambda_n = \Lambda_0 + \sum_{g=h,l,c}\left\{S_{gn} + \frac{\kappa_{g0}n_g}{\kappa_{g0}+n_g}(\bar{y}_g - \mu_{g0})(\bar{y}_g - \mu_{g0})'\right\}$, and $S_{gn} = \sum_{i=1}^{n_g}(y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)'$.

Samples from the joint posterior distribution for $(\mu_g, \Sigma)$, $g = h, l, c$, can be obtained using the following procedure:

Step 1. Draw $\Sigma|D \sim \text{Inv–Wishart}_{\nu_n}(\Lambda_n)$
Step 2. Independently draw $\mu_g|\Sigma, D \sim N(\mu_{gn}, \Sigma/\kappa_{gn}), g = h, l, c$.

The following is a computation algorithm to compute the study type I error or power for given $n_g$, $\delta$, $\gamma$, $M$ (number of Monte Carlo samples), and $N$ (number of simulations):

Step 1. Generate $\theta$ from the sampling prior, i.e., $\theta \sim \pi^{(s)}(\theta)$.
Step 2. Generate data from the multivariate normal distribution, i.e., $y_g \sim N(\mu_g, \Sigma)$, $g = h, l, c$.
Step 3. Generate $M$ samples $\theta^{(m)}$, $m = 1, \ldots, M$, from the joint posterior distribution using the algorithm shown above.
Step 4. Compute $\hat{P}_g = M^{-1}\sum_{m=1}^{M}\mathbf{1}\{\eta_g(\theta^{(m)}) < \eta^*(\delta)\}$, and check whether $\hat{P}_g \geq \gamma$ or not.
Step 5. Repeat steps 1–4 $N$ times, then calculate the proportion of $\{\cup_{g=h,l}\hat{P}_g \geq \gamma\}$ among those $N$ times, which gives an estimate of $\beta_s$, i.e., the type I error or power.

For the sample size determination, we need to repeat the above procedure for other scenarios of different combinations of $n_g$'s and then choose the optimal combination of $n_g$'s as the desired sample size under which both the type I error and power satisfy the design requirement.

**A Special Case: The Noninformative Prior.** In order to facilitate the comparison between the Bayesian approach and the frequentist approach, it is desirable to specify a noninformative prior in the Bayesian approach. A commonly used noninformative prior is the multivariate Jeffreys prior, $\pi(\mu_g, \Sigma) \propto |\Sigma|^{-\frac{J+1}{2}}$, which is the limit of the conjugate prior as $\kappa_{g0} \to 0$, $\nu_0 \to -1$, and $|\Lambda_0| \to 0$, $g = h, l, c$. Consequently, the corresponding posterior distribution can be obtained by

$$\Sigma|D \sim \text{Inv–Wishart}_{n_h+n_l+n_c-1}(S_{hn} + S_{ln} + S_{cn}),$$

$$\mu_g|\Sigma, D \sim N(\bar{y}_g, \Sigma/n_g).$$

The computational algorithm to determine the sample size is exactly same as that described above except for the Monte Carlo sampling steps, where the samples of parameters are drawn from a different posterior distribution. In addition, the marginal distribution for $\mu_g$ is a multivariate $t$ distribution:

$$t_{n_h+n_l+n_c-J}(\bar{y}_g, S_{gn}/(n_g(n_h+n_l+n_c-J))),$$

where $J = 2$.

### 2.3.5 The Power Prior

We extend the power prior of (Ibrahim and Chen 2000) to construct the fitting prior for $\mu_c$ and $\Sigma$. In general, we assume that there are a total of $K$ sets of historical data available for the control group, denoted by $y_{c0k} = (y_{c0ki}, i = 1, 2, \ldots, n_{0k})'$, $k = 1, 2, \ldots, K$, where $n_{0k}$ is the number of samples collected in the $k$th historical dataset. Furthermore, we let $y_{c0} = ((y_{c01})', (y_{c02})', \ldots, (y_{c0K})')'$ denote all the $K$ historical datasets.

We consider the power prior with a fixed power parameter $\mathbf{a}_0$ for $\mu_c$ and $\Sigma$ as

$$\pi(\mu_c, \Sigma|y_{c0}, \mathbf{a}_0)$$
$$\propto \prod_{k=1}^{K}\left[|\Sigma|^{-\frac{n_{0k}}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n_{0k}}(y_{c0ki}-\mu_c)'\Sigma^{-1}(y_{c0ki}-\mu_c)\right)\right]^{a_{0k}} \pi_0(\mu_c, \Sigma), \quad (2.3)$$

where $\mathbf{a}_0 = (a_{01}, \ldots, a_{0K})'$, $0 \le a_{0k} \le 1$, for $k = 1, \ldots, K$, and $\pi_0(\mu_c, \Sigma)$ is an initial prior. When $\pi_0(\mu_c, \Sigma) \propto |\Sigma|^{-(d+1)/2}$, i.e., the noninformative Jeffreys prior, (2.3) can reduce to the conjugate prior

$$\Sigma|y_{c0}, \mathbf{a}_0 \sim \text{Inv–Wishart}_{\nu_0}(\Lambda_0),$$
$$\mu_c|\Sigma, y_{c0}, \mathbf{a}_0 \sim N\left(\sum_{k=1}^{K}(a_{0k}n_{0k}\bar{y}_{c0k})/n_0(\mathbf{a}_0), \Sigma/\kappa_{c0}\right),$$

where

$$\Lambda_0 = \sum_{k=1}^{K} a_{0k}S_{0k} + \sum_{k=1}^{K}(a_{0k}n_{0k}\bar{y}_{c0k}\bar{y}_{c0k}')$$
$$-\sum_{k=1}^{K}(a_{0k}n_{0k}\bar{y}_{c0k})\sum_{k=1}^{K}(a_{0k}n_{0k}\bar{y}_{c0k})'/n_0(\mathbf{a}_0),$$

$S_{0k} = \sum_{i=1}^{n_{0k}}(y_{c0ki}-\bar{y}_{c0k})(y_{c0ki}-\bar{y}_{c0k})'$, $\bar{y}_{c0k} = \left(\sum_{i=1}^{n_{0k}} y_{c0ki}\right)/n_{0k}$, $\nu_0 = n_0(\mathbf{a}_0) - 1$, $\kappa_{c0} = n_0(\mathbf{a}_0) = \sum_{k=1}^{K} a_{0k}n_{0k}$. Then, the computational algorithms developed in Sect. 2.3.4 for the conjugate prior can be applied here correspondingly.

## 2.4 Application to the Design of a Non-inferiority Trial

In this section, we apply the proposed Bayesian approach to the design of the non-inferiority trial described in Sect. 2.2. We use simulations to investigate the performance of the proposed approach in terms of the type I error and power, and compare the Bayesian approach with the conventional frequentist approach with or without the Bonferroni multiplicity adjustment. We assume the data corresponding to the high dose group of the study drug, the low dose group of the study drug, and the control group have the distributions $y_{gi}|\mu_g, \Sigma \sim N_2(\mu_g, \Sigma)$, where $\mu_g$ is the mean vector for the $g$th group, $\Sigma$ is the common variance covariance matrix for all groups, and $i = 1, 2, \cdots, n_g$. Let $\mu_g = (\mu_{g1}, \mu_{g2})'$, where $\mu_{g1}$ and $\mu_{g2}$ are the true means for the two co-primary endpoints, respectively, for the $g$th group. We choose the point mass sampling priors as commonly used in the frequentist trial design and trial analysis. That is, we let

$$\pi^{(s)}(\mu_g) = \begin{cases} 1 & \text{if } \mu_g = \mu_g^{(s)} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \pi^{(s)}(\Sigma) = \begin{cases} 1 & \text{if } \Sigma = \Sigma^{(s)} \\ 0 & \text{otherwise,} \end{cases}$$

where $\mu_c^{(s)}$ and $\Sigma^{(s)}$ are prespecified values. The sample size is also allowed to change in the simulations, which can be used for the sample size determination during the design stage. The design strategy is to find a minimum total size $n$, i.e., $n = n_h + n_l + n_c$, so that the power is at least 80 % and the type I error is controlled at 5 %.

Figure 2.1 shows mean vectors of the two co-primary endpoints for the control group from the two historical data as well as the pooled mean vector. We assume that the mean vector for the co-primary endpoints for the control group for the future data is a linear combination of the mean vector from the first historical data, $\bar{y}_{c01} = c(-5.1, 1.52)'$, and the mean vector from the pooled historical data, $\bar{y}_{c0\cdot} = c(-5.34, 1.54)'$. That is, $\mu_c^{(s)}$ is chosen to be any point on the line interval of $AC$ in Fig. 2.1. Moreover, $\Sigma^{(s)}$ is chosen to be that the variance components are the pooled variances for the two co-primary endpoints from the two historical trials, i.e., $6.34^2$ and $3.70^2$, respectively, and the correlation coefficient to be $-0.29$, estimated from the second historical study. The design value of $\mu_g^{(s)}, g = h, l$ is chosen according to the type I error or power evaluation.

Let $\delta = (\delta_1, \delta_2)' = (1, -0.6)'$. For the type I error evaluation, we simulate the data from the sampling priors with parameters of $\mu_h^{(s)} = \mu_c^{(s)} + \delta$, $\mu_l^{(s)} = \mu_c^{(s)} + \delta$, or $\mu_h^{(s)} = \mu_c^{(s)} + (\delta_1, 0)'$, $\mu_l^{(s)} = \mu_c^{(s)} + (\delta_1, 0)'$, or $\mu_h^{(s)} = \mu_c^{(s)} + (0, \delta_2)'$, $\mu_l^{(s)} = \mu_c^{(s)} + (0, \delta_2)'$, and define the type I error for the design as the maximum type I error. For the power evaluation, we let the sampling prior parameters be $\mu_h^{(s)} = \mu_c^{(s)}$, $\mu_l^{(s)} = \mu_c^{(s)}$, or $\mu_h^{(s)} = \mu_c^{(s)} + (\delta_1, 0)'$, $\mu_l^{(s)} = \mu_c^{(s)}$, or $\mu_h^{(s)} = \mu_c^{(s)} + (0, \delta_2)'$, $\mu_l^{(s)} = \mu_c^{(s)}$, or $\mu_h^{(s)} = \mu_c^{(s)} + (\delta_1, \delta_2)'$, $\mu_l^{(s)} = \mu_c^{(s)}$, and define the power for the design as the minimum power.

We use the power prior in Sect. 2.3.5 to incorporate the two historical data for the control group and use a power prior with an approximately noninformative prior

http://www.springer.com/978-3-319-12693-7

Applied Statistics in Biomedicine and Clinical Trials Design
Selected Papers from 2013 ICSA/ISBS Joint Statistical Meetings
Chen, Z.; Aiyi, L.; Qu, Y.; Tang, L.; Ting, N.; Tsong, Y. (Eds.)
2015, XXIV, 546 p. 107 illus., 35 illus. in color., Hardcover
ISBN: 978-3-319-12693-7