Shengjie Liu¹ / Jun Gao¹ / Yuling Zheng¹ / Lei Huang¹ / Fangrong Yan¹

Bayesian Two-Stage Adaptive Design in Bioequivalence

¹ Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing 210009, P.R. China, E-mail: f.ryan@163.com

Abstract:

Bioequivalence (BE) studies are an integral component of new drug development process, and play an important role in approval and marketing of generic drug products. However, existing design and evaluation methods are basically under the framework of frequentist theory, while few implements Bayesian ideas. Based on the bioequivalence predictive probability model and sample re-estimation strategy, we propose a new Bayesian two-stage adaptive design and explore its application in bioequivalence testing. The new design differs from existing two-stage design (such as Potvin's method B, C) in the following aspects. First, it not only incorporates historical information and expert information, but further combines experimental data flexibly to aid decision-making. Secondly, its sample re-estimation strategy is based on the ratio of the information in interim analysis to total information, which is simpler in calculation than the Potvin's method. Simulation results manifested that the two-stage design can be combined with various stop boundary functions, and the results are different. Moreover, the proposed method saves sample size compared to the Potvin's method under the conditions that type I error rate is below 0.05 and statistical power reaches 80 %.

Keywords: Bayesian two-stage adaptive design, Bioequivalence, Interim analysis, Sample re-estimation strategy, Stop boundary functions

DOI: 10.1515/ijb-2018-0105

Received: October 13, 2018; **Revised:** June 13, 2019; **Accepted:** June 26, 2019

1 Introduction

Bioequivalence studies aim at demonstrating in vivo that two pharmaceutical products are comparable with respect to their bioavailability, that is, the ratios of rate and extent of absorption lie between certain prespecified acceptable limits [1]. Bioequivalence evaluation is a process of using statistical methods to determine whether such requirements are satisfied.

The most common method of bioequivalence evaluation is to compare the efficacy of different drugs by using pharmacokinetic parameters, such as area under the concentration-time curve (AUC), maximum blood concentration ($C_{\max}$) and the peak time ($T_{\max}$), and then to judge whether the difference between different drugs in actual treatment is statistically significant [2]. There are mainly three types of bioequivalence: average bioequivalence (ABE), population bioequivalence (PBE) and individual bioequivalence (IBE). At present, ABE is adopted the most in national laws and regulations, and statistical analysis is carried out to compare the means of pharmacokinetic parameters of the trial drug and the reference drug to infer whether the population parameters satisfy bioequivalence.

The most common trial design used in ABE is cross-over design (such as 2×2 cross-over design). Because the clearance rate of most drugs varies greatly among individuals, the coefficient of variation among individuals is much larger than that within individuals. Cross-over design can effectively avoid the bias caused by variation among individuals [3]. Another advantage, which is probably more important, is the direct ethical and economic benefits. For some highly variable drugs, we may adopt scaled average designs in bioequivalence studies [4].

Statistical testing for ABE is approaching maturity after long period of development. The number of subjects for bioequivalence evaluation is usually between 18 and 24, so the simple and convenient frequentist methods are very advantageous. However, for some drugs with high variability, sample size may be much larger than 24 cases. Therefore, it is necessary to develop novel methods for the evaluation of these drugs.

In the early 1990s [5], the add-on design (AOD) was very popular and was mentioned by many regulatory documents. The design allows recruitment of additional subjects if the results of the original subject group are

Fangrong Yan is the corresponding author.

© 2020 Walter de Gruyter GmbH, Berlin/Boston.

promising (point estimates meet the bioequivalence requirements but interval estimations don't), and bioequivalence can be re-evaluated depending on the combined data. The use of the AOD design should be specified in the protocol in advance, and the two groups of subjects should be treated with the same batch of product and analyzed by the same clinical analysis method. However, since the AOD design does not adjust significance levels during repeated testing, the family-wise type I error rate may exceed the specified level of 0.05.

In 2004, the Product Quality Research Institute (PQRI) established the Sequential Design Working Group under its Biopharmaceutical Technical Committee and proposed a two-stage sequential design framework based on a large number of simulation studies. The simulations proved that the design is able to control risks to subjects. It is a remedy for failure of clinical bioequivalence trial due to insufficient statistical power. The two-stage design by Potvin belongs to this framework [6]. Unlike the AOD design, these designs adjust the significance levels at each stage so that the family-wise type I error rate meets the requirements. The success of two-stage sequential design was approved by government regulators. For instance, the US Food and Drug Administration allow the implement of two-stage group sequential design approach for bioequivalence evaluation. In addition, the latest guidance issued by the European Medicines Agency proposed a two-stage design (TSD) as an alternative to the standard 2×2 or repeated BE design.

All above designs belong to the frequentist framework which cannot fully utilize the prior information of the reference drug, while this can be remedied by the Bayesian method. Bayesian method has gain popularity in clinical trial design for its flexibility, and a growing amount of literature has begun to study two-stage adaptive design under the Bayesian framework. For example, [7] proposed the Optimal Two-Stage design and the Minimax design in 1989. The number of patients in the first and second stage of the trial is set to n_1 and n_2 respectively. If r_1 or fewer responders are observed by the end of the first stage, the trial will be stopped. The product will be rejected and a larger trial is unnecessary. Otherwise, the second stage of the trial will be continued. If r or fewer responders are observed after the trial, the product will be rejected and the subsequent expansion trial will not be continued. Another example is [8], which proposed a Bayesian two-stage design in which parameters are determined based on posterior probability.

The existing Bayesian designs are basically applied to phase II and III clinical trials, and no method has been developed for bioequivalence test yet. Since bioequivalence studies generally use two-crossover design to assign subjects, which is different from the parallel design of phase II and III clinical trials, a design specially for bioequivalence test is needed. In Bayesian sequential monitoring design, a conservative sample size is set before the start of the trial, and if early stop conditions are met, the trial will be stopped early in interim analysis, thus requiring a smaller sample size. The design idea of this paper is to re-adjust sample size according to existing information in the middle of the trial, which controls the type I error rate at a specified level and ensures sufficient statistical power.

The rest of this paper is organized as follows: in Section 2, we illustrate the predictive probability model of bioequivalence and introduces the theory of sample re-estimation. Then a new Bayesian two-stage adaptive design will be proposed and applied to the bioequivalence study, which allows sample re-estimation design during the interim analysis. In Section 3, we calculate the family-wise type I error rate and statistical power by simulation, and the corresponding average sample size to verify the feasibility of our method. In Section 4, we discuss the advantages and disadvantages of the new method in detail and provide insights on future research.

2 Models and methods

2.1 Probability model

The most commonly used statistical design for ABE evaluation between test formulation (T) and reference formulation (R) of a drug is a two-sequence, two-period, crossover design [9], referred to as the standard 2×2 crossover design.

In the 2×2 crossover design, Y is the logarithm of the pharmacokinetic parameters (e. g. AUC , C_{max}). $Y_{Ti} \sim N(\mu_T, \sigma_T^2)$, $Y_{Ri} \sim N(\mu_R, \sigma_R^2)$, $i = 1 \dots n$ where T denotes the test product and R denotes the reference product. Logarithm is used because in most cases response measurements in bioequivalence study follow a lognormal distribution with skewness [9]. Then the hypothesis of the bioequivalence test is:

$$H_0 : \mu_T - \mu_R \leq \theta_L \text{ or } \mu_T - \mu_R \geq \theta_U \quad \text{vs.} \quad H_a : \theta_L < \mu_T - \mu_R < \theta_U \quad (1)$$

Let $X_{1i} = Y_{Ti} - T_{Ri}$ denote the difference in drug absorption of each subject in the TR sequence between two drugs, and $X_{2i} = Y_{Ti} - T_{Ri}$ denotes the difference in drug absorption of each subject in the RT sequence between two drugs.

Assume the following model [10]:

$$X_{1i} \sim N(\theta + \phi, 2\sigma^2) \quad (2)$$

$$X_{2i} \sim N(\theta - \phi, 2\sigma^2) \quad (3)$$

Where $\theta = \mu_T - \mu_R$ represents the difference in formulation, ϕ is the fixed effect of period. We assume $\phi = 0$, thus the above model can be simplified as:

$$Y_{Ti} - Y_{Ri} \sim N(\theta, 2\sigma^2) \quad (4)$$

Where σ^2 denotes the coefficient of variation among individuals, which satisfies $\sigma^2 = \log(1 + CV^2)$, CV is the coefficient of variation.

To perform Bayesian analysis, we choose appropriate prior distributions for the parameters. In clinical trials, reference drugs often have a mass of historical data that can be used to construct a prior distribution. Here, we set a prior distribution of μ_R to $N(\mu_0, \sigma^2/\tau)$, where μ_0 is determined based on prior information of historical data, and parameter τ is used to characterize the similarity between current data μ_R and historical data μ_0 . The smaller τ is, the more uncertain the similarity becomes. However, available prior information of test drugs before the start of the test is generally small, so non-informative distribution is selected as the prior distribution of μ_T , such as uniform distribution. We herein assume $\pi(\mu_T) = 1$.

By the end of the first stage, it is assumed that a total of $n(t_1)$ patients are enrolled where $n(t_1) = n_R(t_1) \cdot t_1$ denotes the information ratio at the end of the first stage and $t_1 = n(t_1)/N_{ori}$, where N_{ori} is the initial maximum sample size. Therefore, after the first stage, the posterior probability distributions of μ_T and μ_R are:

$$\pi(\mu_T | \bar{X}_{T1}) \sim N(\bar{X}_{T1}, \frac{\sigma_T^2}{n_T(t_1)}) \quad (5)$$

$$\pi(\mu_R | \bar{X}_{R1}, \mu_0, \tau) \sim N(\frac{\mu_0\tau + \bar{X}_{R1}n_R(t_1)}{\tau + n_R(t_1)}, \frac{\sigma_R^2}{\tau + n_R(t_1)}) \quad (6)$$

$\bar{X}_{T1}$ is the mean of the test formulation by the end of the first stage. Analogously, $\bar{X}_{R1}$ is the mean of the reference formulation.

The predictive posterior probability is based on data of the first stage, and simulated data is used to estimate the posterior probability of rejecting the null hypothesis when the maximum sample size is reached. The difference between posterior probability and predicted posterior probability is that the former one is based only on existing data, while predictive posterior probability must consider not only the existing data, but also the simulation data generated based on the existing information.

By the end of the first stage, the predicted posterior probability distributions for μ_T and μ_R are:

$$Pre(\mu_T | \bar{X}_{Tj}, n) \sim N(\frac{(n - n_T(t_j))\bar{X}_{Tpi} + n_T(t_j)\bar{X}_{Tj}}{n}, \frac{\sigma_T^2}{n}) \quad (7)$$

$$Pre(\mu_R | \bar{X}_{Rj}, \mu_0, \tau, n) \sim N(\frac{(n - n_R(t_j))\bar{X}_{Rpi} + (n_R(t_j)\bar{X}_{Rj} + \tau\mu_0)}{n + \tau}, \frac{\sigma_R^2}{n + \tau}) \quad (8)$$

Where $\bar{X}_{Tpj} = \frac{1}{n - n_T(t_j)} \sum_{i=1}^{n - n_T(t_j)} X_{Tpi}$, $\bar{X}_{Rpj} = \frac{1}{n - n_R(t_j)} \sum_{i=1}^{n - n_R(t_j)} X_{Rpi}$

$$\begin{aligned} & Pre(|\mu_T - \mu_R| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau, n) \\ & \cong \frac{1}{n} \sum_{r=1}^n I \left[|\Pr e(\mu_{Tr}) - \Pr e(\mu_{Rr})| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau, n \right] \end{aligned} \quad (9)$$

Where $\Pr e(\mu_{Tr})$, $\Pr e(\mu_{Rr})$ are the r^{th} sampled case obtained from the posterior distributions $Pre(\mu_T | \bar{X}_{Tj}, n)$ and $Pre(\mu_R | \bar{X}_{Rj}, \mu_0, \tau, n)$, respectively.

When the posterior probability $P(|\mu_T - \mu_R| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau)$, that is the posterior probability of average bioequivalence (PABE), is greater than the boundary value P_{ul} determined in advance, the null hypothesis is rejected, and the trial is stopped for effectivity with the conclusion that the two drugs are bioequivalent. When $\Pr e(|\mu_T - \mu_R| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau)$ is less than the boundary value P_l determined in advance, the trial is stopped for futility with the conclusion that the two drugs are not bioequivalent.

Here we will introduce PABE. Through the posterior probability in formula eq. (2), we can get the average bioequivalence posterior probability, which is recorded as PABE. The calculation formula is as follows:

$$\begin{aligned} PABE &= \Pr (ABE | data) = \Pr \{ \theta_L < F_T - F_R < \theta_U | data \} \\ &\cong \frac{1}{N_{rep}} \sum_{r=1}^{N_{rep}} I [\log (0.8) < F_{Tr} - F_{Rr} < \log (1.25)] \end{aligned} \quad (10)$$

$F_{Tr} - F_{Rr}$ is a sample obtained from the posterior distribution of $F_T - F_R$. $I(\bullet)$ is an indicator function and N_{rep} is a Gibbs sampling number.

2.2 Sample size re-estimation strategy

In bioequivalence trial, it is crucial to choose a proper initial sample size N_{ori} . If N_{ori} is too small, the trial will fail due to insufficient performance. If it is too large, it will lead to waste of resources. To avoid that, we consider re-estimating the sample size based on available data after the end of the first stage.

There are many strategies for sample re-estimating. The common conditional effective-ness method [11] is to estimate the required new sample size based on available data during the interim analysis in order to ensure final success of the trial and acquire specified power. Unlike the former, the Bayesian predictive efficacy method [12], assumes a prior distribution of therapeutic effect when predictive test is finally successful. In addition, [13] and [14] proposed using information ratio to adjust sample size.

In order to simplify design and improve clinical applicability, this section uses information ratio method to re-estimate the sample size. According to eq. (3), it can be concluded that:

$$\hat{\theta} \sim N \left(\theta, \sigma^2 \left(\frac{1}{n_T} + \frac{1}{n_R} \right) \right) \quad (11)$$

In the j^{th} stage, the amount of information of θ is the inverse of variance in the above equation:

$$I(t_j) = \frac{1}{\sigma^2} \left(\frac{1}{n_T} + \frac{1}{n_R} \right)^{-1} \quad (12)$$

The maximum amount of information I_{max} requires type I error rate below 5 % and power above 80 %.

$$N_{ori} = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{(\delta - \varepsilon)^2} \quad (13)$$

$$I_{max} = \frac{N_{ori}}{\sigma^2} \quad (14)$$

When variance is unknown, the amount of information at the end of the first stage can be calculated by:

$$\hat{I}(t_1) = \frac{1}{S^2} \left(\frac{1}{n_T(t_1)} + \frac{1}{n_R(t_1)} \right)^{-1} \quad (15)$$

$I(t_1)$ denotes the amount of information at the end of the first stage. When the first stage ends:

$$t_1 = \frac{I(t_1)}{I_{max}} = \frac{n(t_1)}{N_{max}^*} \quad (16)$$

Then the sample size can be re-estimated according to the following formula:

$$N_{max}^* = n(t_1) \left/ \frac{I(t_1)}{I_{max}} \right. \quad (17)$$

Where $n(t_1)$ is the sample size of the first stage, and $I(t_1)$ can be estimated based on the data. If it is larger than expected, then N_{max}^* will be smaller than the original N_{ori} , otherwise it will be the opposite. Subsequently, additional $N_{max}^* - n(t_1)$ subjects will be enrolled.

$$N_{max}^* = \begin{cases} N_{max}^* \frac{I(t_1)}{I_{max}} \leq 1, \\ N_{ori} \frac{I(t_1)}{I_{max}} > 1. \end{cases} \quad (18)$$

2.3 Bayesian two-stage adaptive design

This section presents a Bayesian two-stage adaptive design, which is described as follows (as Figure 1 shows):

Algorithm: Bayesian two-stage adaptive in bioequivalence (BTABE)

When the first stage is complete, the first interim analysis is as follows:

Step 1. Distribution updating:

Update the posterior probability distributions of μ_T and μ_R based on prior distribution and current patients data $\bar{X}_{Tj} = (X_{T1}, X_{T2}, \dots, X_{Tj})$, $\bar{X}_{Rj} = (X_{R1}, X_{R2}, \dots, X_{Rj})$. The posterior distribution of μ_T and μ_R are separately $\pi(\mu_T | \bar{X}_{Tj})$, $\pi(\mu_R | \bar{X}_{Rj}, \mu_0, \tau)$.

Step 2. Calculation of posterior probability:

Calculate the posterior probability of rejecting the null hypothesis

$P(|\mu_T - \mu_R| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau)$, that is PABE.

Step 3. Calculation of predicted posterior probability:

Generate simulated data of enrolled patients according to the posterior distribution, and calculate the predicted posterior probability $Pre(|\mu_T - \mu_R| < \theta | \bar{X}_{Tj}, \bar{X}_{Rj}, \mu_0, \tau, n)$ of rejecting the null hypothesis.

Step 4. Stopping for effectivity:

If $PABE \geq P_{u1}$ stop the trial for effectivity and conclude bioequivalence, otherwise jump to Step 6.

Step 5. Stopping for futility:

If $Pre \leq P_l$ stop the trial for futility and conclude bioinequivalence, otherwise continue perform to Step 6.

Step 6. Sample size re-estimation:

If the above stop conditions are not met, re-estimate the sample size according to first stage and continue to enroll subjects into second stage of the trial based on the new sample size.

Step 7. End trial:

Update the posterior probability. If $PABE \geq P_{u2}$, stop the trial and conclude bioequivalence, otherwise conclude bioinequivalence.

2.4 Stopping boundary function

The Bayesian two-stage adaptive design contains futility stop boundary P_l and effectivity stop boundaries P_{u1} , P_{u2} . The selection of boundary parameters shall affect the type I error rate and power.

Controlling type I error rate is a matter of concern in sequential design because multiple adaptive adjustments will cause type I error rate inflation. Pocock [15], O'Brien and Fleming [16] proposed solutions for type I error rate inflation in early sequential designs, which have been proved successful, but both methods have limitations. First, sample size and time points of the interim analysis need to be determined in advance. However, interim analysis may not be performed strictly in accordance with the plan, or it may be necessary to perform additional interim analysis in the trial. Another drawback is patients need to be evenly assigned to various stages. To solve the above problem, [17] proposed the concept of α allocation function. It assumes that functional relationship exists between the cumulative type I error rate at each stage and the ratio of the number of participants to the total sample size. This ratio is called the information ratio, denoted as t^* , and the function relationship is an alpha allocation function. The alpha allocation function has different forms, some of which are similar to Pocock and OBF boundaries, referred to in this section as the OBF type and the Pocock type alpha allocation function. In addition, Power family distribution functions are also frequently used in clinical trials.

Assume that the futility stopping boundary has been determined before the trial begins. We mainly study the influence of three distribution functions, OBF, Pocock and Power family, on the design performance, specifically in terms of type I error rate, power and sample size.

(1) OBF - α distribution function.

$$\alpha(t^*) = 2 - 2\Phi\left(\frac{z_{\alpha/2}}{\sqrt{t^*}}\right) \quad (19)$$

Where ϕ denotes standard normal distribution function.

(2) Pocock- α distribution function

$$\alpha(t^*) = \alpha \log\{1 + (e - 1)t^*\} \quad (20)$$

(3) Power- α distribution function

$$\alpha(t^*) = (t^*)^{\gamma} \alpha \quad (21)$$

We assume $\gamma=1$ in this paper.

3 Simulation studies

3.1 Simulation settings

To explore the performance of Bayesian two-stage adaptive design in this section, the simulation analysis was set up as follows. A two-stage BE bioequivalence evaluation study was conducted using a standard cross-experiment design. Y is the logarithm of pharmacokinetic parameter (e. g. AUC , C_{max}). $Y_{Ti} \sim N(\mu_T, \sigma_T^2)$ and $Y_{Ri} \sim N(\mu_R, \sigma_R^2)$ denotes the test product data and reference product data.

Assume the following model:

$$Y_{Ti} - Y_{Ri} \sim N(\theta, 2\sigma^2) \quad (22)$$

Where $\theta = \log(\mu_T/\mu_R)$ represents the difference in formulation. σ^2 is difference among individuals.

In the simulation we assume $Y_R \sim N(\mu_R, \sigma_R^2)$, $\mu_R=0$, $Y_T \sim N(\mu_T, \sigma_T^2)$, $\mu_R = -0.223$, -0.0512 . We follow the 0.8–1.25 criterion that if $-0.223 < \mu_T - \mu_R < 0.223$ then T and R are bioequivalence. Thus, when $\mu_T = -0.0512$, the two drugs are bioequivalence, and when $\mu_T = -0.223$, the two drugs are bioinequivalence.

Further, we assume that the prior distribution of the reference drug is $N(\mu_0, \sigma^2/\tau)$ $\mu_0 = 0$, $\tau = 0.1$. In order to study the influence of different boundary functions, our effectivity boundary values will obey the three boundary functions of Pocock, OBF and Power family.

Through simulation studies, we determine the boundary value that controls family-wise type I error rate below 0.05 while ensuring power above 80 %.

3.2 Simulation results

For each parameter setting, 10000 simulations were performed. To study the effect of Bayesian two-stage adaptive design, four cases of $P_I = 0.1, 0.2, 0.3$, and 0.4 were set. At the same time, in order to study the impact of time points on the results in the interim analysis, the sample size is respectively 1/3, 1/2 and 2/3 of the initial sample size, that is, to make interim decision at $t^* = 1/3, 1/2$, and $2/3$. The results obtained by the simulation and the parameters under different boundary functions are shown in Figure 2 and Figure 3 (only results at $P_I = 0.3$ are given in this section, and the others are given in Appendix A: Figure 4, Figure 5, Figure 6).

Figure 1: Bayesian two-stage adaptive design in bioequivalence.

Figure 2: P_{u1} value under different CVs (left: $t_1 = 1/3$; medium: $t_1 = 1/2$; right: $t_1 = 2/3$).

Figure 3: When $P_l = 0.3$, the P_{u2} boundary value under different CVs (left: $t_1 = 1/3$; medium: $t_1 = 1/2$; right: $t_1 = 2/3$).

As Figure 2 shows:

1. When $t_1 = 1/3, 1/2$ and $2/3$, the P_{u1} boundary under OBF is always the largest, Power family is in between, and Pocock is always the smallest, indicating that among the three methods, OBF is the most conservative with the minimal probability of early effectivity stopping in interim analysis.
2. With the same boundary shape, the earlier the interim analysis is performed, the smaller the corresponding P_{u1} is, which indicates that if the sample size for making decisions gets smaller, effectivity stopping boundary needs to be expanded to ensure overall power.
3. With the increase of inter-individual variation CV, P_{u1} tends to decrease continuously, indicating that when the inter-individual variation CV is large, the result estimation is unstable, so lower stopping condition is required.
4. The downward trend of P_{u1} will be continuously compensated by the increase in the sample size during the interim analysis.

As Figure 3 shows:

1. Contrary to P_{u1} , the P_{u2} boundary under OBF is the smallest, Power family is in between, and Pocock is always the biggest.

- Like P_{u1} , with the same boundary shape, the earlier the interim analysis is performed, the smaller the corresponding boundary value is. Thus, overall, the earlier the time for interim analysis is, the less available sample information is, and it is necessary to improve performance by lowering the effective threshold.
- With the increase of inter-individual variation CV, P_{u2} also decreasing, indicating that when the inter-individual variation CV is large, a lower stopping criterion is required.
- The decreasing trend of P_{u2} will also be compensated by the increase of sample size during the interim analysis, and the compensation will be more than that of the P_{u1} .

After designing the parameters, 10000 simulated tests were performed again under H_0 and H_1 conditions, and the results of type I error rate and power were obtained. See Table 1 and Table 2 (this section only gives the results when $P_1 = 0.3$, other results are shown in Appendix B: Table 6, Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13, Table 14, Table 15, Table 16, Table 17)

Table 1: Type I error rate at futility stop thresholds $P_1 = 0.3$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.0508	0.0508	0.0508	0.0499	0.0499	0.0499	0.0486	0.0486	0.0389
20 %	0.0493	0.0406	0.0498	0.0336	0.0476	0.0493	0.0333	0.0431	0.0461
30 %	0.0346	0.0463	0.0500	0.033	0.0448	0.0496	0.0304	0.0412	0.0472
40 %	0.0445	0.0471	0.0498	0.0331	0.0427	0.0459	0.0318	0.0461	0.0464
50 %	0.0398	0.0481	0.0499	0.033	0.0459	0.0496	0.0316	0.0417	0.0455
60 %	0.0396	0.0448	0.0499	0.0361	0.0477	0.0492	0.0347	0.0445	0.0491
70 %	0.0467	0.0497	0.496	0.0374	0.0495	0.0494	0.0364	0.0441	0.0461
80 %	0.0495	0.0494	0.0497	0.0362	0.479	0.0497	0.0370	0.0449	0.0492
90 %	0.0492	0.0494	0.0495	0.0361	0.0436	0.0499	0.0361	0.0435	0.0460
100 %	0.0495	0.0498	0.0497	0.0397	0.0472	0.0495	0.0321	0.0419	0.0463

Table 2: Power at futility stop thresholds $P_1 = 0.3$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.2282	0.2282	0.2282	0.8469	0.8469	0.8272	0.8463	0.8106	0.4736
20 %	0.8376	0.8100	0.8129	0.8019	0.8061	0.7869	0.8048	0.8079	0.8014
30 %	0.8007	0.8024	0.7699	0.8013	0.8011	0.7982	0.8005	0.8088	0.8012
40 %	0.8007	0.8008	0.7505	0.8013	0.8059	0.8036	0.8009	0.8103	0.8075
50 %	0.8020	0.8001	0.7788	0.8015	0.8012	0.7951	0.8019	0.8095	0.8020
60 %	0.8006	0.8011	0.7798	0.8028	0.8007	0.7954	0.8022	0.8085	0.8027
70 %	0.8019	0.7896	0.7653	0.8015	0.8035	0.7800	0.8027	0.8133	0.8035
80 %	0.8293	0.7853	0.7705	0.8011	0.8054	0.7848	0.8025	0.8029	0.8056
90 %	0.8100	0.8062	0.7848	0.8005	0.8001	0.8023	0.8010	0.8047	0.8033
100 %	0.8127	0.8070	0.7893	0.8017	0.8028	0.7974	0.8006	0.8029	0.8055

Table 1 shows that:

- When interim analysis includes 1/3 of the initial sample size, type I error rate is controlled below 0.05 except when CV=10 %, in which case the rate slightly exceeds 0.05.
- Under $t_1 = 1/3, 1/2$ and $2/3$, type I error rate of the OBF boundary function is basically slightly smaller than other functions, which also proves that the OBF is more conservative than the other two.

Table 2 shows that:

- When CV=10 % and interim analysis includes 1/3 of the initial sample size, statistical power is around 0.2 which fails to meet the requirements. It suggests that if interim analysis is too early and sample information is insufficient, effectiveness of the decision will be reduced. Therefore, in practice it is necessary to prudently select the appropriate time point for interim analysis.

2. Under $t_1 = 1/3, 1/2$ and $2/3$, the power of OBF and Power family boundary functions is similar. The power of Pocock boundary function in the cases of 0.33M and 0.5M is slightly lower than 0.8, indicating that the other two functions perform better.

As can be seen from Table 3 and Table 4:

1. As CV continues to increase, the average sample size of the test will also increase.
2. Under conditions H_0 and conditions H_1 , Pocock requires the smallest sample size when CV is the same, and the OBF requires the largest sample size and Power family is in between.
3. No obvious regularity can be observed in the impact of interim analysis time point on the sample size. According to the results, earlier interim analysis does not necessarily make sample size smaller, because premature interim analysis contains little information in existing data, which is not enough to make a decision.

Table 3: Sample size of bioequivalence at $P_I = 0.3, H_0$.

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	5.85	5.85	5.85	5.66	5.66	5.66	5.57	5.57	5.54
20 %	17.54	17.62	17.50	17.23	17.35	17.20	17.04	17.14	17.01
30 %	35.86	36.05	35.79	35.48	35.72	35.40	35.39	35.62	35.52
40 %	59.49	59.82	59.36	58.78	59.18	58.63	58.65	59.16	58.63
50 %	87.56	88.03	87.35	87.02	87.62	86.79	86.75	87.33	86.56
60 %	118.89	119.52	118.54	119.09	119.92	118.76	117.59	118.42	117.32
70 %	153.96	154.82	153.56	153.33	154.32	152.86	152.93	154.00	153.64
80 %	189.25	190.43	188.87	189.37	190.74	188.84	188.64	189.94	188.24
90 %	225.88	227.11	225.3	224.47	226.00	223.72	224.94	226.51	224.47
100 %	262.91	264.37	262.17	262.68	264.28	261.82	262.28	264.08	261.73

Table 4: Sample size of bioequivalence at $P_I = 0.3, H_1$.

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.74	6.74	6.74	6.31	6.31	6.31	6.09	6.09	5.72
20 %	20.22	20.65	20.05	19.01	20.03	18.84	17.06	18.27	16.86
30 %	41.01	41.88	40.68	38.55	40.68	37.85	34.26	37.03	33.81
40 %	67.87	69.30	67.27	62.43	67.28	62.34	56.23	61.72	55.93
50 %	98.48	101.20	97.46	93.97	99.38	92.12	82.90	89.67	81.81
60 %	134.42	137.69	132.92	127.28	134.35	124.54	113.15	123.55	111.14
70 %	172.72	176.45	170.89	164.64	173.86	160.99	144.41	156.62	142.85
80 %	212.91	217.29	211.22	202.39	213.43	197.68	181.67	197.71	178.25
90 %	253.18	260.07	250.86	238.58	253.79	231.42	214.19	232.69	209.52
100 %	294.58	302.17	291.03	282.08	297.01	274.91	247.18	264.86	243.02

We compare the proposed method with the fixed design, Potvin's two-stage design method B and C.

Table 5 shows the sample size required for the fixed design at different CV levels. Compared to Table 3 and Table 4, with a few exceptions, fixed design has a larger sample size. However, sample size varies little when CV is small, in which case two-stage design is not needed. In addition, compared with the results of Potvin, sample size of the proposed method is also better. In clinical trials, it should be noted that if CV is miscalculated, there will be a price in applying two-stage design. If sample size of interim analysis is too large for the real CV, it will result in a waste of sample size. If sample size of interim analysis is too small for the real CV, the amount of information will be too small and result in a larger sample size in re-estimation.

Table 5: Fixed design sample size.

CV	n
10	8
20	20
30	40
40	66
50	98
60	134
70	174
80	216
90	258
100	300

4 Conclusion

This paper mainly studies a Bayesian two-stage adaptive design and explores its application in bioequivalence trial. Unlike existing two-stage designs (such as Potvin's method B, C), Bayesian two-stage adaptive design incorporates historical information and expert information, and is flexibly combined with trial data to aid decision making. Secondly, the sample re-estimation strategy used in this paper is based on the ratio of the information in interim analysis to total information, which is easier to calculate than the Potvin method. Moreover, the two-stage design can be combined with diversified stop boundary functions, generating different results. Finally, under the condition of type I error below 0.05 and power above 80 %, the proposed method requires smaller sample size compared to Potvin method. Simulation results have proved that Bayesian two-stage adaptive design has certain application value in clinical practice, yet there remain limitations worthy of further discussion.

Above all, the simulation in this paper only considers the case of non-informative prior distribution, so prior information is not included, but the selection of prior information will have an impact on trial results. If prior information is inconsistent with experimental data, the Bayesian method will lose its advantages and instead lead to erroneous results. Therefore, the correct selection of different prior distributions is also a focus of future research.

Secondly, there are many methods for sample re-estimation strategy in clinical design, such as re-estimating sample size according to conditional efficiency, or continuously adjusting sample size until Bayesian prediction probability reach the specified level, but the calculation process is complicated. The sample re-estimation strategy used in this paper is relatively simple and is under the frequentist theory, which lacks flexibility compared to the first two methods, so it requires further attention in considering other better sample re-estimation strategies and simplifying calculations.

Finally, this paper studies three common stopping boundary functions whose advantages and disadvantages are explored, and the optimal one is selected. There are actually many other forms of boundary functions. How to choose the appropriate boundary function according to actual needs is also a problem worthy of further study.

Acknowledgements

The research of F. Y. was supported in part by National Social Science Fund No.16BTJ021.

Appendix

A

Figure 4: When $P_l = 0.1$, the P_{u2} boundary value under different CV (left: $t_1 = 1/3$; medium: $t_1 = 1/2$; right: $t_1 = 2/3$).

Figure 5: When $P_l = 0.2$, the P_{u2} boundary value under different CV (left: $t_1 = 1/3$; medium: $t_1 = 1/2$; right: $t_1 = 2/3$).

Figure 6: When $P_I = 0.4$, the P_{u2} boundary value under different CV (left: $t_1 = 1/3$; medium: $t_1 = 1/2$; right: $t_1 = 2/3$).**B****Table 6:** Type I error rate at futility stop thresholds $P_I = 0.1$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.0515	0.0515	0.0515	0.0499	0.0499	0.0499	0.0495	0.0494	0.0392
20 %	0.0361	0.0409	0.0478	0.0393	0.0494	0.0494	0.0374	0.0463	0.0488
30 %	0.0349	0.0454	0.0498	0.0378	0.0488	0.0493	0.0361	0.0428	0.0495
40 %	0.0423	0.0455	0.0495	0.0363	0.0457	0.0487	0.0371	0.0481	0.0484
50 %	0.0418	0.0499	0.0493	0.0382	0.0488	0.0487	0.0359	0.0435	0.0481
60 %	0.0373	0.0457	0.0499	0.0424	0.0499	0.0499	0.0401	0.0474	0.0488
70 %	0.0440	0.0499	0.0493	0.0426	0.0500	0.0490	0.0392	0.0450	0.0481
80 %	0.0408	0.0499	0.0497	0.0403	0.0491	0.0498	0.0426	0.0481	0.0484
90 %	0.0430	0.0497	0.0499	0.0420	0.0496	0.0495	0.0397	0.0458	0.0480
100 %	0.0418	0.0496	0.0499	0.0443	0.0496	0.0497	0.0354	0.0439	0.0474

Table 7: Type I error rate at futility stop thresholds $P_I = 0.2$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.0509	0.0509	0.0509	0.0493	0.0493	0.0493	0.0495	0.0495	0.0391
20 %	0.0361	0.0410	0.0470	0.0379	0.0477	0.0494	0.0352	0.0443	0.0470
30 %	0.0345	0.0491	0.0500	0.0342	0.0468	0.0498	0.0334	0.0410	0.0481
40 %	0.0423	0.0483	0.0497	0.0341	0.0459	0.0465	0.0346	0.0466	0.0470
50 %	0.0408	0.0494	0.0493	0.0350	0.0471	0.0497	0.0339	0.0417	0.0470
60 %	0.0371	0.0464	0.0495	0.0381	0.0492	0.0493	0.0375	0.0465	0.0479
70 %	0.0452	0.0499	0.0495	0.0412	0.0500	0.0493	0.0374	0.0449	0.0467
80 %	0.0392	0.0493	0.0500	0.0383	0.0495	0.0495	0.0405	0.0461	0.0497
90 %	0.0435	0.0495	0.0500	0.0392	0.0443	0.0492	0.0383	0.0452	0.0472
100 %	0.0431	0.0499	0.0491	0.0408	0.0498	0.0499	0.0342	0.0432	0.0467

Table 8: Type I error rate at futility stop thresholds $P_I = 0.4$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.0506	0.0506	0.0506	0.0497	0.0497	0.0497	0.0498	0.0489	0.0486
20 %	0.0420	0.0404	0.0462	0.0499	0.0461	0.0494	0.0318	0.0483	0.0462
30 %	0.0352	0.0465	0.0474	0.0500	0.0432	0.0485	0.0295	0.0428	0.0466
40 %	0.0476	0.0480	0.0500	0.0495	0.0409	0.0451	0.0293	0.0459	0.0472
50 %	0.0384	0.0491	0.0494	0.0500	0.0446	0.0500	0.0306	0.0409	0.0446
60 %	0.0426	0.0472	0.0499	0.0500	0.0467	0.0495	0.0339	0.0425	0.0487
70 %	0.0485	0.0496	0.0496	0.0495	0.0465	0.0493	0.0353	0.0431	0.0452
80 %	0.0410	0.0500	0.0500	0.0499	0.0463	0.0493	0.0337	0.0439	0.0477
90 %	0.0459	0.0500	0.0497	0.0496	0.0433	0.0482	0.0340	0.0422	0.0454
100 %	0.0429	0.0497	0.0492	0.0494	0.0458	0.0498	0.0303	0.0411	0.0456

Table 9: Power at futility stop thresholds $P_I = 0.1$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.2294	0.2294	0.2294	0.7989	0.7989	0.7989	0.8266	0.7632	0.4737
20 %	0.8008	0.8022	0.8016	0.8023	0.7947	0.7643	0.8018	0.8023	0.8037
30 %	0.8025	0.8002	0.8022	0.8010	0.7987	0.7775	0.8043	0.8047	0.8050
40 %	0.8029	0.8018	0.7854	0.8024	0.8033	0.8028	0.8044	0.8138	0.8028
50 %	0.8025	0.7982	0.7710	0.8005	0.8028	0.7678	0.8017	0.8018	0.8055
60 %	0.8014	0.8013	0.7828	0.8018	0.7950	0.7754	0.8001	0.8064	0.7917
70 %	0.8005	0.7894	0.7718	0.8009	0.7881	0.7613	0.8032	0.8047	0.8023
80 %	0.8008	0.7840	0.7694	0.8024	0.7924	0.7667	0.8002	0.8030	0.7925
90 %	0.8029	0.8039	0.7826	0.8003	0.8008	0.7838	0.8011	0.8037	0.8004
100 %	0.8007	0.8072	0.7789	0.8016	0.8009	0.7758	0.8010	0.8007	0.8031

Table 10: Power at futility stop thresholds $P_f=0.2$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.2286	0.2286	0.2286	0.8193	0.8193	0.8193	0.8320	0.7992	0.4737
20 %	0.8018	0.8112	0.8009	0.8026	0.8019	0.7805	0.8010	0.8014	0.8031
30 %	0.8006	0.8133	0.8004	0.8018	0.8016	0.7944	0.8034	0.8035	0.8040
40 %	0.8018	0.8046	0.7873	0.8019	0.8151	0.8001	0.8032	0.8057	0.8018
50 %	0.8027	0.8014	0.7720	0.8004	0.8013	0.7842	0.804	0.8001	0.8037
60 %	0.8005	0.8067	0.7823	0.8016	0.8017	0.7856	0.8020	0.8043	0.7903
70 %	0.8003	0.7887	0.7677	0.8009	0.7995	0.7699	0.8015	0.8160	0.8008
80 %	0.8013	0.7907	0.7718	0.8021	0.8095	0.7795	0.8023	0.8013	0.7975
90 %	0.8004	0.8056	0.7968	0.8005	0.8012	0.7925	0.8036	0.8099	0.8049
100 %	0.8013	0.8063	0.7812	0.8003	0.8029	0.7881	0.8029	0.8049	0.8016

Table 11: Power at futility stop thresholds $P_f=0.4$.

CV	0.33M			0.5M			0.67M		
	OBF	Power	Pocock	OBF	Power	Pocock	OBF	Power	Pocock
10 %	0.2275	0.2275	0.2275	0.8620	0.8503	0.8433	0.8446	0.8382	0.7367
20 %	0.8008	0.8081	0.8006	0.8649	0.8051	0.7994	0.8007	0.8118	0.8070
30 %	0.8008	0.8020	0.8021	0.8676	0.8037	0.8021	0.8017	0.8063	0.8085
40 %	0.8002	0.8006	0.7896	0.8624	0.8021	0.8021	0.8019	0.8073	0.8035
50 %	0.8002	0.8009	0.7767	0.8545	0.8020	0.7999	0.8007	0.8026	0.8059
60 %	0.8016	0.8004	0.7878	0.8577	0.8016	0.8019	0.8016	0.8030	0.8047
70 %	0.8011	0.7891	0.7621	0.8500	0.8013	0.7877	0.8036	0.8095	0.8003
80 %	0.8005	0.7855	0.7686	0.8438	0.8054	0.7888	0.8031	0.8012	0.8030
90 %	0.8017	0.7984	0.7868	0.8476	0.8016	0.8016	0.8042	0.8008	0.8052
100 %	0.8015	0.8069	0.7871	0.8335	0.8005	0.7990	0.8005	0.8040	0.8020

Table 12: Sample size of bioequivalence at $P_f=0.1$, H_0 .

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.33	6.33	6.33	6.10	6.10	6.10	5.98	5.98	5.94
20 %	19.01	19.09	18.98	18.71	18.83	18.68	18.53	18.63	18.51
30 %	38.68	38.87	38.61	38.40	38.64	38.32	38.31	38.54	38.24
40 %	64.03	64.36	63.90	63.70	64.10	63.54	63.36	63.87	63.33
50 %	94.10	94.58	93.90	93.71	94.30	93.47	93.90	94.48	93.72
60 %	127.14	127.77	126.79	128.36	129.19	128.03	127.39	128.22	127.12

70 %	164.16	165.03	163.76	164.35	165.34	163.88	165.03	166.09	164.74
80 %	200.97	202.16	200.59	202.71	204.07	202.18	203.54	204.84	203.13
90 %	241.07	242.30	240.48	241.91	243.44	241.16	242.25	243.82	241.78
100 %	279.39	280.85	278.65	282.40	284.00	281.54	282.84	284.64	282.28

Table 13: Sample size of bioequivalence at $P_I=0.2$, H_0 .

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.08	6.08	6.08	5.89	5.89	5.89	5.78	5.78	5.75
20 %	18.26	18.33	18.22	17.94	18.06	17.91	17.76	17.86	17.74
30 %	37.15	37.34	37.08	36.81	37.04	36.72	36.82	37.05	36.75
40 %	61.50	61.83	61.37	61.05	61.45	60.90	60.90	61.41	60.88
50 %	90.50	90.98	90.29	90.18	90.78	89.95	90.15	90.73	89.97
60 %	122.70	123.33	122.35	123.38	124.21	123.05	122.07	122.90	121.80
70 %	158.60	159.46	158.20	158.48	159.47	158.01	158.59	159.66	158.31
80 %	194.62	195.81	194.24	195.72	197.09	195.20	195.73	197.02	195.32
90 %	232.68	233.91	232.09	232.50	234.03	231.75	233.53	235.10	233.06
100 %	270.37	271.84	269.64	271.62	273.21	270.76	272.07	273.87	271.51

Table 14: Sample size of bioequivalence at $P_I=0.4$, H_0 .

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	5.60	5.60	5.60	5.40	5.40	5.40	5.32	5.32	5.29
20 %	16.92	17.00	16.88	16.55	16.68	16.53	16.36	16.46	16.34
30 %	34.68	34.86	34.61	34.22	34.46	34.14	33.97	34.20	33.90
40 %	57.74	58.07	57.61	56.67	57.07	56.52	56.41	56.91	56.38
50 %	84.83	85.31	84.62	84.14	84.73	83.90	83.58	84.16	83.40
60 %	115.79	116.42	115.44	115.32	116.15	114.99	113.41	114.24	113.14
70 %	149.76	150.62	149.36	148.18	149.17	147.72	147.65	148.71	147.36
80 %	184.25	185.44	183.87	183.49	184.85	182.96	182.36	183.65	181.95
90 %	220.35	221.58	219.76	217.83	219.36	217.08	217.33	218.90	216.86
100 %	255.93	257.39	255.19	254.43	256.03	253.58	253.53	255.33	252.97

Table 15: Sample size of bioequivalence at $P_I=0.1$, H_1 .

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.78	6.78	6.78	6.32	6.32	6.32	6.09	6.09	5.72
20 %	20.45	20.88	20.28	19.10	20.11	18.93	17.09	18.30	16.88
30 %	41.52	42.39	41.19	38.75	40.87	38.04	34.36	37.13	33.91
40 %	68.69	70.12	68.09	63.76	67.60	62.67	56.36	61.85	56.05
50 %	99.61	102.34	98.59	94.48	99.89	92.63	83.10	89.88	82.02
60 %	136.02	139.28	134.52	127.96	135.03	125.22	113.49	123.90	111.49
70 %	174.86	178.59	173.03	165.50	174.72	161.85	144.84	157.05	143.28
80 %	215.51	219.89	213.82	203.68	214.72	198.97	182.14	198.18	178.72
90 %	255.93	262.82	253.60	240.07	255.28	232.91	214.65	233.15	209.98
100 %	297.85	305.44	294.30	283.51	298.44	276.34	247.89	265.57	243.72

Table 16: Sample size of bioequivalence at $P_I=0.2$, H_1 .

--	--	--	--	--	--	--	--	--	--

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.76	6.76	6.76	6.32	6.32	6.32	6.09	6.09	5.72
20 %	20.35	20.77	20.18	19.06	20.08	18.89	17.08	18.29	16.87
30 %	41.27	42.15	40.95	38.68	40.81	37.98	34.33	37.10	33.88
40 %	68.30	69.73	67.71	63.62	67.47	62.54	56.32	61.81	56.01
50 %	99.09	101.81	98.07	94.29	99.70	92.44	83.01	89.79	81.93
60 %	135.23	138.50	133.73	127.72	134.80	124.99	113.34	123.74	111.33
70 %	173.75	177.48	171.91	165.50	174.32	161.45	144.71	156.92	143.15
80 %	214.39	218.77	212.70	203.17	214.21	198.46	181.96	198.00	178.54
90 %	254.59	261.47	252.26	239.41	254.61	232.24	214.49	232.98	209.82
100 %	296.22	303.81	292.67	282.83	297.76	275.66	247.55	265.23	243.38

Table 17: Sample size of bioequivalence at $P_I=0.4$, H_1 .

CV	0.33M			0.5M			0.67M		
	Power	OBF	Pocock	Power	OBF	Pocock	Power	OBF	Pocock
10 %	6.71	6.71	6.71	6.29	6.29	6.29	6.08	6.08	5.71
20 %	20.06	20.48	19.89	18.92	19.94	18.75	17.02	18.23	16.81
30 %	40.62	41.50	40.29	38.40	40.52	37.69	34.20	36.98	33.76
40 %	67.31	68.74	66.71	63.17	67.02	62.08	56.07	61.56	55.77
50 %	97.82	100.55	96.80	93.54	98.94	91.68	82.65	89.42	81.56
60 %	133.36	136.62	131.86	126.73	133.80	123.99	112.79	123.20	110.79
70 %	171.51	175.25	169.68	163.93	173.15	160.28	144.04	156.25	142.48
80 %	211.30	215.68	209.61	201.55	212.59	196.84	181.31	197.35	177.89
90 %	251.33	258.21	249.00	237.57	252.78	230.41	213.64	232.14	208.97
100 %	292.59	300.17	289.04	280.87	295.80	273.70	246.60	264.28	242.43

References

- [1] Kieser M, Rauch G. Two-stage designs for cross-over bioequivalence trials. *Stat Med.* 2015;34:2403–16.
- [2] Selwyn MR, Dempster AP, Hall NR. A bayesian approach to bioequivalence for the 2 x 2 changeover design. *Biometrics.* 1981;37:11–21.
- [3] Grieve AP. A bayesian analysis of the two-period crossover design for clinical trials. *Biometrics.* 1985;41:979–90.
- [4] Molins E, Cobo E, Ocaña J. Two-stage designs versus European scaled average designs in bioequivalence studies for highly variable drugs: Which to choose? *Stat Med.* 2017;4777–88.
- [5] Melander H. Problems and possibilities with the add-on subject design. *Bio-Int. Bioavailability Bioequivalence pharmacokinetics. Med-pharm Stuttgart.* 1992;35:85–90.
- [6] Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. Sequential design approaches for bioequivalence studies with crossover designs. *Pharm Stat.* 2010;7:245–62.
- [7] Simon R. Optimal two-stage designs for phase ii clinical trials. *Contemp Clin Trials.* 1989;10:1–10.
- [8] Tan S, Machin D. A bayesian re-assessment of two phase ii trials of gemcitabine in metastatic nasopharyngeal cancer. *Br J Cancer.* 2002;86:843.
- [9] Ghosh P, Rosner GL. A semi-parametric bayesian approach to average bioequivalence. *Stat Med.* 2007;26:1224–36.
- [10] Bandyopadhyay N, Dragalin V. Implementation of an adaptive group sequential design in a bioequivalence study. *Pharm Stat: J Appl Stat Pharm Ind.* 2007;6:115–22.
- [11] Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics.* 1995;51:1315–24.
- [12] Tsiatis AA. Information-based monitoring of clinical trials. *Stat Med.* 2006;25:3236–44.
- [13] Mehta CR, Tsiatis AA. Flexible sample size considerations using information-based interim monitoring. *Drug Inf J.* 2001;35:1095–112.
- [14] Wang, MD. Sample size reestimation by bayesian prediction. *Biometrical J: J Math Methods Biosci.* 2007;49:365–77.
- [15] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika.* 1977;64:191–9.
- [16] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics.* 1979;35:549–56.
- [17] Gordon Lan K, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika.* 1983;70:659–63.