MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
Vol. 12, No. 3, pp. 295–309, 2002

ON STATISTICAL POWER FOR AVERAGE BIOEQUIVALENCE TESTING UNDER REPLICATED CROSSOVER DESIGNS

Hansheng Wang* and Shein-Chung Chow

StatPlus, Inc., Heston Hall, Suite 206, 1790 Yardley-Langhorne Road, Yardley, PA 19067

ABSTRACT

In its recent guidance on bioequivalence, the U.S. Food and Drug Administration (FDA) recommends a two-sequence, four-period (2×4) replicated crossover design be used for assessment of population and individual bioequivalence [FDA. *Guidance for Industry on Statistical Approaches to Establishing Bioequivalence*; Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2001]. The recommended replicated crossover design not only allows estimates of both the inter-subject and the intra-subject variabilities and the variability due to subject-by-formulation interaction, but also provides an assessment of average bioequivalence (ABE). In this article, power function for assessment of ABE under a general replicated crossover design (i.e., a $2 \times 2m$ replicated crossover design) based on the traditional analysis of variance model and the mixed effects model as suggested by the FDA are studied. It is found that the power of a $2 \times 2m$ replicated crossover design depends upon the variability due to subject-by-formulation interaction and the number of replicates. Based on the derived power function, formula for sample size calculation for assessment of ABE under a $2 \times 2m$ replicated crossover design is also provided.

*Corresponding author. E-mail: hwang@statplusinc.com

DOI: 10.1081/BIP-120014560
1054-3406 (Print); 1520-5711 (Online)

www.dekker.com

Key Words: Replicated crossover design; Average bioequivalence; Subject-by-formulation interaction; Power

INTRODUCTION

In the United States, when a brand-name drug is going off patent, generic companies may file an abbreviated new drug application (ANDA) for generic approval by providing evidence of bioequivalence in average bioavailability between the generic drug and the brand-name drug. The bioavailability of a drug is defined as the rate and extent to which the active drug ingredient or therapeutic moiety is absorbed and becomes available at the site of drug action. The rate and extent of drug absorption are usually measured by the maximum blood or plasma concentration ($C_{\max}$) and area under the blood or plasma concentration–time curve (AUC). As indicated in the guidances on bioequivalence published by the Food and Drug Administration (FDA), evidence of bioequivalence in average bioavailability can be demonstrated through the conduct of a bioequivalence trial under a standard two-sequence, two-period (2×2) crossover design.[1,2] Two formulations of the same drug or two drug products are said to be bioequivalent in average bioavailability if the 90% confidence interval of the ratio of means of the primary pharmacokinetic parameters such as AUC and $C_{\max}$ are within the interval of 80 and 125% based on log-transformed data.[2]

In recent years, as more generic drug products become available in the marketplace, it is a concern whether the generic drug products are comparable in quality, safety, and efficacy as compared to the brand-name drug. As a result, it is of concern whether the generic drug products and the brand-name drug can be used interchangeably. Drug interchangeability is usually classified as drug prescribability and drug switchability. Drug prescribability is referred to as the physician's choice for prescribing an appropriate drug for his/her new patients among the generic drug products and the brand-name drug, while drug switchability is related to the switch from a drug product to an alternative drug product within the same patient whose concentration of the drug product has been titrated to a steady, efficacious, and safe level.[3,4] To address drug prescribability and switchability, population bioequivalence (PBE) and individual bioequivalence (IBE) are proposed, respectively (see, e.g., Ref. [5]). In its recent guidance on PBE and IBE, the FDA recommends a two-sequence, four-period (2×4) replicated crossover design be used.[6] The recommended replicated crossover design not only allows estimates of both the inter-subject and the intra-subject variabilities and the variability due to subject-by-formulation interaction, but also provides a more accurate and reliable assessment of average bioequivalence (ABE). In this article, the power functions for assessment of ABE under (i) the standard 2×2 crossover design based on the analysis of variance model[1,2] and (ii) a general replicated crossover design (i.e., a $2 \times 2m$ replicated crossover design) based on the mixed effects model[6] are studied. We will refer to the analysis of variance model for assessment of ABE under the standard 2×2 crossover design as the traditional model.

The remaining part of this article is organized as follows. In the next section, a detailed discussion regarding the traditional model is provided, which includes statistical model, power analysis, sample size calculation, and an example. Similar discussion is also presented for the mixed effects model in the third section. The article is concluded with a discussion in the fourth section.

TRADITIONAL MODEL

In practice, ABE is usually established under a standard two-sequence, two-period (2×2) crossover design, which assumes that there is no subject-by-formulation interaction. It should be noted this assumption does not imply that there is no random subject effect. Instead, no subject-by-formulation interaction is an indication that the random subject effect is the same under both treatments. Thus, a similar model can be applied to a more general replicated crossover design, i.e., a $2 \times 2m$ crossover design.

Statistical Model

Consider a $2 \times 2m$ ($m > 1$) replicated crossover design comparing mean responses of a test drug and a reference drug. Let y_{ijkl} be the lth replicate or response ($l = 1, \ldots, m$) observed from the jth subject ($j = 1, \ldots, n$) in the ith sequence ($i = 1, 2$) under the kth treatment ($k = \mathrm{T}, \mathrm{R}$). The following analysis of variance model as suggested in the FDA guidance[1,2] is considered:

$$y_{ijkl} = \mu_k + \gamma_{ik} + s_{ij} + e_{ijkl}, \tag{1}$$

where μ_k is the kth treatment effect, γ_{ik} is the fixed effect of the ith sequence under treatment k, and s_{ij} is the random effect of the jth subject in the ith sequence. The $e_{ij\mathrm{T}l}$ and $e_{ij\mathrm{R}l}$ are assumed to be independent normal random variables with mean 0 and variances σ^2_{WT} or σ^2_{WR}, respectively. Let ϵ be the difference in mean pharmacokinetic responses between the test and the reference product, i.e., $\epsilon = \mu_{\mathrm{T}} - \mu_{\mathrm{R}}$ (test–reference). Also, let

$$\bar{y}_{ijk\cdot} = \frac{1}{m}(y_{ijk1} + \cdots + y_{ijkm}) \quad \text{and} \quad d_{ij} = \bar{y}_{ijT\cdot} - \bar{y}_{ijR\cdot}.$$

An unbiased estimator for ϵ is given by

$$\hat{\epsilon} = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} d_{ij\cdot}$$

Under model (1), $\hat{\epsilon}$ follows a normal distribution with mean ϵ and variance $\sigma^2_m/2n$, where

$$\sigma^2_m = \frac{1}{m}(\sigma^2_{\mathrm{WT}} + \sigma^2_{\mathrm{WR}}). \tag{2}$$

An unbiased estimator of σ_m^2 can be obtained by

$$\hat{\sigma}_m^2 = \frac{1}{2(n-1)} \sum_{i=1}^{2} \sum_{j=1}^{n} (d_{ij} - \bar{d}_{i\cdot})^2,$$

where

$$\bar{d}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} d_{ij}.$$

In order to establish ABE, the following interval hypotheses are usually considered:

$$H_0: |\epsilon| \geq \delta \text{ vs. } H_a: |\epsilon| < \delta.$$

The test drug is concluded bioequivalent to the reference in average bioavailability if the null hypothesis H_0 is rejected at the α level of significance when

$$\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m} < -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m} > t_{\alpha,2n-2}, \tag{3}$$

where $t_{\alpha,2n-2}$ is the αth upper percentile of a t-distribution with $2n - 2$ degrees of freedom.

Power Analysis

Under the alternative hypothesis that $|\epsilon| < \delta$, both

$$\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m} \quad \text{and} \quad \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m}$$

follow a noncentral t-distribution with noncentrality parameters given by

$$\frac{\sqrt{2n}(\epsilon - \delta)}{\sigma_m} \quad \text{and} \quad \frac{\sqrt{2n}(\delta + \epsilon)}{\sigma_m},$$

respectively. The power (denoted by $1 - \beta$) of Eq. (3) is then given by:

$$\begin{aligned} 1 - \beta &= P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m} < -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m} > t_{\alpha,2n-2} \right) \\ &= P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m} < -t_{\alpha,2n-2} \right) - P\left(\frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m} < t_{\alpha,2n-2} \right) \\ &\quad + P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m} > -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m} < t_{\alpha,2n-2} \right). \end{aligned}$$

The second term of the above probability can be rewritten as

$$P\left(\frac{\sqrt{2n}(\hat{\epsilon}-\delta)}{\hat{\sigma}_m} > -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon}+\delta)}{\hat{\sigma}_m} < t_{\alpha,2n-2}\right)$$
$$= P\left(\delta - \frac{t_{\alpha,2n-2}\hat{\sigma}_m}{\sqrt{2n}} < \hat{\epsilon} < \frac{t_{\alpha,2n-2}\hat{\sigma}_m}{\sqrt{2n}} - \delta\right),$$

which is relatively small. By ignoring this small probability, the power of Eq. (3) can be approximated by

$$1 - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta-\epsilon)}{\sigma_m}\right.\right) - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta+\epsilon)}{\sigma_m}\right.\right), \tag{4}$$

where $\mathcal{T}_{2n-2}(\cdot|\delta)$ represents the cumulative distribution function of a noncentral t-distribution with $2n-2$ degrees of freedom and the noncentrality parameter δ. As it can be seen, the power for establishing ABE under the traditional analysis of variance model mainly depends upon σ_m (the intra-subject variability). When m is arbitrarily large, σ_m could be arbitrarily small. Hence, the power could be arbitrarily close to 1.

To provide a better understanding, as an example, Fig. 1 plots the power against m for the case of $n = 20$, $\delta = 0.223$, $\epsilon = 0$, and $\sigma_{WT} = \sigma_{WR} = \sigma_e$. As it can be seen, the power increases as the number of replicates (i.e., m) increases. The impact of m on power is substantial when σ_e is large (e.g., $\sigma_e = 0.50$). When σ_e is small (e.g., $\sigma_e \leq 0.20$), the gain in power by increasing m is not significant.

Sample Size Calculation

Under the traditional analysis of variance model, sample size needed for achieving a desired power of $1-\beta$ can be estimated by solving the following equation:

$$1 - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta-\epsilon)}{\sigma_m}\right.\right) - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta+\epsilon)}{\sigma_m}\right.\right) = 1-\beta.$$

Note that the left-hand side of the above equation is larger than or equal to

$$1 - 2\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta-|\epsilon|)}{\sigma_m}\right.\right).$$

Thus, a conservative approximation to n can be obtained by solving

$$\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta-|\epsilon|)}{\sigma_m}\right.\right) = \frac{\beta}{2},$$

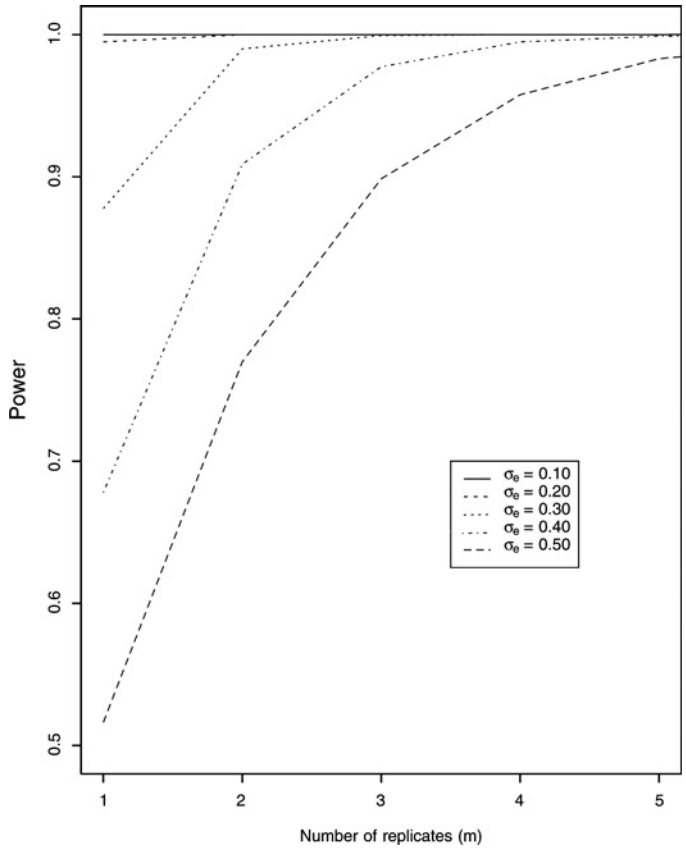

Figure 1. Power function for establishing ABE—traditional model.

which can be done by using Table 1 with $\theta = 2(\delta - |\epsilon|)/\sigma_m$. When n is large, this leads to

$$n = \frac{(z_\alpha + z_{\beta/2})^2 \sigma_m^2}{2(\delta - |\epsilon|)^2}. \tag{5}$$

However, when $\epsilon \neq 0$, the above sample size formula may be too conservative to be of practical interest. Alternatively, similar to the one proposed by Chow and Liu,[3] we may consider the following approach. Note that in Eq. (4), when $\epsilon \neq 0$,

$$\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta - |\epsilon|)}{\sigma_m}\right.\right)$$

is much larger than

$$\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta + |\epsilon|)}{\sigma_m}\right.\right).$$

Thus, the power can be further approximated by

$$1 - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left|\frac{\sqrt{2n}(\delta - |\epsilon|)}{\sigma_m}\right.\right).$$

The sample size needed can be obtained by setting the above equation equal to 0. Similarly, Table 1 with $\theta = 2(\delta - |\epsilon|)/\sigma_m$ is useful. More specifically, this can be done by referring to the column with significance level α and the power $1 - \beta$. Then, the sample size is given by the row with $\theta = 2(\delta - |\epsilon|)/\sigma_m$. When n is large, this leads to

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma_m^2}{2(\delta - |\epsilon|)^2}. \tag{6}$$

An Example

Consider a 2×4 crossover design (TRTR, RTRT) (i.e., $m = 2$). Suppose that $\sigma_e = 40\%$ based on log-transformed AUC data from pilot studies. According to the 2001 FDA guidance, the bioequivalence limit for log-transformed PK data is given by $(-0.223, 0.223)$. Suppose that the true mean difference between the test and reference is expected to be 5% (i.e., $|\epsilon| = 5\%$). The investigator wishes to have an 80% (i.e., $\beta = 0.20$) power for establishing ABE at 5% level of significance. It follows that

$$\sigma_m^2 = \frac{1}{2}(\sigma_{\mathrm{WT}}^2 + \sigma_{\mathrm{WR}}^2) = \sigma_e^2 = \frac{1}{2}(0.40^2 + 0.40^2) = 0.16.$$

According to Eq. (6), the sample size required can be obtained as follows:

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma_m^2}{2(\delta - |\epsilon|)^2} = \frac{(1.64 + 0.84)^2 0.16}{2(0.223 - |0.05|)^2} = 16.4 \approx 17.$$

As a result, 17 subject per sequence is needed in order to achieve an 80% for establishment of ABE.

On the other hand, the sample size can also be estimated based on noncentral t-distribution as described. It follows that

$$\theta = \frac{2(\delta - |\epsilon|)}{\sigma_m} = \frac{2(0.223 - 0.05)}{\sqrt{0.16}} = 0.865 \approx 0.87.$$

By referring to Table 1 under the column with $\alpha = 5\%$ and $1 - \beta = 80\%$, the required sample size per sequence is given by 18.

MIXED EFFECTS MODEL

As discussed earlier, the traditional model ignores the effect of the variability due to the subject-by-formulation interaction. As indicated by Chen et al.,[7]

Table 1. Smallest n with $\mathcal{T}_{2n-2}(t_{\alpha,2n-2}|\sqrt{n}\theta/\sqrt{2}) \leq \beta$

	$\alpha = 2.5\%$		$\alpha = 5\%$			$\alpha = 2.5\%$		$\alpha = 5\%$	
θ	$1-\beta = 80\%$	$1-\beta = 90\%$	$1-\beta = 80\%$	$1-\beta = 90\%$	θ	$1-\beta = 80\%$	$1-\beta = 90\%$	$1-\beta = 80\%$	$1-\beta = 90\%$
0.31	165	220	130	179	0.66	38	50	30	41
0.32	155	207	122	168	0.67	36	48	29	39
0.33	146	194	115	158	0.68	35	47	28	38
0.34	137	183	108	149	0.69	34	46	27	37
0.35	130	173	102	141	0.70	34	44	26	36
0.36	123	164	97	133	0.71	33	43	26	35
0.37	116	155	92	126	0.72	32	42	25	34
0.38	110	147	87	120	0.73	31	41	24	33
0.39	105	140	82	114	0.74	30	40	24	32
0.40	100	133	78	108	0.75	29	39	23	32
0.41	95	126	75	103	0.76	29	38	23	31
0.42	90	121	71	98	0.77	28	37	22	30
0.43	86	115	68	94	0.78	27	36	22	29
0.44	83	110	65	90	0.79	27	35	21	29
0.45	79	105	62	86	0.80	26	34	21	28
0.46	76	101	60	82	0.81	25	34	20	27
0.47	73	97	57	79	0.82	25	33	20	27
0.48	70	93	55	76	0.83	24	32	19	26
0.49	67	89	53	73	0.84	24	31	19	25

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

0.50	64	86	51	70	0.85	23	31	18	25		
0.51	62	82	49	67	0.86	23	30	18	24		
0.52	60	79	47	65	0.87	22	29	18	24		
0.53	57	76	45	62	0.88	22	29	17	23		
0.54	55	74	44	60	0.89	21	28	17	23		
0.55	53	71	42	58	0.90	21	27	16	22		
0.56	52	68	41	56	0.91	20	27	16	22		
0.57	50	66	39	54	0.92	20	26	16	21		
0.58	48	64	38	52	0.93	20	26	16	21		
0.59	47	62	37	50	0.94	19	25	15	21		
0.60	45	60	36	49	0.95	19	25	15	20		
0.61	44	58	34	47	0.96	19	24	15	20		
0.62	42	56	33	46	0.97	18	24	14	19		
0.63	41	54	32	44	0.98	18	23	14	19		
0.64	40	53	31	43	0.99	18	23	14	19		
0.65	39	51	30	42	1.00	17	23	14	18		

the variability due to the subject-by-formulation interaction will have an impact on drug switchability. Thus, it is suggested that the variability due to the subject-by-formulation interaction should be taken into consideration when assessing ABE.

Statistical Model

For assessment of ABE under a $2 \times 2m$ replicated crossover design, the FDA recommends the following mixed effects model be considered:

$$y_{ijkl} = \mu_k + \gamma_{ik} + s_{ijk} + e_{ijkl}, \tag{7}$$

where μ_k is the kth treatment effect, γ_{ik} is the fixed effect of the ith sequence under treatment k, and s_{ijk} is the random effect of the jth subject in the ith sequence under treatment k. $(s_{ij\mathrm{T}}, s_{ij\mathrm{R}})$, $i = 1, 2$, $j = 1, \ldots, n$ are assumed to be independent and identically distributed as bivariate normal random variables with mean 0 and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{\mathrm{BT}}^2 & \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}} \\ \rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}} & \sigma_{\mathrm{BR}}^2 \end{pmatrix},$$

where σ_{BT}^2 and σ_{BR}^2 are the inter-subject variabilities of the test drug and the reference drug, respectively, and ρ is the inter-subject correlation coefficient between the $s_{ij\mathrm{T}}$ and $s_{ij\mathrm{R}}$. The $e_{ij\mathrm{T}l}$ and $e_{ij\mathrm{R}l}$ are assumed to be independent normal random variables with mean 0 and variance σ_{WT}^2 and σ_{WR}^2, respectively (depending on the treatment). Define

$$\sigma_{\mathrm{D}}^2 = \sigma_{\mathrm{BT}}^2 + \sigma_{\mathrm{BR}}^2 - 2\rho\sigma_{\mathrm{BT}}\sigma_{\mathrm{BR}}.$$

σ_{D}^2 is usually referred to as the variability due to the effect of subject-by-formulation interaction, which reflects the heteroscedasticity of the subject random effect between the test drug and the reference drug. Similarly, an unbiased estimator of ϵ can be obtained as

$$\hat{\epsilon} = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} d_{ij}.$$

Under model (7), $\hat{\epsilon}$ follows a normal distribution with mean ϵ and variance $\sigma_m^{*2}/2n$, where

$$\sigma_m^{*2} = \sigma_{\mathrm{D}}^2 + \frac{1}{m}(\sigma_{\mathrm{WT}}^2 + \sigma_{\mathrm{WR}}^2). \tag{8}$$

An unbiased estimator of $\hat{\sigma}_m^{*2}$ can be obtained as

$$\hat{\sigma}_m^{*2} = \frac{1}{2(n-1)}\sum_{i=1}^{2}\sum_{j=1}^{n}(d_{ij} - \bar{d}_{i\cdot})^2,$$

where

$$\bar{d}_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} d_{ij}.$$

We then reject the null hypothesis and conclude ABE at the α level of significance if

$$\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m^*} < -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m^*} > t_{\alpha,2n-2}. \tag{9}$$

Power Analysis

Under the mixed effects model, the power can be similarly obtained as follows

$$\begin{aligned} 1 - \beta &= P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m^*} < -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m^*} > t_{\alpha,2n-2}\right) \\ &= P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \delta)}{\hat{\sigma}_m^*} < -t_{\alpha,2n-2}\right) - P\left(\frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m^*} < t_{\alpha,2n-2}\right) \\ &\quad + P\left(\frac{\sqrt{2n}(\hat{\epsilon} - \hat{\delta})}{\hat{\sigma}_m^*} > -t_{\alpha,2n-2} \text{ and } \frac{\sqrt{2n}(\hat{\epsilon} + \delta)}{\hat{\sigma}_m^*} < t_{\alpha,2n-2}\right). \end{aligned}$$

Thus, the power of Eq. (9) can be approximated by

$$1 - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\middle|\frac{\sqrt{2n}(\delta - \epsilon)}{\sigma_m^*}\right) - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\middle|\frac{\sqrt{2n}(\delta + \epsilon)}{\sigma_m^*}\right). \tag{10}$$

The power for establishing ABE depends upon σ_m. When σ_m can be arbitrarily small, the power can also be arbitrarily close to 1. However, according to Eq. (8), $\sigma_m \to \sigma_{\mathrm{D}}$ if $m \to \infty$. Hence, unless $\sigma_{\mathrm{D}} = 0$, which means there is no subject-by-formulation interaction, the power cannot be arbitrarily large. An approximate upper limit is given by

$$1 - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\middle|\frac{\sqrt{2n}(\delta - \epsilon)}{\sigma_{\mathrm{D}}}\right) - \mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\middle|\frac{\sqrt{2n}(\delta + \epsilon)}{\sigma_{\mathrm{D}}}\right).$$

To provide a better understanding, a plot of the power for the case where $n = 20$, $\delta = 0.223$, $\epsilon = 0$, and $\sigma_{\mathrm{WT}} = \sigma_{\mathrm{WR}} = 0.30$ is shown in Fig. 2. Figure 2 exhibits a similar pattern as that of the power under the traditional model. For example, the power increases as the number of replicates increases. The gain of power is relatively significant as m increases from $m = 1$ to $m = 2$. However, it should also be noted from Fig. 2 that the power cannot be arbitrarily close to 1 by

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

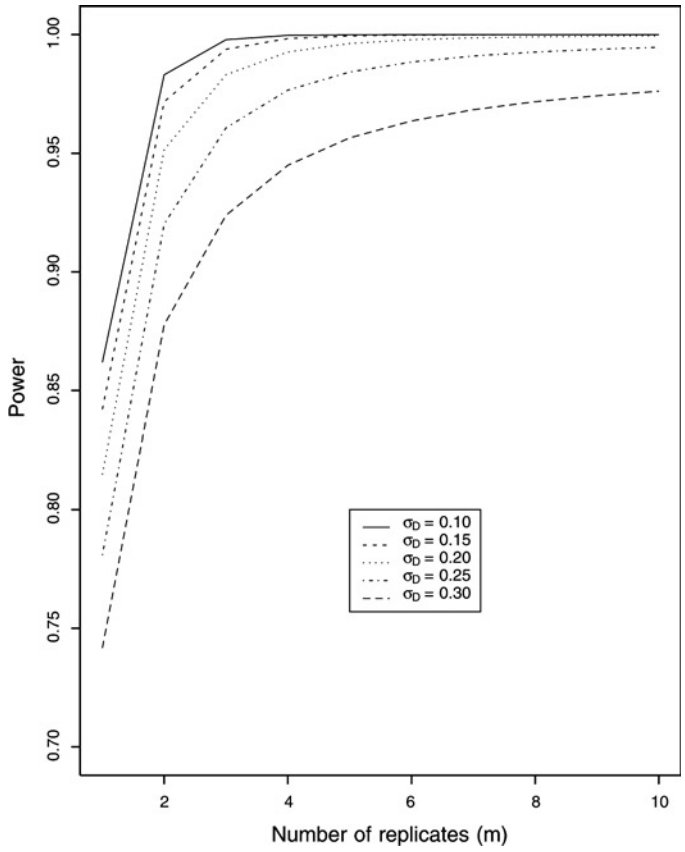

Figure 2. Power function for establishing ABE—mixed effects model.

increasing m if σ_D is not negligible. The power goes to a limit, which is smaller than 1, as m goes to infinity. It should also be noted that this limit decreases as σ_D increases.

Sample Size Calculation

To achieve a desired power of $1 - \beta$, the sample size needed can be obtained as follows:

$$n = \frac{(z_\alpha + z_{\beta/2})^2 \sigma_m^{*2}}{2(\delta - |\epsilon|)^2}. \tag{11}$$

Similarly, when $\epsilon \neq 0$, the alternative approach as described in the previous section leads to

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma_m^{*2}}{2(\delta - |\epsilon|)^2}. \tag{12}$$

An Example

Consider the same example as described in the previous section but with $\sigma_D^2 = 0.20$. We have

$$\sigma_m^{*2} = \sigma_D^2 + \frac{1}{2}(\sigma_{WT}^2 + \sigma_{WR}^2) = 0.20^2 + \frac{1}{2}(0.40^2 + 0.40^2) = 0.20.$$

According to Eq. (12), the sample size required for achieving an 80% power at the 5% level of significance is given by

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma_m^{*2}}{2(\delta - |\epsilon|)^2} = \frac{(1.64 + 0.84)^2 0.20}{2(0.223 - |0.05|)^2} = 20.5 \approx 21.$$

As a result, 21 subjects per sequence are needed in order to achieve 80% power in establishing ABE.

On the other hand, the required sample size based on the approach of a noncentral *t*-distribution can also be obtained. It follows that

$$\theta = \frac{2(\delta - |\epsilon|)}{\sigma_m} = \frac{2(0.223 - 0.05)}{\sqrt{0.20}} = 0.77.$$

By referring to Table 1 under the column with $\alpha = 5\%$ and $1 - \beta = 80\%$, the sample size per sequence is given by 22.

Discussion

In this article, we focus on $2 \times 2m$ replicated crossover designs. When $m = 1$, it reduces to the standard two-sequence, two-period crossover design. The standard 2×2 crossover design suffers the following disadvantages: (i) it does not allow independent estimates for the intra-subject variabilities because each subject receives each treatment only once; (ii) the effects of sequence, period, and carryover are confounded and cannot be separated under the study design. The $2 \times 2m$ ($m \geq 2$) replicated crossover design, on the other hand, not only provides independent estimates of the intra-subject and inter-subject variabilities and variabilities due to subject-by-formulation interaction, but also allows separate tests of the sequence, period, and carryover effects under appropriate statistical contrasts.

For the purpose of comparison, the power of ABE testing and the variability of the point estimate ($\hat{\delta}$) under different models (the traditional model and the mixed effects model) are given in Tables 2 and 3, respectively. As it can be seen from Tables 2 and 3, the variance of the mean difference estimates ($\hat{\delta}$) decreases as m increases for both the traditional model and the mixed effects model. It, however, should be noted that in the traditional model, the variance can be arbitrarily small by increasing m for a fixed sample size. As a result, the power can be arbitrarily close to 1. For the mixed effects model, since

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

Table 2. Powers of ABE Testing

Model	Design: 2×2	Design: $2\times 2m$
Traditional model	$1-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2n}(\delta-\epsilon)}{\sqrt{\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$ $-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2n}(\delta+\epsilon)}{\sqrt{\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$	$1-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2mn}(\delta-\epsilon)}{\sqrt{\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$ $-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2mn}(\delta+\epsilon)}{\sqrt{\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$
Mixed effects model	$1-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2n}(\delta-\epsilon)}{\sqrt{\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$ $-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2n}(\delta+\epsilon)}{\sqrt{\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$	$1-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2mn}(\delta-\epsilon)}{\sqrt{m\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$ $-\mathcal{T}_{2n-2}\left(t_{\alpha,2n-2}\left\|\frac{\sqrt{2mn}(\delta+\epsilon)}{\sqrt{m\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2}}\right.\right)$

Table 3. Variances of Mean Difference Estimates

Model	Design: 2×2	Design: $2\times 2m$
Traditional model	$\frac{1}{2n}(\sigma_{WT}^2+\sigma_{WR}^2)$	$\frac{1}{2nm}(\sigma_{WT}^2+\sigma_{WR}^2)$
Mixed effects model	$\frac{1}{2n}(\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2)$	$\frac{1}{2nm}(m\sigma_D^2+\sigma_{WT}^2+\sigma_{WR}^2)$

the subject-by-formulation interaction cannot be reduced by simply increasing the number of the replicates (m), the variance of the $\hat{\delta}$ cannot be arbitrarily small for a fixed sample size. There is a lower limit for a fixed sample size. Hence, the power cannot be arbitrarily close to 1. It is also noted that the variability due to the subject-by-formulation interaction (σ_D) plays an important role in determining the power of passing ABE testing. When σ_D decreases, the actual power will increase.

In practice, under a $2\times 2m$ replicated crossover design, it should also be noted that ABE assessment based on Schuirmann's two one-sided tests procedure,[8] which is based on the traditional analysis of variance model (by ignoring σ_D) and the mixed effects model could result in a totally different conclusion regarding the passage of ABE. This issue could be controversial especially when the test results are marginal.

REFERENCES

1. FDA. *Guidance on Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design*; Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 1992.
2. FDA. *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations*; Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2000.
3. Chow, S.C.; Liu, J.P. *Design and Analysis of Bioavailability and Bioequivalence Studies*, 2nd Ed.; Marcel Dekker, Inc.: New York, 2000.
4. Chow, S.C.; Shao, J. *Statistics in Drug Research—Methodologies and Recent Development*; Marcel Dekker, Inc.: New York, 2002.
5. Chen, M.L. Individual Bioequivalence—A Regulatory Update. J. Biopharm. Stat. **1997**, *7*, 5–11.
6. FDA. *Guidance for Industry on Statistical Approaches to Establishing Bioequivalence*; Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2001.
7. Chen, M.L.; Patnaik, R.; Hauck, W.W.; Schuirmann, D.J.; Hyslop, T.; Williams, R. An Individual Bioequivalence Criterion: Regulatory Considerations. Stat. Med. **2000**, *19*, 2821–2842.
8. Schuirmann, D.J. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. J. Pharmacokinet. Biopharm. **1987**, *15*, 657–680.