Sankhyā : The Indian Journal of Statistics
Special Issue on Biostatistics
2000, Volume 62, Series B, Pt. 1, pp. 162–174

DETERMINING AVERAGE BIOEQUIVALENCE AND CORRESPONDING SAMPLE SIZES IN GENERAL ASYMMETRIC SETTINGS

By SAURABH MUKHOPADHYAY
and
JAMES A. BOLOGNESE
Merck Research Laboratories, Rahway

SUMMARY. A new approach has been developed for bioequivalence testing to avoid impractical rejections which may occur with the commonly used Westlake's (1972) symmetric interval procedure. These methods are shown to cover the original symmetric confidence interval approach as a special case. Also, it is of considerable interest from the manufacturer's point of view, to obtain the minimum sample size which guarantees a pre-specified power to the test. The calculation of this minimum sample size accounts for the plausible asymmetry in the equivalence range and plausible non-zero expected value of the true difference. Algorithms are provided with illustrations for both testing and sample size problems.

1. **Introduction**

Bioequivalence trials compare the pharmacological endpoints of different formulations of a given drug. These endpoints include different measurements for the absorption, distribution, metabolism, and elimination of the drug following administration into the human body.

In practice it is recognized that no two formulations will have exactly equal bioavailability profiles. Therefore, a bioequivalence trial specifies clinically meaningful limits and declares bioequivalence if there is sufficient evidence that the two formulations differ by no more than the specified limits. It is recognized that a different statistical formulation is required than the usual F-test from ANOVA to obtain significant evidence to declare bioequivalence. Westlake (1972, 76) first introduced a confidence interval approach for a formal assessment of bioequivalence.

AMS (1991) *subject classification.* 62F25, 62F03.
Key words and phrases. Bioequivalence; clinical trials; confidence intervals; crossover design; minimum sample size

Anderson and Hauck (1983) presented a hypothesis testing approach where the probability of the null hypothesis (no bioequivalence) is computed. If the probability is less than a predetermined level, bioequivalence is declared. A Bayesian procedure for bioequivalence was presented by Rodda and Davis (1980). Given the observed results of the study, this procedure computes the probability that the true difference is less than the clinically important difference. Bioequivalence is declared if a high value of the probability is observed. All these approaches declare or reject bioequivalence similarly in relatively clear cases. Since the confidence interval approach is used most commonly and since the FDA has adopted this approach as the regulatory requirement, this article only deals with this approach.

The usual symmetric confidence interval approach based on t-distribution was suggested by Westlake (1972). These confidence intervals are symmetric about the sample estimate of the unknown difference. If this confidence interval is contained within the acceptable interval, bioequivalence is decided. These procedures ignored the need for adjusting the confidence interval with respect to the unknown difference.

There are practical concerns for using a symmetric confidence interval approach. For example, this symmetric interval procedure rejects bioequivalence unintuitively when the confidence interval just overlaps any of the boundaries. Also, it is unlikely to have exact bioequivalence in any given situation; thus, we may have reasons to believe that there may exists some non-zero difference. Since in that case the distance between the true difference and one of the boundaries decreases, we may need a larger sample size to detect the bioequivalence. However, the usual computation available for the symmetric interval approach produces a minimum sample size which can have enough power only in the most optimistic case of exact bioequivalence.

This article considers a general method of constructing the confidence interval and uses it appropriately to declare or refute bioequivalence. An optimization procedure is adopted to adjust the interval to accommodate plausible asymmetry of the target range and plausible non-zero difference between the formulations and still maintain the power level.

For successful implementations, along with the derivation of minimum sample size, we also provide the algorithm for computation. This sample size accounts for the plausible asymmetry of the target range and possible non-zero difference to protect the users from failing the objective. It is pointed out how the conventional approaches can be obtained from this procedure as a special case.

Section 2 describes the symmetric confidence interval approach. In Section 3, we develop the optimal method for assessing bioequivalence. We derive the minimum sample size for the optimal method in Section 4. The methods described in these sections are appropriately illustrated with examples and algorithms. Finally, in Section 5 we discuss how the new methods can be useful in different situations from practical point of view.

2. Usual Symmetric Procedure

The objective of a bioequivalence trial is to find out whether or not the difference

between the formulations, θ, of a pharmacokinetic endpoint is within a pre-specified acceptance region $(L,\ U)$. With standard notation, denote $\hat{\theta}$ as the estimator of θ, where $\hat{\theta}$ follows normal distribution with mean θ and variance $\sigma^2_{\hat{\theta}}$. If $\hat{\sigma}_{\hat{\theta}}$ denotes the estimated standard error of $\hat{\theta}$ with ν degrees of freedom (d.f.), then using standard distribution theory, the quantity $(\hat{\theta}-\theta)/\hat{\sigma}_{\hat{\theta}}$ follows the Student's t-distribution with associated d.f. ν. Usually, the estimate $\hat{\theta}$ is obtained from linear combination of appropriately transformed responses from the experiment.

For example, consider a two-period crossover design, commonly used in bioequivalence trial. Let y_{1i} and y_{2i} denote the observed AUC from subject i, following administrations of formulation $F1$ and formulation $F2$, respectively. Commonly, for AUC and C_{max} parameters the acceptance bounds are specified in terms of ratios, but with the usual log-transformation, the bounds can be specified in terms of the differences. Suppose that $\log y_{ki}$, $(1 \le i \le n,\ k = 1, 2)$, are normally distributed with mean μ_k and common variance σ^2. Then denoting $x_i = \log y_{1i} - \log y_{2i}$, we can use the sample mean of the differences, $\bar{x}$, as $\hat{\theta}$ to estimate $\theta = \mu_1 - \mu_2$. With the ANOVA model used for a two-period crossover study, $\sigma^2_{\hat{\theta}} = \frac{2}{n}\,\sigma^2$ is estimated by $\frac{2}{n}\,s^2$, where, s^2 is the MSE from the ANOVA with associated d.f. $\nu = n - 2$.

Westlake's (1972, '76, '88) approach computes a $100(1-2\alpha)\%$ confidence interval for θ which is symmetric around the observed $\hat{\theta}$. With the notation described before, distribution of $(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}}$ is used to construct the symmetric interval:

$$\left(\hat{\theta} - \hat{\sigma}_{\hat{\theta}} \cdot t_{\nu;\alpha},\ \hat{\theta} + \hat{\sigma}_{\hat{\theta}} \cdot t_{\nu;\alpha}\right) \tag{1}$$

where $t_{\nu;\alpha}$, (now onwards will be denoted as t_α for notational convenience), is the upper α cut-off point from the t-distribution with associated d.f. ν. The test declares bioequivalence if the entire confidence interval (1) falls within the target interval $(U,\ L)$; and rejects bioequivalence otherwise. While the simplicity of the procedure is attractive, it opens criticism for impractical rejections of the bioequivalence in *marginal* cases as illustrated below.

Illustrative Example: Suppose that a 20% acceptance rule is to be used on the ratio of AUC means to declare bioequivalence. Thus, the target interval on the ratio is given by (0.80, 1.25). With a usual log-transformation to the AUC endpoint, the target interval for the difference becomes (-0.223, 0.223). Also suppose that a two-period crossover design was used with $n = 15$. Now, with an observed mean and standard error (both in the log-scale) of 0.10 and 0.06, respectively, the symmetric 95%[1] confidence interval (1) is given by (-0.03, 0.23). This confidence interval rejects the bioequivalence as it just overlaps with the boundary of the target interval. However, this rejection is not so clear from practical viewpoint. The observation suggests that with a very high probability the true mean θ is well within the target range. In the following we develop a rejection rule which avoids this type of marginal rejection.

[1]Because much current interest in clinical comparability methods exists and uses bioequivalence methodology, the more stringent $\alpha = 0.025$ level is used for the examples. The same feature of the proposed method in this paper apply similarly to $\alpha = 0.05$ level tests of bioequivalence using the conventional 90% confidence intervals.

3. Proposed Method on Testing Bioequivalence

Other methods also available in the literature to test bioequivalence are the two one-sided test (TOST), BHM method (Brown, Hwang and Munk, 1995), and the new union intersection test (IUT) method by Berger and Hsu (1996). Berger and Hsu also compared these methods. While both BHM and the IUT methods were shown to have more power than the TOST method, none of these two methods is uniformly more powerful than the other. More importantly, there is no minimum power guaranteed for any of these tests which is essential from both the manufacturer and regulatory points of view. In the proposed method described below we provide a confidence interval based method with a guaranteed minimum power.

There are several choices in obtaining a confidence interval with a fixed confidence level. A general definition that covers all such confidence intervals for θ with a $100(1-2\alpha)\%$ confidence level is given by

$$\left(\hat{\theta} - \hat{\sigma}_{\hat{\theta}} \cdot t_{\alpha_1},\ \hat{\theta} + \hat{\sigma}_{\hat{\theta}} \cdot t_{(2\alpha-\alpha_1)}\right), \tag{2}$$

for any α_1, $(0 < \alpha_1 < 2\alpha)$. The symmetric confidence interval given by (1) is only a special case of (2) with $\alpha_1 = \alpha$. From the sampling theory viewpoint, the symmetric interval is motivated from producing the smallest length confidence interval with a fixed α-level. However, in the bioequivalence trials the acceptance or rejection of the bioequivalence only depends on whether or not the entire interval is within the target range. The width of the confidence interval becomes irrelevant as soon as the interval overlaps with any boundary of the target range. Thus, the symmetric confidence interval (1) may not always be the practical choice for bioequivalence trials.

In the bioequivalence trial, we are required to verify whether the observed distribution guarantees a minimum pre-specified probability on the target range. We can use any of the confidence intervals of the form (2) suitably to declare the bioequivalence. It still would keep the overall type-I probability *small*. However, note that the testing procedure and hence the power of the test is determined with the choice of the α_1. Therefore, a prudent choice on α_1 in (2) should be made to maintain a desirably high power. Even with this restriction, we can avoid some impractical rejections illustrated in the previous section.

How to choose α_1? It can be shown from the derivations in the appendix that the power of detecting bioequivalence using a confidence interval of the form (2) is given by

$$\Psi(\theta) \quad = \quad P\left(\frac{L-\theta}{\hat{\sigma}_{\hat{\theta}}} + t_{\alpha_1} < T_\nu < \frac{U-\theta}{\hat{\sigma}_{\hat{\theta}}} - t_{(2\alpha-\alpha_1)}\right). \tag{3}$$

Thus, although α_1 can be any where from 0 to 2α to satisfy (2), the choice of α_1 influences the power $\Psi(\theta)$. Therefore, we should choose α_1 which will assure a guaranteed power. For fixed n and a predefined minimum power of $1-\beta$, we define the set I_{α_1} as the candidate values of α_1 which satisfies

$$\Psi(\theta) \quad \geq \quad 1-\beta, \tag{4}$$

for some $\theta \in (L,\ U)$. Thus, bioequivalence will be detected if for some $\alpha_1 \in I_{\alpha_1}$ the confidence interval falls within the target range (L, U). In other words, we declare bioequivalence if there exists valid $\alpha_1 (0 < \alpha_1 < 2\alpha)$ so that the confidence interval (2) falls within the equivalence range $(L,\ U)$ and α_1 satisfies (4). We can avoid computing I_{α_1} in the practical implementation using the algorithm given below.

ALGORITHM 1.

0. **BEGIN**

   If $\hat{\theta}$ is outside the target range

   then reject bioequivalence and go to Step III.

   else go to Step I.

I. **If** $\hat{\theta}$ is closer to right boundary (i.e.$U - \hat{\theta} \leq \hat{\theta} - L$),

   then solve $\hat{\theta} + \hat{\sigma}_{\hat{\theta}} \cdot t_{(2\alpha - \alpha_1)} \ = \ U$ for α_1 in $(0, 2\alpha)$;

   else solve $\hat{\theta} - \hat{\sigma}_{\hat{\theta}} \cdot t_{\alpha_1} \ = \ L$ for α_1 in $(0, 2\alpha)$.

   EndIf

II. **If** a valid solution exists from the above, say $\hat{\alpha_1}$,

   then { obtain from (3) the power curve $\Psi(\theta)$ at $\alpha_1 = \hat{\alpha_1}$.

   If the curve exceeds the minimum value $1 - \beta$ at some $\theta \in (L,\ U)$,

   then accept the bioequivalence and go to Step III.

   else reject bioequivalence and go to Step III.}

   else (i.e., if no valid solution exists from Step I) reject bioequivalence and go to Step III.

   EndIf

III. **STOP**

To compute the power in step II, we have the option of using the observed standard error estimate $(\hat{\sigma}_{\hat{\theta}})$ or using the historical estimate which was used at the design stage to power the study appropriately. As long as the procedure clearly describes which one to be used, either procedure should be acceptable.

Illustrative example: Working with the same example of Section 2, note that we can avoid the *marginal* rejection by adopting the above confidence interval procedure. Straightforward calculation shows that the 95% confidence interval with $\alpha_1 = 0.01852$ becomes (-.039, 0.222) and therefore it fits within the target range to claim the expected bioequivalence. Note that according to our rule, we still require to verify that this choice of α_1 will produce more than 80% power for some θ within the target range. Using $\hat{\sigma}_{\hat{\theta}} = 0.06$, $\alpha_1 = 0.01852$, and $n = 15$, the power curve $\Psi(\theta)$ computed on the interval (-0.223,0.223) by using the expression from (3) intercepts

the 80% level. Hence we declare bioequivalence. Also note that, the chance of false rejection in this case is bounded above by $\max(\alpha_1, 2\alpha - \alpha 1) = 0.03148$ which is marginally higher than $\alpha = 0.025$ of the symmetric case. While this slight increase in the chance of false acceptance has no real implication, declaring bioequivalence makes a practical difference towards a more intuitive choice.

As described in the algorithm and as illustrated above, our testing method followed two major steps to conclude bioequivalence. Firstly, we required an α_1 so that the entire confidence interval fits within the target range. Secondly, we required a guaranteed 80% power with that α_1.

Sometimes we may have an α_1 so that the confidence interval fits within the acceptance range, but we may not have enough power to accept bioequivalence. This draws a sharp contrast with the Bayesian method proposed by Rodda and Davis (1980). That approach would accept bioequivalence if such an α_1 exists without paying any regard to the power. However, note that while we use an asymmetric procedure to fit the confidence interval within the stipulated range, the type I error probability increases slightly. Thus a guaranteed power warrants such compromise which was absent in the Bayesian approach mentioned above. Following is an illustration where the power requirement is not met and hence we failed to conclude bioequivalence, and must be so.

Suppose that we observe the same mean (= 0.10) and slightly increased standard error (= 0.063), then at $\alpha_1 = 0.01258$ the confidence interval becomes (-0.059, 0.222) and thus fits within the target range. Note that with $\alpha_1 = 0.025$ (symmetric case), the power curve exceeds the 80% level (power at 0 is approximately 81%). However, the power curve stays below the required 80% level for the computed $\alpha_1 = 0.01258$. Thus we cannot declare bioequivalence with this α_1. In fact, in this case, there does not exist any α_1 which meets the power requirement and at the same time fits the confidence interval within the target range. Hence we can never declare bioequivalence in this case.

4. **Sample Size Determination**

New methods for bioequivalence testing described in previous sections are important from a practical point of view. Consequently, from the manufacturer's point of view it is important to find out the minimum sample size in order to implement the procedure appropriately.

Our goal in this section is to find the minimum sample size to declare bioequivalence with guaranteed minimum power of $1 - \beta$ when the target interval (L, U) is not necessarily symmetric and/or the investigator believes that, or wants to design as if, a non-zero difference may exist between the formulations. Let us define the sample size n as the total number of subjects in the study design. Note that finding the minimum sample size requires explicit expression of the standard error term showing it's relationship with the sample size. We will therefore use the common two-period crossover design where the relationship is $\hat{\sigma}^2_{\hat{\theta}} = \frac{2}{n} s^2$. While this specific design is important from a practical point of view, the method is easily extendible

to other designs as long as $\hat{\sigma}_{\hat{\theta}} \propto n^{-1/2}$. Also note that at the design stage we do not observe the sample variance and therefore we use a historical estimate or *guess* for s^2. Therefore, to distinguish this estimate with the unobserved s^2, it will be denoted by $\hat{\sigma}^2$, and whence $\hat{\sigma}^2_{\hat{\theta}}$ will be $\frac{2}{n}\,\hat{\sigma}^2$, for rest of this section.

In the special case of $\theta = 0$ and symmetric acceptance range $(-\Delta, \Delta)$, (i.e., $0 < -L = U = \Delta$), computation is available for the method described in Section 2 (see Westlake, 1988) and is given by the following iterative equation

$$n \quad \geq \quad \frac{2(t_{\nu;\alpha} + t_{\nu;\beta/2})^2 \hat{\sigma}^2}{\Delta^2} \tag{5}$$

This minimum sample size guarantees a power of $(1-\beta)$ if no true difference exists between the formulations (i.e. when $\theta = 0$). However, in many practical situations, a real non-zero difference exists between the formulations which can still be well within the equivalence range. Therefore, the minimum sample size computed at $\theta = 0$ may no longer assure the desired power of the test. Clearly, we need more samples to maintain the same power as the real θ approaches either boundary of the target range. Thus in case a non-zero difference is warranted at the design stage, it will be useful to adjust the minimum sample size accordingly.

Another practical concern relates to using the symmetric confidence interval. As discussed in Section 3, using the general form given by (2) rules out some impractical rejections of the symmetric interval approach. An expression of the minimum sample size assuring a pre-specified power is obtained in the following theorem:

THEOREM 1. *The minimum sample size $n^\star$ with a guaranteed minimum power of $(1-\beta)$ at a pre-specified $\theta \in (L,\ U)$ is given by $[\tilde{n}^\star]$ (minimum integer $\geq \tilde{n}^\star$), where*

$$\tilde{n}^\star \quad = \quad \min_{\substack{0<\alpha_1<2\alpha \\ 0<\beta_1<\beta}} \frac{2(t_{\alpha_1} + t_{\beta_1})^2}{(L-\theta)^2} \cdot \hat{\sigma}^2 \tag{6}$$

subject to

$$\frac{t_{\alpha_1} + t_{\beta_1}}{t_{(2\alpha-\alpha_1)} + t_{(\beta-\beta_1)}} \quad = \quad \frac{|L-\theta|}{|U-\theta|}. \tag{7}$$

The proof of the theorem is given in the Appendix. Note that the theorem is obtained by solving a constrained MINIMAX problem (A.5) which is hard to solve numerically. Also note that the minimum sample size in Theorem 1 actually depends on the relative position of the user specified difference within the target range. Thus, unlike the expression in (5), it adjusts for the plausible asymmetry of the target range with respect to the specified value of θ.

In the special case when the stipulated θ is at the mid-point of the target range, it is intuitive that a symmetric confidence interval should be used to maximize the power. Indeed we see that from the following corollary.

COROLLARY 1. *In case of $\theta = \frac{(L+U)}{2}$, the minimum sample size $[\tilde{n}^\star]$ from (6) is obtained at $\alpha_1 = \alpha$ and $\beta_1 = \beta/2$.*

While we again defer the proof to the Appendix, it is immediately recognized from the corollary that when θ is the mid-point of the target interval, the expression (6) subject to (7) reduces to the following single expression

$$n_0^\star \quad = \quad \left[\frac{2(t_\alpha + t_{\beta/2})^2}{\Delta^2} \cdot \hat{\sigma}^2\right]. \tag{8}$$

where Δ is the half width of the target range. Note that the above expression is identical to (5). Thus, sample size calculation for Westlake's symmetric confidence interval approach is obtained as a special case when a symmetric interval is used and the study is powered at $\theta = 0$.

Computation: While the Theorem 1 provides an understanding on the relationship of the minimum sample size with the other user-specified parameters in the problem, the expression is still in implicit form. Therefore, an iterative procedure must be applied to obtain the minimum sample size. An *S-Plus* code was made useful to implement the computation which is available on request from one of the authors.

Figure 1: Minimum sample size $n^\star$ with target range (-0.223, 0.223)

For illustration purpose we computed the minimum sample size $n^\star$ for some specimen input values and presented them in Figure 1 and Table 1. For optimal method, a 80% power is guaranteed throughout these computations. In Figure 1, we illustrate the increase in the value of $n^\star$ when either the SD increases or the speculated difference θ approaches either boundary of the target range (-0.223, 0.223).

Table 1: The minimum sample size using 95% confidence interval with guaranteed 80% power to detect bioequivalence.

				Optimal Method		Usual Method		
					Power at θ	Power at θ		Power at θ
L	U	SD	θ	$n^\star$	(using $n = n^\star$)	(using $n = n^\star$)	$n_0^\star$	(using $n = n_0^\star$)
-0.223	0.223	0.1	0.00	7	0.830	0.830	7	0.830
			0.02	7	0.820	0.810	7	0.810
			0.04	8	0.878	0.849	7	0.751
			0.06	8	0.832	0.767	7	0.653
			0.08	9	0.829	0.735	7	0.526
			0.10	11	0.833	0.725	7	0.390
			0.12	14	0.815	0.703	7	0.268
			0.14	20	0.806	0.697	7	0.174
			0.16	33	0.801	0.696	7	0.108
-0.223	0.223	0.15	0.00	12	0.812	0.812	12	0.812
			0.02	12	0.801	0.792	12	0.792
			0.04	13	0.818	0.787	12	0.735
			0.06	14	0.805	0.742	12	0.644
			0.08	17	0.823	0.736	12	0.531
			0.10	21	0.813	0.710	12	0.409
			0.12	29	0.812	0.711	12	0.295
			0.14	43	0.807	0.706	12	0.200
			0.16	72	0.800	0.700	12	0.128
-0.223	0.182	0.1	0.00	8	0.827	0.816	7	0.694
			0.02	9	0.863	0.827	7	0.615
			0.04	9	0.802	0.724	7	0.500
			0.06	11	0.819	0.717	7	0.372
			0.08	14	0.807	0.694	7	0.255
			0.10	21	0.816	0.710	7	0.165
			0.12	34	0.801	0.696	7	0.103
			0.14	72	0.800	0.700	7	0.063
			0.16	259	0.801	0.703	7	0.039
-0.223	0.182	0.15	0.00	15	0.838	0.828	12	0.669
			0.02	16	0.834	0.796	12	0.601
			0.04	18	0.826	0.756	12	0.502
			0.06	21	0.801	0.703	12	0.390
			0.08	29	0.805	0.702	12	0.282
			0.10	44	0.807	0.706	12	0.191
			0.12	75	0.804	0.704	12	0.122
			0.14	160	0.800	0.701	12	0.075
			0.16	579	0.800	0.702	12	0.044

As noted earlier, this acceptance interval comes from the usual log-transformed 20% acceptance range. By back transforming this to the original scale we obtain (0.80, 1.25). However, sometimes the stipulated 20% target range is (0.8, 1.2) which is symmetric in the original scale but not in the log-scale. In the log-scale it becomes (-0.223, 0.0.182). Table 1 computes $n^\star$ for each of theses two acceptance intervals and different choices of θ and SD. Both figure and table demonstrate that the $n^\star$ is smallest when the difference is the mid-point of the target range and it grows slowly at the beginning when the true difference approaches either boundary of the target range. Thus, in many situations when the investigator suspects, or wishes to design for, a small departure from the bioequivalence, a relatively small pay off in terms of increase in the sample size will guarantee the desired power. Table 1 also

compares the proposed optimal method results with the usual symmetric approach.

Under the usual method we first obtain the power of the test using the minimum sample size, $n^\star$, of the optimal method. With the symmetric acceptance region of (-0.223, 0.223), we have the same power at $\theta = 0$ in either method. This demonstrates that the symmetric interval method is a special case of the proposed optimal method when the true θ is at the mid-point of the acceptance range. Notice that when the asymmetric acceptance range of (-0.223, 0.182) is used, $\theta = 0$ is no longer at the mid-range value. Consequently, there is less power at $\theta = 0$ in the usual method as compared to the optimal method. However, we also notice that, as long as the true θ is very close to the mid-point of acceptance range, the usual approach still guarantees 80% or more power. A more drastic drop in the power is observed with this procedure if the conventional minimum sample size $n_0^\star$ is used. Since the computation of $n_0^\star$ ignores the relative position of θ within the target range, it remains the same at different values of θ. Thus, even if we want to stick to the usual procedure, it is always safe to use the sample size $n^\star$. Finally note that, although this shows a gain in power over using $n_0^\star$, it does not always satisfy the minimum power requirement. In order to achieve that, we should use the proposed optimal method.

5. **Discussion**

Our proposed method for constructing bioequivalence tests as described in Section 3 relaxes the symmetry condition around the observed estimate of the difference without compromising the overall type-I error-probability or power. Thus, if a 95% confidence interval is used then the overall type I error-probability is always bounded by 5% to conform with the regulatory requirement. To avoid marginal unintuitive rejection, this type I probability in the proposed method can exceed slightly from $\alpha = 0.025$, however, that is thought to be only a small compromise for increased power. This adjustment range is usually tight around 0.025 unless the width of the confidence interval is very tight as compared to the width of the target range. On the other hand, when the interval width is very small compared to the length of the acceptance range, it is unlikely that the adjustment will make any difference in the final inference. Finally, the proposed method guarantees a minimum power for using the adjusted interval which is an added assurance to both regulatory agencies and manufacturers.

Minimum sample size described in Section 4 guarantees a pre-specified power at some user specified difference between the formulations. As noted in the introduction it is unlikely that the bioavailabilities of two different formulations are exactly the same. From a practical point of view, therefore, appropriate consideration should be given to this fact throughout this study, especially at the design stage. For example, the commonly used sample size guarantees a power at exact bioequivalence. This sample size often does not have enough power to detect bioequivalence with a clinically insignificant departure from exact equality. As a result,

it becomes harder to detect bioequivalence from this study, and failure of any of these expensive studies is certainly undesirable to the manufacturers.

The odds of detecting bioequivalence could be improved with a prudent choice on the sample size. Often the investigator can provide with some realistic guesses on the difference which may exist between the formulations. Consequently, the method described in Section 4 can be used to compute the minimum sample sizes. Like other statistical procedures, we need to decide on the testing method before we observe the data. Since the optimal method has greater power than usual method to detect bioequivalence, going with the optimal method seems to be the rational choice. With a speculated non-zero difference, if it requires an unrealistically large number of subjects to detect bioequivalence, then it may be wise for the investigator to retreat from the study. Some other times, running a successful bioequivalence trial can be far more important than some modest increase in the sample size. In any case, the computations of these minimum sample sizes will be helpful to investigators in making an informed choice on the sample size and whether to run the study.

Appendix

PROOF OF THEOREM 1. Let the power of testing bioequivalence for any θ be denoted by $\Psi(\theta) = P(A|\theta)$, where A be the event that the confidence interval (2) is within the equivalence range $(L,\ U)$ for some α_1, $(0 < \alpha_1 < 2\alpha)$. Using routine algebraic simplification and noting that $\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \sim T_\nu$, the Student's t-distribution with ν d.f., the power $\Psi(\theta)$ is given by

$$\Psi(\theta) \quad = \quad P\left(\frac{L-\theta}{\hat{\sigma}_{\hat{\theta}}} + t_{\alpha_1} < T_\nu < \frac{U-\theta}{\hat{\sigma}_{\hat{\theta}}} - t_{(2\alpha-\alpha_1)}\right). \tag{A.1}$$

To guarantee a minimum power of $1-\beta$, using the above equation, there must exist β_1, $(0 < \beta_1 < \beta)$ such that

$$\frac{L-\theta}{\hat{\sigma}_{\hat{\theta}}} + t_{\alpha_1} \leq -t_{\beta_1} \quad \text{and} \quad \frac{U-\theta}{\hat{\sigma}_{\hat{\theta}}} - t_{(2\alpha-\alpha_1)} \geq t_{(\beta-\beta_1)}. \tag{A.2}$$

Now, when $\hat{\sigma}_{\hat{\theta}} = \sqrt{\frac{2}{n}}\hat{\sigma}$, (in a two-period crossover trial), then (A.2) can be expressed as

$$n \geq \max\left\{\frac{2(t_{\alpha_1}+t_{\beta_1})^2}{(L-\theta)^2},\ \frac{2(t_{(2\alpha-\alpha_1)}+t_{(\beta-\beta_1)})^2}{(U-\theta)^2}\right\}\hat{\sigma}^2. \tag{A.3}$$

The required minimum sample size, therefore, will be obtained by minimizing the iterative in equation (A.3) with respect to α_1 and β_1 and will be given by $[\tilde{n}^\star]$ (minimum integer $\geq \tilde{n}^\star$), where

$$\tilde{n}^\star \quad = \quad \min_{\substack{0<\alpha_1<2\alpha \\ 0<\beta_1<\beta}} \left[\max\left\{h(\alpha_1,\beta_1|\theta,L,\tilde{n}^\star),\ h(2\alpha-\alpha_1,\beta-\beta_1|\theta,U,\tilde{n}^\star)\right\}\right]\ \hat{\sigma}^2, \tag{A.4}$$

where $h(\gamma_1, \gamma_2|\theta, K, n) = \frac{2(t_{\gamma_1}+t_{\gamma_2})^2}{(K-\theta)^2}$ (note that t_{γ_i} implicitly includes n through the d.f.).

Suppose that the solution of (A.4) is attained at $(\alpha_1^\star, \beta_1^\star)$. Then we claim that $h(\alpha_1^\star, \beta_1^\star|\theta, L, \tilde{n}^\star) = h(2\alpha - \alpha_1^\star, \beta - \beta_1^\star|\theta, U, \tilde{n}^\star)$ holds, which is equivalent to

$$\frac{t_{\alpha_1^\star} + t_{\beta_1^\star}}{t_{(2\alpha-\alpha_1^\star)} + t_{(\beta-\beta_1^\star)}} \quad = \quad \left|\frac{L-\theta}{U-\theta}\right|. \tag{A.5}$$

If the claim is true, the minimization problem in (A.4) reduces to the minimization of any one of the two $h(\cdot)$ functions :

$$\tilde{n}^\star \quad = \quad \min_{\substack{0<\alpha_1<2\alpha \\ 0<\beta_1<\beta}} \left[\frac{2(t_{\alpha_1} + t_{\beta_1})^2}{(L-\theta)^2}\right] \hat{\sigma}^2 \tag{A.6}$$

subject to (A.5). Hence the assertion of the theorem will follow. Therefore it remains to prove that the claim is true. Suppose that the claim is false; then for some fixed L, U, and θ, there exists a solution $\tilde{n}^\star$ attained at $(\alpha_1^\star, \beta_1^\star)$ such that $d(\alpha_1^\star, \beta_1^\star) \neq 0$ where $d(\alpha_1, \beta_1) = h(\alpha_1, \beta_1|\theta, L, \tilde{n}^\star) - h(2\alpha - \alpha_1, \beta - \beta_1|\theta, U, \tilde{n}^\star)$. Without loss of generality if we can assume that $d(\alpha_1^\star, \beta_1^\star) > 0$ then the solution $\tilde{n}^\star$ using (A.4) will be given by $\tilde{n}^\star = h(\alpha_1^\star, \beta_1^\star|\theta, L, \tilde{n}^\star)\,\hat{\sigma}^2$. But, since $d(\alpha_1, \beta_1)$ is continuous with respect to α_1 and $d(\alpha_1^\star, \beta_1^\star) > 0$, there exists an $\varepsilon(> 0)$ such that $d(\alpha_1, \beta_1^\star) > 0$ for any $\alpha_1 \in (\alpha_1^\star - \varepsilon, \alpha_1^\star + \varepsilon)$. In particular, $d(\alpha_1^\star - \frac{\varepsilon}{2}, \beta_1^\star) > 0$. However, by definition, $h(\alpha_1, \beta_1^\star|\theta, L, \tilde{n}^\star)$ is increasing with α_1 and hence

$$h(\alpha_1^\star - \frac{\varepsilon}{2}, \beta_1^\star|\theta, L, \tilde{n}^\star)\,\hat{\sigma}^2 \quad < \quad h(\alpha_1^\star, \beta_1^\star|\theta, L, \tilde{n}^\star)\,\hat{\sigma}^2 = \tilde{n}^\star. \tag{A.7}$$

Thus, (A.7) together with $d(\alpha_1^\star - \frac{\varepsilon}{2}, \beta_1^\star) > 0$ guarantees existence of a lower solution than $\tilde{n}^\star$ at $\alpha_1 = \alpha_1^\star - \varepsilon/2$ and $\beta_1 = \beta_1^\star$ which contradicts that $\tilde{n}^\star$ is solution to (A.4). Hence, the claim is true and the theorem follows.

PROOF OF COROLLARY 1. We first minimize the quantity $(t_{\alpha_1} + t_{\beta_1})$ subject to

$$t_{\alpha_1} + t_{\beta_1} \quad = \quad t_{(2\alpha-\alpha_1)} + t_{(\beta-\beta_1)} \quad (0 < \alpha_1 < 2\alpha,\ 0 < \beta_1 < \beta). \tag{A.8}$$

Using Lagrange's multiplier, λ, and taking partial derivatives with respect to α_1 and β_1, respectively, we obtain

$$\frac{\partial t_{\alpha_1}}{\partial \alpha_1} - \lambda\left(\frac{\partial t_{\alpha_1}}{\partial \alpha_1} - \frac{\partial t_{(2\alpha-\alpha_1)}}{\partial \alpha_1}\right) = 0 \quad \text{and} \quad \frac{\partial t_{\beta_1}}{\partial \beta_1} - \lambda\left(\frac{\partial t_{\beta_1}}{\partial \beta_1} - \frac{\partial t_{(\beta-\beta_1)}}{\partial \beta_1}\right) = 0. \tag{A.9}$$

Now denoting $F(\cdot)$ and $f(\cdot)$ as distribution function and density, respectively, of the underlying t-distribution, we have for any $(0 < p < 1)$, $\frac{\partial F(t_p)}{\partial p} = f(t_p) \cdot \frac{\partial t_p}{\partial p}$; also, since $F(t_p) = 1 - p$, we get $\frac{\partial t_p}{\partial p} = -1/f(t_p)$. Thus we obtain from (A.9) upon simplification

$$1/\lambda = 1 + f(t_{\alpha_1})/f(t_{(2\alpha-\alpha_1)}) \quad \text{and} \quad 1/\lambda = 1 + f(t_{\beta_1})/f(t_{(\beta-\beta_1)}). \tag{A.10}$$

Eliminating λ from equations in (A.10) we obtain the required condition for the minimization:

$$f(t_{\alpha_1})/f(t_{(2\alpha-\alpha_1)}) \quad = \quad f(t_{\beta_1})/f(t_{(\beta-\beta_1)}). \tag{A.11}$$

Notice that $(\alpha_1, \beta_1) = (\alpha, \beta/2)$ satisfies above and (A.8) hence a solution. To prove that this solution is unique, suppose that there exists $\alpha_1 \neq \alpha$ as a solution to the minimization problem. Without loss of generality, we assume that $\alpha_1 > 2\alpha - \alpha_1$. Since the cut-off point decreases with the increase in its argument, then applying (A.8), we will have $\beta_1 < \beta - \beta_1$. However in that case, $f(t_{\alpha_1}) > f(t_{(2\alpha-\alpha_1)})$ and $f(t_{\beta_1}) < f(t_{(\beta-\beta_1)})$ which contradicts the requirement (A.11). Hence, the constrained minimization is attained uniquely at $(\alpha_1, \beta_1) = (\alpha, \beta/2)$.

Finally, when $\theta = \frac{U+L}{2}$ then the equality constraint in Theorem 1 reduces to (A.8). Since $\dfrac{2(t_{\alpha_1} + t_{\beta_1})^2}{(L-\theta)^2}\,\hat{\sigma}^2$ is an increasing function of $(t_{\alpha_1} + t_{\beta_1})$, the required minimum sample size $n^\star$ will attain where $(t_{\alpha_1} + t_{\beta_1})$ is maximized and hence the proof follows.

Acknowledgments. The authors thank Professor Pranab Sen and Professor Bimal Sinha for their useful comments on the article.

References

Anderson, S. and Hauck, W. W. (1983), A new procedure for testing equivalence in comparative bioavailability and other clinical trials, *Communication in Statistics - Theory and Methods*, **12**, 2663-2692.

Berger, R. L. and Hsu, J. C. (1996), Bioequivalence trials, intersection union tests and equivalence confidence sets (with comments and rejoinder), *Statistical Sciences*, **11**, 283-319.

Brown, L. D., Hwang, J. T. G., and Munk, A. (1995), An unbiased test for the bioequivalence problem., *Tehnical Report*, Cornell Univ.

Rodda, B. E. and Davis, R. L. (1980), Determining the probability an important difference in bioavailability, *Clinical Pharmacology and Therapeutics*, **28**, 247-252.

Westlake, W. J. (1988), Bioavailability and bioequivalence of pharmaceutical formulations, in Peace, K. E. (ed.), *Biopharmaceutical Statistics for Drug Development*, Mercel Dekker, New York, 1988, 329-352.

– – ––– (1981), Comments on Bioequivalence testing – A need to rethink, *Biometrics*, **37**, 591-593.

– – ––– (1979), Statistical aspects of comparative bioavailability trials, *Biometrics*, **35**, 273-280.

– – ––– (1976), Symmetrical confidence intervals for bioequivalence trials, *Biometrics*, **32**, 741-744.

Saurabh Mukhopadhyay and James A. Bolognese
Clinical Biostatistics
Department at Merck Research Laboratories
P.O. Box 2000, RY 33-404
Rahway, NJ 07065-0900, U.S.A.
E-mail: saurabh_mukhopadhyay@merck.com