

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture 2000 - 12th Annual Conference Proceedings


ESTABLISHING POPULATION AND INDIVIDUAL BIOEQUIVALENCE CONFIDENCE INTERVALS

Feng Yu

Linda J. Young

Gary R. Stevens

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Yu, Feng; Young, Linda J.; and Stevens, Gary R. (2000). "ESTABLISHING POPULATION AND INDIVIDUAL BIOEQUIVALENCE CONFIDENCE INTERVALS," *Conference on Applied Statistics in Agriculture*.
<https://doi.org/10.4148/2475-7772.1242>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

ESTABLISHING POPULATION AND INDIVIDUAL BIOEQUIVALENCE CONFIDENCE INTERVALS

Feng Yu¹, Linda J. Young¹, Gary R. Stevens²

¹Departments of Mathematics & Statistics and Biometry, University of Nebraska-Lincoln

²Mallinckrodt, Inc., St. Louis, Missouri

Abstract

Average bioequivalence is used to assess pharmacokinetic properties of proposed generic drug before they are marketed. The limitations of average bioequivalence have led the U.S. Food and Drug Administration to propose the use of population bioequivalence and individual bioequivalence. In this study, bootstrap confidence intervals were used to evaluate population bioequivalence and individual bioequivalence in the context of a 2 x 4 crossover experimental design. Two bioequivalence criteria were compared: the mean-squared difference criterion and a probability-based criterion. Simulations were conducted to study the properties of the bootstrap confidence intervals under each criterion in establishing population bioequivalence or individual bioequivalence. Various inter-subject, within-subject, and subject-by-formulation interaction variance components were considered.

1. Introduction

Suppose that a generic drug T has been developed and is proposed as an alternative to the reference drug R. Before marketing formulation T, it must be determined whether the two drugs are bioequivalent. When two formulations of the same drug or two drug products are claimed to be *bioequivalent*, it is assumed that they either provide the same therapeutic effect or are therapeutically equivalent. To better understand how bioequivalence is assessed, a few terms should be introduced. If two drug products contain identical amounts of the same active ingredient, they are *pharmaceutical equivalents*. If, instead, two drug products both contain an identical therapeutic moiety, but not necessarily in the same amount or dosage form or as the same salt or ester, they are identified as *pharmaceutical alternatives* to each other. Two drug products are identified as bioequivalent if (1) they are either pharmaceutical equivalents or pharmaceutical alternatives and (2) their rates and extents of absorption do not show a significant difference when administered at the same molar dose of the therapeutic moiety under similar experimental conditions (Chow and Liu 1992). This work will focus on the statistical methods used in the latter of these two criteria. In general, bioequivalence studies focus on the comparison of bioavailabilites using measurements of the blood concentration-time profile, such as the extent of drug absorption as measured by the Area Under the Curve (AUC), the maximum drug concentration (C_{\max}), and the time required to reach the maximum drug concentration (T_{\max}) (see Figure 1). In the United States, standards for bioequivalence studies are set by the U.S. Food and Drug Administration (FDA).

Traditionally, simple measures such as average bioequivalence (ABE) have been used to satisfy FDA standards of bioequivalence. Suppose two drug products are to be compared. Let R be the reference formulation and T the test formulation. These could be two formulations of the same drug or two different drug products. ABE is declared if the mean values of pharmacokinetic parameters measured after administration of the two formulations are sufficiently close; that is, ABE is declared if

$$-\Delta \leq \mu_T - \mu_R \leq \Delta, \quad (1)$$

where μ_T and μ_R are the means of the response variables Y_T and Y_R for the drug products T and R, respectively, and Δ is determined by the drug regulatory bodies, such as the FDA. When using AUC to measure bioavailability, equation (1) usually represents bioequivalence on the logarithmic scale. If τ_T and τ_R are the mean bioavailabilities of the test and reference drug products on the original scale, the bioequivalence may be defined as

$$\exp(-\Delta) \leq \frac{\tau_T}{\tau_R} \leq \exp(\Delta). \quad (2)$$

A typical choice for $\exp(\Delta)$ is 1.25. Thus, the means are generally taken to be sufficiently close if a 90% confidence interval for the ratio of the observed means of the formulations R and T fall within a bioequivalence limit, generally 80 to 125% for the ratio of the formulation averages. (Note: 0.80 is the reciprocal of 1.25.)

Anderson and Hauck (1990) noted limitations of ABE and proposed using population bioequivalence (PBE) and individual bioequivalence (IBE). For example, when a patient begins a new drug therapy and two formulations are available, the choice does not matter, at least with respect to bioavailability, if the distributions of between-subject responses for the two formulations are sufficiently close. Now ABE is based on a comparison of the mean responses and ignores other features of the distributions, such as the variance. Population bioequivalence (PBE) or prescribability is claimed if the response distributions of the formulations R and T are sufficiently similar. For some random variables, including normal ones, the distribution is determined by the first two moments. Thus measures of PBE have included comparisons of both the means and variances

Let Y_i be the response associated with formulation i . For two drug products to be PBE, Schall and Luus (1993) suggested that the between-subject differences $Y_T - Y_R$ and $Y_R - Y_R'$ should be sufficiently close; that is, the distribution of the between-subject difference of bioavailabilities in subjects receiving the test and reference drug products should be sufficiently close to the between-subject difference of bioavailabilities in subjects who receive the reference drug product. Based on this concept, they suggested two measures of bioequivalence. The first claims PBE if

$$E(Y_T - Y_R)^2 - E(Y_R - Y'_R)^2 \leq \Delta_P^2, \quad (3)$$

or equivalently, if

$$(\mu_T - \mu_R)^2 + \sigma_{BT}^2 - \sigma_{BR}^2 \leq \Delta_P^2 \quad (4)$$

where σ_{BT}^2 and σ_{BR}^2 are the between-subject variances of Y_T and Y_R , respectively, and Δ_P is the bioequivalence limit. The FDA (1999) draft guidelines recommend that the criterion in equations (3) and (4) be standardized giving the mean-squared difference criterion:

$$\frac{(\mu_T - \mu_R)^2 + (\sigma_{BT}^2 - \sigma_{BR}^2)}{\max(\sigma_{BR}^2, \sigma_{B0}^2)} \leq \theta_P \quad (5)$$

where σ_{B0}^2 is a specified constant and θ_P is the bioequivalence limit on the standardized scale. The left-side of equation (5) has the form

$$\frac{ABE \text{ limit}^2 + \text{variance factor}}{\text{variance}}. \quad (6)$$

This led the FDA (1999) in its draft guidelines to suggest taking

$$\theta_P = \frac{(\ln 1.25)^2 + \epsilon_P}{\sigma_{B0}^2}. \quad (7)$$

The $(\ln 1.25)^2$ is the square of the usual ABE limit. The recommendations are to set $\epsilon_P = 0.02$ and $\sigma_{B0}^2 = 0.2$, leading to $\theta_P = 0.35$.

The second criterion, the probability-based criterion, again uses the between-subject differences $Y_T - Y_R$ and $Y_R - Y'_R$ and declares PBE if

$$Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r) \geq \Delta_2 \quad (8)$$

where r and Δ_2 are bioequivalence limits. Schall and Luus (1993) suggested $r = \ln(1.25)$ and Δ_2 between -20 and -25%. Here $\Delta_2 = -20\%$ is used.

Even though two drugs may be PBE, the presence of subject-by-formulation interaction may

cause some patients who have been taking one formulation to have significantly different responses if they are switched to the other formulation. This led Anderson and Hauck (1990) to introduce the concept of individual bioequivalence (IBE) or switchability. Suppose a patient is on the reference formulation. If the distribution of within-subject differences $Y_T - Y_R$ is sufficiently close to that of the within-subject differences $Y_R - Y_{R'}$, then it does not matter whether the patient continues with the reference formulation or changes to the test formulation; that is, the patient may be switched.

Schall and Luus (1993) proposed a mean-squared difference criterion and a probability-based criterion for assessing IBE that are analogous to those for evaluating PBE. The primary difference in the criteria for PBE and IBE is that when considering the differences $Y_T - Y_R$ and $Y_R - Y_{R'}$, the PBE criteria use between-subject differences while the IBE criteria use within-subject differences. Thus, when working with within-subject differences, the mean-squared difference criterion for IBE based on equation (3) is

$$(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2 \leq \Delta_I^2 \quad (9)$$

where σ_{WT}^2 and σ_{WR}^2 are the within-subject variances of Y_T and Y_R , respectively, σ_D^2 is the subject-by-formulation interaction, and Δ_I is the bioequivalence limit. In a manner analogous to that in equation (5), equation (9) can be standardized, giving

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2)}{\max(\sigma_{WR}^2, \sigma_{W0}^2)} \leq \theta_I \quad (10)$$

where σ_{W0}^2 is the specified lower threshold for within-subject variance, and θ_I is the standardized IBE limit. Again noting the relationship to ABE, FDA (1999) suggests

$$\theta_I = \frac{(\ln 1.25)^2 + \epsilon_I}{\sigma_{W0}^2} \quad (11)$$

and recommends setting $\epsilon_I = 0.05$ and $\sigma_{W0}^2 = 0.2$, producing $\theta_I = 0.50$. The probability-based criterion for IBE may be stated as in equation (8), but the probabilities should now be based on within-subject differences instead of the between-subject differences used in assessing PBE.

The mean-squared difference criterion is recommended in the draft FDA guidelines (FDA 1999), but the probability-based criterion is not. In this study, the properties of bootstrap confidence intervals for the assessment of PBE and IBE using a 2 x 4 crossover experimental design are explored.

2. Methods

The model for a 2 x 4 crossover experimental design is

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl}, \quad i = 1, 2; j = 1, 2, \dots, n_i; k = R, T; l = 1, 2 \quad (12)$$

where Y_{ijkl} denotes the response of the j th subject within the i th sequence, receiving the k th treatment, in the l th treatment replicate for the subject in sequence i ; μ_k is the mean response from treatment k ; γ_{ikl} is the fixed effect of replicate l on treatment k in sequence i ; δ_{ijk} is the random effect of the j th subject receiving the k th treatment within the i th sequence; and ϵ_{ijkl} is the random error associated with Y_{ijkl} . Assume that the ϵ_{ijkl} are independently and identically distributed as

$$\epsilon_{ijkl} \sim D(0, \sigma_{WK}^2) \quad (13)$$

where σ_{WK}^2 is the within subject variance for treatment k . Note that the assumption of normality is not needed for the bootstrap method so a general distribution D is assumed. The random subject effects, $\delta_{ij} = (\delta_{ijR}, \delta_{ijT})$, is a bivariate variable with distribution

$$\delta_{ij} \sim D_2 \left(\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BT}^2 \end{pmatrix} \right) \quad (14)$$

where σ_{BR}^2 and σ_{BT}^2 are the between-subject variances for the reference R and test T formulations, respectively; and ρ is the correlation between the reference formulation R and test formulation T responses from subject j . The correlation coefficient ρ is related to the subject-by-formulation interaction variance component σ_D^2 as follows:

$$\sigma_D^2 = (\sigma_{BT} - \sigma_{BR})^2 + 2(1 - \rho)\sigma_{BT}\sigma_{BR}. \quad (15)$$

Note that for model (12), it is assumed that a sufficiently long wash-out period exists between the consecutive points in time that either formulation R or T is administered to a subject so that no carryover effects are present.

The statistical hypotheses of interest when evaluating PBE are

$$H_0: \frac{(\mu_T - \mu_R)^2 + (\sigma_{BT}^2 - \sigma_{BR}^2)}{\max(\sigma_{BR}^2, \sigma_{BT}^2)} \geq \theta_P \quad (16)$$

versus

$$H_1: \frac{(\mu_T - \mu_R)^2 + (\sigma_{BT}^2 - \sigma_{BR}^2)}{\max(\sigma_{BR}^2, \sigma_{B0}^2)} < \theta_P \quad (17)$$

Thus, PBE is claimed if the null hypothesis (16) is rejected in favor of the alternative (17). Note that the PBE equation (5), consistent with the alternative hypothesis (17), may be written as

$$(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2) - \theta_P \max(\sigma_{TR}^2, \sigma_{T0}^2) \leq 0. \quad (18)$$

Thus, PBE could be declared if the upper limit of an upper one-sided confidence interval on the left-side of equation (18) is less than zero.

Similarly, the statistical hypotheses being tested in evaluating IBE are

$$H_0: \frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WTR}^2)}{\max(\sigma_{WR}^2, \sigma_{W0}^2)} \geq \theta_I \quad (19)$$

versus

$$H_1: \frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WTR}^2)}{\max(\sigma_{WR}^2, \sigma_{W0}^2)} < \theta_I. \quad (20)$$

Hence rejection of the null hypothesis (19) in favor of the alternative (20) leads to a conclusion of IBE. Also, the IBE equation (10) may be written as

$$(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WTR}^2) - \theta_I \max(\sigma_{WR}^2, \sigma_{W0}^2) \leq 0. \quad (21)$$

If the upper limit of an upper one-sided confidence interval on the left-side of equation (21) is less than zero, IBE can be declared.

When working with the probability-based criterion, the hypotheses of interest based on equation (8) are

$$H_0: Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r) \leq \Delta_2 \quad (22)$$

versus

$$H_1: Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r) > \Delta_2. \quad (23)$$

Recall that $\Delta_2 = -0.20 = -20\%$. Thus, if the lower limit of a lower one-sided confidence interval on the left-side of equation (8) is greater than -0.20 , the null hypothesis (22) may be rejected in favor of the alternative (23). If the estimated probabilities are based on between-subject differences, PBE is declared. If, instead, they are based on within-subject differences, IBE is claimed.

Bootstrap confidence intervals will be used to set the confidence intervals when testing the hypotheses for both PBE and IBE. Three steps are involved in generating a bootstrap confidence interval for this setting. First, draw n_1 bootstrap samples $(Y_R^*, Y_T^*, Y_R^*, Y_T^*)$ with replacement from sample observations (Y_R, Y_T, Y_R, Y_T) in sequence 1 and n_2 bootstrap samples $(Y_T^*, Y_R^*, Y_T^*, Y_R^*)$ with replacement from sample observations (Y_T, Y_R, Y_T, Y_R) in sequence 2. The total sample size is $n = n_1 + n_2$. The second step is to calculate the estimates of $E(Y_T - Y_R)^2 - E(Y_R - Y'_R)^2$ and $Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r)$ from the bootstrap sample. Steps one and two are repeated N times where N is generally large, say 1000. The final step is to take the 95th percentile of the N bootstrap estimates of $E(Y_T - Y_R)^2 - E(Y_R - Y'_R)^2$ as the upper bound of a one-sided 95% confidence interval for the left-side of equation (18) when working with between-subject differences and equation (21) when working with within-subject differences. Similarly, the 5th percentile of the estimates of $Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r)$ is taken as the lower bound of the one-sided 95% confidence interval for the left-side of equation (8) where separate computations are made for between- and within-subject differences..

In the above description of the process of setting bootstrap confidence intervals, the need for estimates of $E(Y_T - Y_R)^2 - E(Y_R - Y'_R)^2$ and $Pr(|Y_T - Y_R| \leq r) - Pr(|Y_R - Y'_R| \leq r)$ on both the population and individual level was mentioned. Estimators for these quantities will now be presented. If there is no period effect, an unbiased estimator of the left-side of equation (18) is

$$(\bar{Y}_T^* - \bar{Y}_R^*)^2 - \hat{\sigma}_D^{2*}/n + (\hat{\sigma}_{TT}^{2*} - \hat{\sigma}_{RT}^{2*}) - \theta_p \max(\hat{\sigma}_{TR}^{2*}, \sigma_{T0}^2) \quad (24)$$

where \bar{Y}_T^* and \bar{Y}_R^* are the sample means of the bootstrap sample values of Y_T^* and Y_R^* ,

respectively; $\hat{\sigma}_{TT}^{2*}$ and $\hat{\sigma}_{RT}^{2*}$ are the sample variances of Y_T^* and Y_R^* pooled across sequences;

and $\hat{\sigma}_D^{2*}$ is the sample variance of $(Y_T^* - Y_R^*)$ pooled across sequences.

When using equation (8) to assess PBE, the two replicates of formulations T and R for each

subject were averaged. Thus, n pairs of (Y_T^*, Y_R^*) were formed. An unbiased estimator of the population quantity on the left-side of equation (8) is then

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i=1}^n I(|\bar{Y}_{Ti}^* - \bar{Y}_{Rj}^*| \leq r) - \frac{1}{n(n-1)/2} \sum_{i=1}^n \sum_{j=i+1}^n I(|\bar{Y}_{Ri}^* - \bar{Y}_{Rj}^*| \leq r) \quad (25)$$

where $I(\cdot)$ is the indicator function that takes the value 1 if the statement within the parentheses is true and 0 otherwise. The period effect was tested for significance and estimated using SAS's mixed procedure with the first-order factor analytic variance-covariance structure as suggested in the draft FDA (1999) guidelines. Values of Y_T^* and Y_R^* were adjusted to eliminate any significant period effect, and computations proceeded as described above.

In considering IBE, an unbiased estimator of the mean-squared difference criterion on the left-side of equation (8) is

$$\frac{1}{4n} \sum_{i=1}^n \left[(Y_{T1i}^* - Y_{R1i}^*)^2 + (Y_{T2i}^* - Y_{R1i}^*)^2 + (Y_{T1i}^* - Y_{R2i}^*)^2 + (Y_{T2i}^* - Y_{R2i}^*)^2 \right] - \frac{1}{n} \sum_{i=1}^n (Y_{R1i}^* - Y_{R2i}^*)^2 \quad (26)$$

An unbiased estimator of the left-side of equation (8) for the evaluation of IBE using the probability-based criterion is

$$\frac{1}{4n} \sum_{i=1}^n \left[I(|Y_{T1i}^* - Y_{R1i}^*| \leq r) + I(|Y_{T1i}^* - Y_{R2i}^*| \leq r) + I(|Y_{T2i}^* - Y_{R1i}^*| \leq r) + I(|Y_{T2i}^* - Y_{R2i}^*| \leq r) \right] - \frac{1}{n} \sum_{i=1}^n I(|Y_{R1i}^* - Y_{R2i}^*| \leq r) \quad (27)$$

If a significant period effect was present, the bootstrap sample values were adjusted to remove the period effect as described above when considering the estimate associated with PBE. These adjusted values were then used in equations (26) and (27).

A Monte Carlo study was conducted to study the properties of the mean-squared difference criteria and the probability-based criteria for PBE and IBE. Only the 2 x 4 crossover design was considered. The two treatments were the reference formulation R and the test formulation T. Two sequences were used in each simulated experiment. In sequence 1, the treatments were arranged in the order of RTRT. The treatment order was TRTR in sequence 2. Current interest

is in the analysis of the log-transformed response of AUC. Because these values are usually normally distributed, the normal distribution was used in simulating both the errors and subject effects given in equations (13) and (14). The numbers of subjects per treatment simulated were 3, 5, 10, 20, and 40. The treatment means for formulations R and T were both 100. Values of σ_D simulated were 0, 0.01, and 0.1. $\sigma_{WR} = \sigma_{WT} = 0.15$ in all cases. $\sigma_{BR} = \sigma_{BT} = 0.15, 0.30$ were simulated. In all cases simulated, the conditions for PBE and IBE were satisfied as described in equations (5), (8), and (10). For each combination of the parameters, 100 experiments were simulated. A hundred bootstrap samples were drawn from each simulated data set. The simulations were run using The SAS System, Version 8.0. The probability of concluding formulations R and T were bioequivalent was estimated based on the proportion of simulated samples for which the null hypothesis was rejected.

3. Results

The effect of inter-subject variability on the probability of concluding PBE and IBE is illustrated in Figure 2. No subject-by-formulation interaction is present ($\sigma_D = 0$). Recall that in all cases, the parameters being used should lead to the conclusion of bioequivalence. As would be expected, the probability of concluding either PBE (Figure 2a) or IBE (Figure 2b) increases with sample size. For both methods, increasing the between-subject variability led to reduced power. In assessing PBE, the mean-squared difference criterion has consistently higher power than the probability-based criterion. When $\sigma_{BR} = \sigma_{BT} = 0.15$, approximately eight subjects per sequence are sufficient to ensure at least 80% power for the mean-squared difference criterion whereas it takes approximately twelve subjects to achieve this level of power when $\sigma_{BR} = \sigma_{BT} = 0.30$. In contrast, the probability-based criterion requires about 54 and 60 subjects per sequence to attain 80% power when $\sigma_{BR} = \sigma_{BT} = 0.15$ and 0.30, respectively. Although the mean-squared difference criterion continues to have more power than the probability-based criterion when considering IBE, the difference in the two criteria is less than that observed for PBE. For the mean-squared difference criterion, fewer than ten subjects/treatment are needed to achieve at least 80% power for $\sigma_{BR} = \sigma_{BT} = 0.15$ or 0.30. The results for the probability-based criterion are somewhat confusing as the estimated power is greater for $\sigma_{BR} = \sigma_{BT} = 0.30$ than for $\sigma_{BR} = \sigma_{BT} = 0.15$. This may be due to either the low number of simulated data sets or the small number N of bootstrap samples per data set. Certainly further study is needed to determine whether this observation is an anomaly or in fact a property of this criterion.

The effect of subject-by-formulation interaction on the probability of concluding PBE and IBE is illustrated in Figure 3. In all cases, $\sigma_{WR} = \sigma_{WT} = 0.15$ and $\sigma_{BR} = \sigma_{BT} = 0.30$. The subject-by-formulation interaction was varied with $\sigma_D = 0.0, 0.01, \text{ and } 0.1$. For both PBE and IBE, the mean-squared difference criterion consistently had greater power than the probability-based criterion. Neither criteria showed a large change in power as σ_D increased from 0 to 0.01, but both showed a major drop in power as σ_D was increased to 0.1. The drop was especially great when assessing PBE. It should be noted that as σ_D continues to increase, the drugs R and T are no longer considered either PBE or IBE. Thus, this drop in power is anticipated.

4. Discussion

The power of the mean-squared difference and the probability-based criteria in evaluating PBE and IBE for a 2 x 4 crossover design was investigated through a Monte Carlo study. Because only 100 data sets were generated for each combination of parameters and 100 bootstrap samples were used to set the confidence limits for each data set, the results of this study should be considered to be preliminary. However, few simulation studies on bioequivalence trials have been reported in scientific journals (Sheiner 1992). More study is needed to provide a full understanding of the properties of the statistical methods being used.

Although the draft FDA (1999) guidelines do not suggest adjusting for significant sequence or period effects, those adjustments were made here. The effect of not adjusting these values should also be evaluated. The mean-squared difference criteria is presented in the draft guidelines, but the probability-based criteria is not.

The power associated with the mean-squared difference criterion was consistently higher than that of the probability-based criterion. One reason for that may have been the -20% threshold value that was used in the study. Schall and Luus (1993) noted that “it is more difficult to suggest and motivate a bioequivalence range for” the probability-based criteria “because there is little experience with statistics of that form.” Based on the results of this study, further consideration must be given to setting this value before the probability-based criteria can be considered suitable alternatives to the mean-squared difference criteria in evaluating PBE and IBE. Other bioequivalence criteria, such as the therapy window criterion (Lyer and Chu 2000), should also be explored.

Although subject-by-formulation interaction is not explicitly present in the development of PBE (see equations (5) and (10)), the impact of such interaction on power appears to be present and greater for PBE than for IBE. The reason for this is not yet clear. Increases in between-subject variability caused a reduction in power for both PBE and IBE.

Finally, the method of moments was used to estimate the variance components. Other approaches, such as REML (Chinchilli 1996, Endrenyl and Tothfalusi 1999), should also be explored.

5. Summary

Bootstrap confidence intervals were used to evaluate population bioequivalence and individual bioequivalence in the context of a 2 x 4 crossover experimental design. Two bioequivalence criteria were compared through Monte Carlo simulation: the mean-squared difference criterion and a probability-based criterion. The numbers of subjects per treatment simulated were 3, 5, 10, 20, and 40. The treatment means for formulations R and T were both 100. Values of the subject-by-formulation interaction (σ_{D}) simulated were 0, 0.01, and 0.1. The within-subject standard deviations were $\sigma_{\text{WR}} = \sigma_{\text{WT}} = 0.15$ in all cases. The between-subject standard deviations $\sigma_{\text{BR}} =$

$\sigma_{BT} = 0.15, 0.30$ were simulated. In all cases, the data were normally distributed, and the conditions for population and individual bioequivalence were satisfied. For each combination of the parameters, 100 experiments were simulated. SAS's mixed procedure was used to perform the analysis of each data set, assuming a first-order factor analytic variance-covariance structure. Adjustments were made for significant sequence and period effects. A hundred bootstrap samples were drawn from each simulated data set. The probability of concluding formulations R and T were bioequivalent was estimated based on the proportion of simulated samples for which the null hypothesis was rejected. The power associated with the mean-squared difference criterion was consistently higher than that of the probability-based criterion. Although subject-by-formulation interaction is not explicitly present in the development of population bioequivalence, such interaction appears to affect power, and the effect seems to be greater for population than for individual bioequivalence. Increases in between-subject variability caused a reduction in power for both population and individual bioequivalence. The current study is limited in that only 100 samples were generated for each combination of parameters, and only 100 bootstrap samples were used to set the confidence limits for each data set. In addition, the effects of non-normality, of unequal sample sizes, of assuming different covariance structures, and of using other estimation methods (such as REML) were not evaluated. The probability of declaring treatments to be either population or individual bioequivalent when in fact they are not should also be considered.

6. Acknowledgments

The authors want to express their appreciation to Mallinckrodt, Inc. for financial support and to Daryl Travnicek for programming assistance.

7. References

Albert, K.S. and R.B. Smith. 1980. Bioavailability assessment as influenced by variation in drug absorption. In K.S. Albert (ed), *Drug Absorption and Dispersion: Statistical Considerations*, American Pharmaceutical Association, Washington, D.C. Pp. 87-113.

Anderson, S. and W.W. Hauck. 1990. Consideration of individual bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* **18**: 259-273.

Chinchilli, V.M. 1996. Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine* **15**: 1619-1634.

Chow, S. and J. Lui. 1992. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, Inc: New York.

Endrenyl, L. and L. Tothfalusi. 1999. Subject-by-formulation interaction in determinations of individual bioequivalence: Bias and prevalence. *Pharmaceutical Research* **16**: 186-190.

FDA. 1999. *Guidance for Industry BA and BE Studies for Orally Administered Drug Products--General Considerations, Draft Guidance*. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER).

Liu, J. 1995. Use of the repeated cross-over design in assessing bioequivalence. *Statistics in Medicine* **14**: 1067-1078.

Lyer, H., P. Chapman, and Y.-L. Chu. 2000. Mixed linear models for individual bioequivalence. Proceedings of the Kansas State University Conference on Applied Statistics in Agriculture. Manhattan, Kansas, April 30-May 2. *In press*.

Schall, R. and H.G. Luus. 1993. On population and individual bioequivalence. *Statistics in Medicine* **12**: 1109-1124.

Sheiner, L.B. 1992. Bioequivalence revisited. *Statistics in Medicine* **11**: 1777-1788.

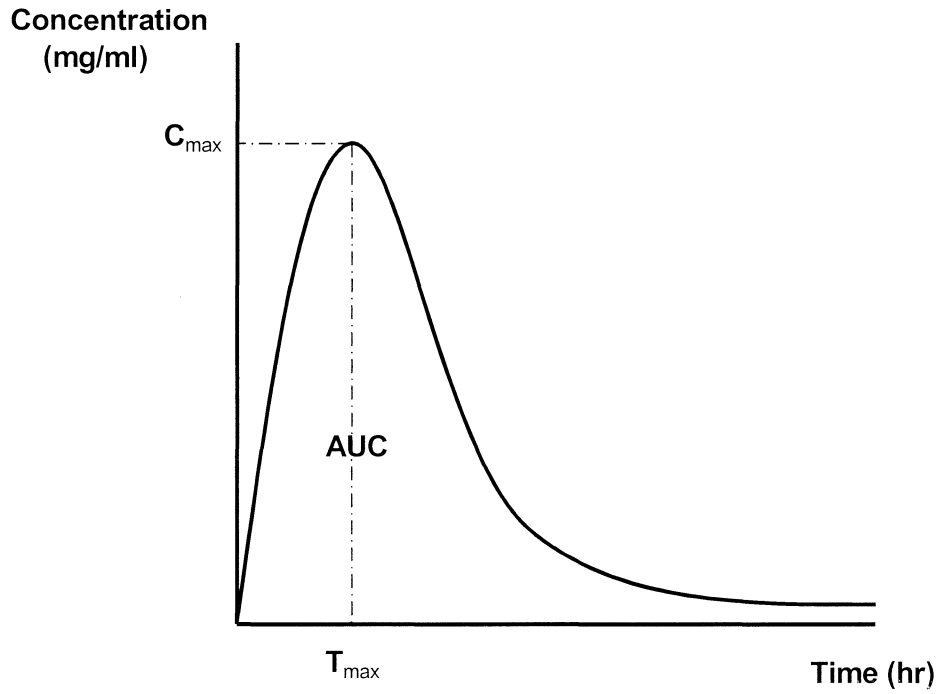
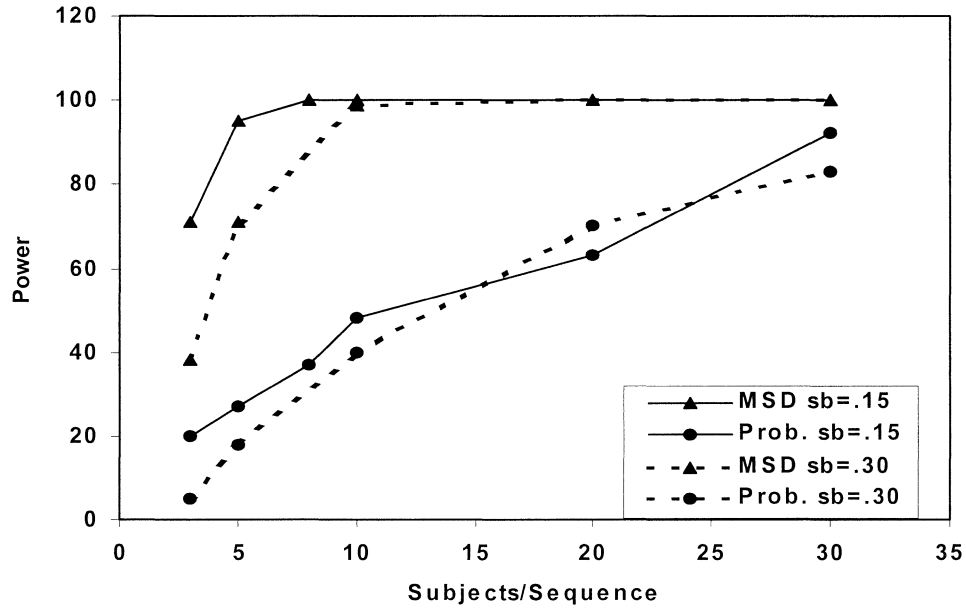
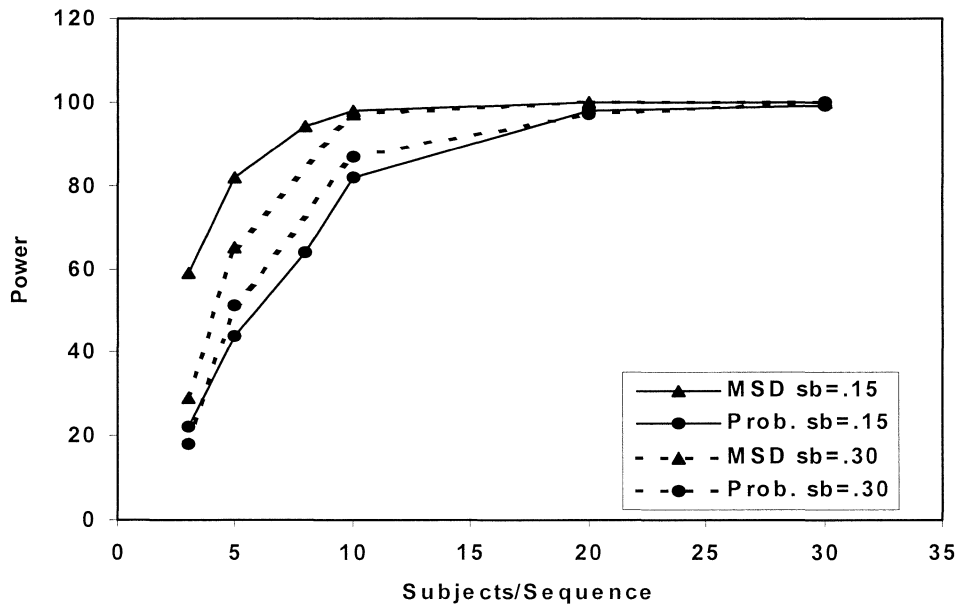


Figure 1: Common pharmacokinetic measurements from the blood concentration-time profile used in evaluating bioequivalence are the area under the curve (AUC), C_{\max} , and T_{\max} .

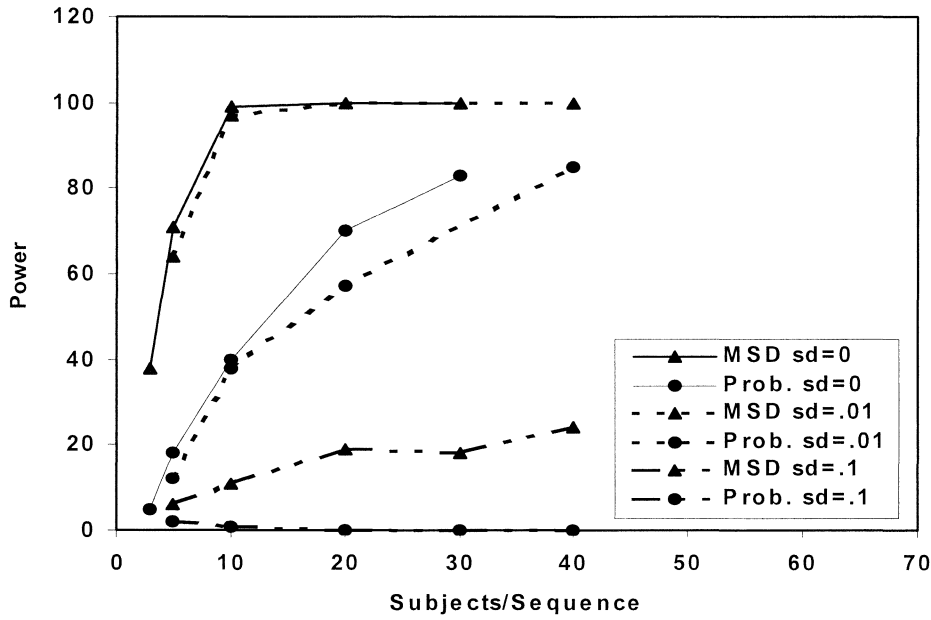


(2a)

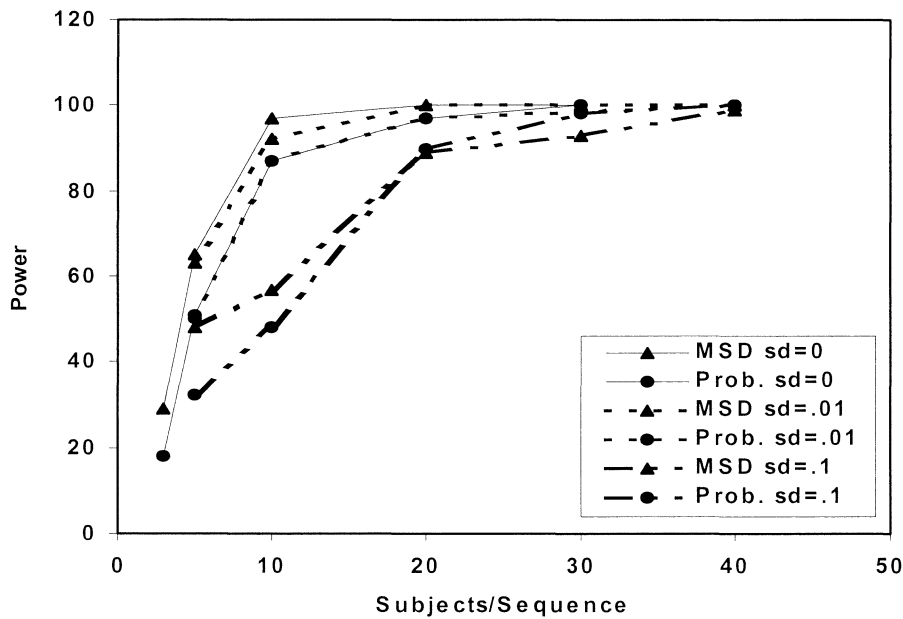


(2b)

Figure 2. The effect of inter-subject variance on population (2a) and individual (2b) bioequivalence assessment.



(3a)



(3b)

Figure 3. The effect of subject-by-formulation interaction on population (3a) and individual (2b) bioequivalence assessment.