6-2009

Statistical Procedures for Bioequivalence Analysis

Srinand Ponnathapura Nandakumar
Western Michigan University

STATISTICAL PROCEDURES FOR BIOEQUIVALENCE ANALYSIS

by

Srinand Ponnathapura Nandakumar

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor : Joseph W. McKean, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2009

STATISTICAL PROCEDURES FOR BIOEQUIVALENCE ANALYSIS

Srinand Ponnathapura Nandakumar, Ph.D.

Western Michigan University, 2009

Applicants submitting a new drug application (NDA) or new animal drug application (NADA) under the Federal Food, Drug, and Cosmetic Act (FDC Act) are required to document bioavailability (BA). A sponsor of an abbreviated new drug application (ANDA) or abbreviated new animal drug application (ANADA) must document first pharmaceutical equivalence and then bioequivalence (BE) to be deemed therapeutically equivalent to a reference listed drug (RLD). The Average (ABE), Population (PBE) and Individual (IBE) bioequivalence have been used to establish the equivalence in the pharmaco-kinetics of drugs.

The current procedure of PBE uses Cornish Fisher's (CF) expansion on small samples. Since area under the curve (AUC) and maximum dose (Cmax) are inherently skewed, a least squared (LS) normality based analysis is suspect. A bootstrap procedure is proposed which uses scale estimators. Since this bootstrap procedure works best for large samples, we propose a small sample analysis which uses robust scale estimators to compare least squares CF with Gini mean difference and inter quartile range.

Traditional ABE is univariate, two one-sided test which follows strict LS normality assumptions. We suggest small sample ABE utilizing AUC and Cmax in a multivariate setting with or without outliers using Componentwise rank method.

ACKNOWLEDGMENTS

I thank my family who has had a remarkable influence on my Ph.D. The financial and moral support by my parents and the constant urging by my sister have helped me complete my work in a planned timeline. I would also like to thank my uncle and aunt who enticed me with the idea of a Ph.D. I owe a lot to my fiancée, for she bore the brunt of my frustration and yet, encouraged me to complete.

I would like to express my deep and sincere gratitude to my supervisor, Dr. Joe McKean. His patience and guidance helped shape my research. I also thank him for sparing his personal time on my research. I am also deeply grateful to my supervisor, Dr. Gary Neidert for instigating this thesis. I was able to understand much about the industry's requirement from his guidance. I also thank my supervisor, Dr. Joshua Noranjo. His ideas set the direction of my research.

I owe my most sincere gratitude to the people at Innovative Analytics Inc. It was heartening to see their constant interest in my progress. During this work I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work.

Srinand Ponnathapura Nandakumar

TABLE OF CONTENTS

Table of Contents—continued

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Two pharmaceutical products are considered to be bioequivalent(BE) when their concentration versus time profiles, for the same molar dose, are so similar that they are unlikely to produce clinically relevant differences in therapeutic and/or adverse effects (Skelly *et al.*, 1995). A formal definition of bioequivalence by the FDA (2003a) is

> "Bioequivalence is defined as the the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study."

Applicants submitting a new drug application (NDA) or new animal drug application (NADA) under the provisions of section 505(b) in the Federal Food, Drug, and Cosmetic Act (FDC Act) are required to document bioavailability (BA). If approved, an NDA drug product may subsequently become a reference listed drug (RLD). Under section 505(j) of the Act, a sponsor of an abbreviated new drug application (ANDA) or abbreviated new animal drug application (ANADA) must first document pharmaceutical equivalence and then bioequivalence (BE) to be deemed therapeutically equivalent to an RLD. BE is documented by comparing the performance of the new or reformulated (test) and listed (reference) products (Niazi, 2007).

Pharmaceutical equivalents are drugs that have the same active ingredient in the same strength, dosage form, route of administration, have comparable labeling and meet compendia or other standards of identity, strength, quality, purity, and potency.

1.1 Metrics to characterize concentration-time profiles

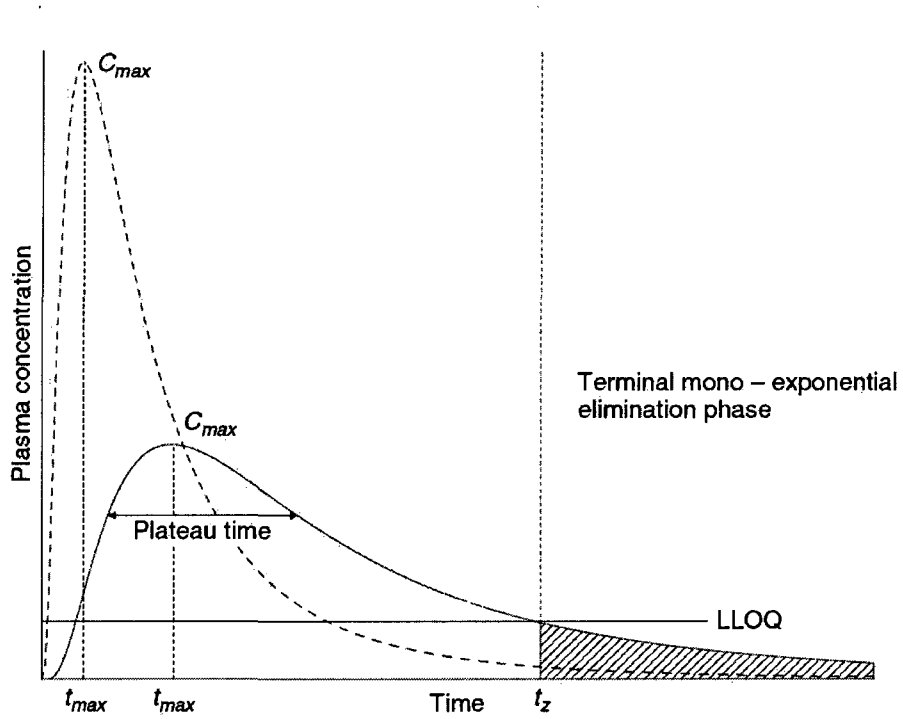


Figure 1: Typical concentration-time profile after a single dose

In figure 1 the dotted curve refers to an immediate release formulation and the solid curve to a prolonged release formulation. The metrics to characterize the concentration-time profiles are :

1. Area under the curve, AUC, is universally accepted as characteristic of the extent of drug absorption or total drug exposure. AUC is calculated using the trapezoidal rule.

2. Maximum drug absorbed, Cmax, is the peak plasma or the serum drug concentration which is an indirect metric for the rate of absorption.

3. Time of maximum concentration, Tmax, is the time to reach Cmax and is a direct metric for the rate of absorption.

The two most frequently used metrics are AUC and Cmax. The rationale (FDA, 2001) for log transformation of the metrics are:

1. Clinical Rationale: In a BE study, the ratio, rather than the difference between average parameter data from the test (T) and reference (R) formulations is of interest. With logarithmic transformation the FDA proposes a general linear model (glm) for inferences about the difference between the two means on the log scale.

2. Pharmacokinetic Rationale: A multiplicative model is postulated for pharmacokinetic measures AUC and Cmax. AUC is calculated as $\frac{FD}{CL}$ and Cmax as $\frac{FD}{V}e^{-k_e T_{max}}$. F is the fraction absorbed, D is the administered dose, and CL is the clearance of a given subject for the apparent volume of distribution V with a constant elimination rate k_e. Thus log transformations linearize AUC and Cmax.

1.2 Applications of bioequivalence studies

Hauschke et al. (2007) sight significant areas where bioequivalence studies are applied. These include

1. Applications for products containing new active substances.

2. Applications for products containing approved active substances.

   (a) Exemptions from bioequivalence studies in the case of oral immediate release forms (in vitro dissolution data as part of a bioequivalence waiver).

   (b) Post approval changes.

   (c) Dose proportionality of immediate release oral dosage forms.

   (d) Suprabioavailability (necessitates reformulation to a lower dosage strength, otherwise the suprabioavailable product may be considered as a new medicinal

product, the efficacy and safety of which have to be supported by clinical studies).

3. Applications for modified release forms essentially similar to a marketed modified release form.

   (a) The test formulation exhibits the claimed prolonged release characteristics of the reference.

   (b) The active drug substance is not released unexpectedly from the test formulation (dose dumping).

   (c) Performance of the test and reference formulation is equivalent after single dose and at a steady state.

   (d) The effect of food on the in vivo performance is comparable for both formulations when a single-dose study is conducted comparing equal doses of the test formulation with those of the reference formulation administered immediately after a predefined high fat meal.

In the statistical approaches to bioequivalence, the FDA (2003a) recognized three types of bioequivalence studies. They are:

- Average bioequivalence, ABE, used as a simple test of location equivalence. The mean differences are tested using Schuirmann's two one-sided procedure.

- Population bioequivalence, PBE, to compute the mean differences and variances for the BE criterion suggested by Chinchilli and Esinhart over a population group.

- Individual bioequivalence, IBE, to compare the mean differences and variances for the BE criterion on replicated crossover designs for an individual.

The order of testing these are ABE followed by either PBE or IBE. If ABE fails, then the remaining two are not tested. For the bioequivalence analysis, the interest lies in the ratio of the geometric means between the test(T) and the reference(R) drugs. This is stated in the FDA (2001) document that suggests the use of log-transformed data for the analysis.

1.3 Average bioequivalence (ABE)

The FDA (1992) suggests parametric (normal-theory) methods for the analysis of log transformed BE measures. For ABE, the general approach constructs a 90% confidence interval for the quantity $\mu_T - \mu_R$. If this confidence interval is contained in the interval $(-\theta_A, \theta_A)$, ABE is concluded.

1.3.1 Current procedure : Schuirmann's two one-sided t-tests

The ABE hypothesis tests are conducted with two one-sided t-tests. The hypothesis are:

$$\begin{aligned} H_{01} : \mu_T - \mu_R \leq \ln 0.8 \quad & or \quad H_{02} : \mu_T - \mu_R \geq \ln 1.25 \\ H_{A1} : \mu_T - \mu_R > \ln 0.8 \quad & \& \quad H_{A2} : \mu_T - \mu_R < \ln 1.25 \end{aligned} \tag{1.1}$$

A two period, two sequence, randomized double blind study is generally setup for testing ABE. We use Schuirmann's (1987) two one-sided t-tests and calculate the test statistics for each of the two nulls as $T_1 = \frac{\overline{Y_T} - \overline{Y_R} - log(0.80)}{\sqrt{\frac{MSE}{n_T + n_R - 2}}} > t_{1-\alpha,\nu}$ and $T_2 = \frac{\overline{Y_T} - \overline{Y_R} - log(1.25)}{\sqrt{\frac{MSE}{n_T + n_R - 2}}} < -t_{1-\alpha,\nu}$. If we reject either H_{01} or H_{02} then we reject H_0. By rejecting the null, we conclude ABE.

1.3.2 Issues with the current procedure

1. The FDA in its guidance for industry (2001) states

   "Sponsors and/or applicants are not encouraged to test for normality of

error distribution after log-transformation, nor should they use normality of error distribution as a reason for carrying out the statistical analysis on the original scale."

This suggests that there is considerable doubt regarding the distribution of the log transformed data. Schuirmann's (1987) t-test may fail if there were outliers or if the normality condition was not sufficiently satisfied.

2. Ghosh et al. (2007) state that histograms of the AUC and Cmax measures suggest non-normality in their distributions as well as the strong presence of outliers. Since AUC is calculated by extrapolating the concentration curve to infinity in time, this may lead to an outlier in extended release drugs. So, in studies involving small samples, Schuirmann's (1987) t-test may fail.

3. The adaptive procedure with Bonferroni confidence intervals used to address the multivariate setting of AUC and Cmax by Hui et al. (2001) has not been widely used. But the case of assessment of equivalence on multiple endpoints has been strongly suggested.

4. Multiple endpoints are suggested (Berger & Hsu, 1996), with pharmacokinetic parameters (Sunkara *et al.*, 2007) such as Tmax, $t_{1/2}$, MRT, etc (Yates *et al.*, 2002) and univariate Schuirmanns two one-sided tests are conducted on them.

Due to the above issues, we propose the use of the Componentwise rank method in analyzing ABE and address outliers with a distribution free approach on a multivariate setup of AUC and Cmax.

1.3.3 Proposed procedure : Componentwise rank method

The two treatment, two period crossover trial is routinely used to establish average bioequivalence of two drugs. We construct Schuirmann's (1987) two one-sided hypothesis (TOST) test in a multivariate setting as

$$H_{01}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \leq \ln 0.8 \cup H_{02}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \geq \ln 1.25$$

$$H_{A1}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} > \ln 0.8 \cap H_{A2}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} < \ln 1.25$$

Following a procedure outlined by Devan et al. (2008) we consider the difference between the Test and Reference drug responses for both AUC and Cmax there by eliminating the random factors. Following this approach, the hypothesis is bounded by (log(0.80),log(1.25)) and the rejection region is represented by a rectangle. We now calculate the robust estimates of location as the Hodges Lehmann estimate and the variance by Componentwise rank method.

The confidence region is an ellipse centered at the location estimates and the axes are determined by $\sqrt{\frac{\lambda_i}{n}}c = \sqrt{\lambda_i}\sqrt{\frac{p(n-1)F_{p,n-p}(\alpha)}{n(n-p)}}$ units along the eigen vectors e_i (Johnson & Wichern, 1992). If the ellipse is completely enclosed in the rejection region, we conclude PBE. The sensitivity analysis and the simulation results of the proposed procedure are discussed in Chapter 5.

1.4 Population bioequivalence (PBE)

As previously noted, current practice is to first test ABE. If ABE is concluded, PBE or IBE are tested. PBE is assessed to approve bioequivalence of a to-be-marketed formulation when a major formulation change has been made prior to approval of a new drug. It is

tested by administering the new drug to the patient who will be taking the drug formulation for the first time. Population bioequivalence will be considered only if average equivalence is approved. Chinchilli et al. (1996) have proposed a two sequence, four period cross-over design which the FDA has recommended for both PBE and IBE analysis.

1.4.1 Current procedure : LS Cornish Fisher's expansion (LSCF)

The FDA (2001), Hwang et al. (1996), Westlake (1988) have suggested the PBE hypothesis as

$$\begin{aligned} H_0 &: \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_0^2, \sigma_R^2)} \geq \theta_P \\ H_1 &: \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_0^2, \sigma_R^2)} < \theta_P \end{aligned} \tag{1.2}$$

where $\sigma_T^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ and $\sigma_R^2 = \sigma_{WR}^2 + \sigma_{BR}^2$ are the total variances of the test and the reference drugs. 'W' and 'B' refer to within and between subjects. The constants σ_0^2=0.04 and θ_P=1.744826 are fixed regulatory standards (FDA, 2001).

Setting $\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2)$, the hypothesis is rewritten as

$$\begin{aligned} H_0 &: \eta \geq 0 \\ H_1 &: \eta < 0 \end{aligned}$$

where $\widehat{\eta}$ is calculated using $\widehat{\eta} = \left(\widehat{\mu_T - \mu_R}\right)^2 + \widehat{\sigma_T^2} - \widehat{\sigma_R^2} - \theta_P max(\widehat{\sigma_R^2}, 0.04)$. The upper confidence interval of the linear combination of means and variances(η) is given by Cornish-Fisher's(CF) expansion. CF (Cornish & Fisher, 1938) is a procedure of combining sample quantiles for an upper limit approximate confidence interval. If $\widehat{\eta_{95\%}} \geq 0$, then we fail to reject H_0. When we reject H_0, PBE is concluded.

1.4.2 Issues with the present LSCF procedure

Ghosh et al. (2007) state that histograms of the AUC and Cmax measures suggest non normality of their distributions as well as the strong presence of outliers. The bootstrap procedure was initially suggested but was immediately dropped due to the complexity and the rigor involved in such analysis (Schall & Luus, 1993). Cornish-Fisher's expansion in Hyslop et al.(2000) was then proposed as the method of moments (MM) procedure.
The FDA (2001) notes that

> "One consequence of Cornish-Fisher(MM) expansion is that the estimator of $\widehat{\sigma_D^2}$ (the difference in within variances for IBE) is unbiased but could be negative."

The forced non negativity has the effect of making the estimate positively biased and introduces a small amount of conservatism to the confidence bound. In Lee et al. (2004),

> "A key condition assumed in all previously published works on modified large sample(MLS) is that the estimated variance components are independent. In some applications, however, variance component estimators are dependent. This occurs, in particular, when the study design is a crossover design, which is chosen by the FDA for bioequivalence studies."

The FDA (2003a) and the EC-GCP (2001) proposed the use of the non-parametric procedure of univariate Wilcoxon tests as a replacement to t-tests. Thus, alternative procedures to least squared Cornish Fisher's (LSCF) seem necessary to handle these issues. We, therefore propose two robust procedures that better handle outliers. Since we were not able to obtain consistent covariance structure with small samples, we separate the PBE analysis into large sample and small sample procedures.

1.4.3 Proposed robust bootstraps for large sample PBE

We decided to investigate PBE using robust bootstraps. Large sample PBE analysis worked best with samples of size sixty and above. This procedure involved the estimation of the upper confidence limit, $\widehat{\eta}_{95\%}$, using the median and five different variance estimates : Gini's mean difference (Gini), median absolute deviation (MAD), inter quartile range (IQR), median of absolute differences (S_n) and the k^{th} order statistic of the pairwise differences (Q_n).

Using the FDA (2001) proposed design, a two sequence, four period cross-over study was considered. Details of the bootstrap procedure are described in Chapter 3. For the variance, Gini, MAD, IQR, S_n and Q_n were used and η was estimated for each of the five cases as $\hat{\eta} = \widehat{\delta}^2 + \widehat{\sigma_T^2} - \widehat{\sigma_R^2} - 1.744826 \max\left(\widehat{\sigma_R^2}, 0.04\right)$ where $\widehat{\sigma}$ and $\widehat{\delta}$ were the scale and location analogue for LS. The 95^{th} highest η for each procedure gave $\widehat{\eta_{95\%}}$. The sensitivity analysis and the simulation results of the proposed procedure are discussed in Chapter 3.

1.4.4 Proposed procedure of small sample PBE

For samples of size twenty to thirty-six, we looked at PBE using Cornish Fisher's expansion. Continuing with the procedure similar to LSCF, we replaced the location estimates with medians and variance estimates from the IQR and the Gini procedures.

We estimated η by replacing the LS mean differences with median differences and the variances with the unbiased estimates of Gini and IQR. The upper 95% confidence interval of the Test and Reference location difference was estimated by Wilcoxon's rank-sum confidence interval. The sensitivity analysis and the simulation results of the proposed procedure are discussed in Chapter 4.

CHAPTER II

PRESENT PROCEDURE

2.1 Average bioequivalence (ABE)

The ABE hypothesis tests are conducted with two one-sided t-tests. The hypothesis is

$$\begin{aligned} H_{01} &: \mu_T - \mu_R \leq \ln 0.8 \quad or \quad H_{02} : \mu_T - \mu_R \geq \ln 1.25, \\ H_{A1} &: \mu_T - \mu_R > \ln 0.8 \quad \& \quad H_{A2} : \mu_T - \mu_R < \ln 1.25. \end{aligned} \tag{2.1}$$

This hypothesis is constructed in this manner because we are not just testing if the test and reference drugs are sufficiently close, but if they are "therapeutically equivalent" as well. Westlake (1976) stated that

> "The test of the hypothesis $H_0 : \mu_N = \mu_S$ is of scant interest since the practical problem is that of determining whether or not μ_N is sufficiently "therapeutically equivalent" to S. One approach, proposed by Westlake and Metzler is based on confidence intervals $\mu_S + C_2 \leq \mu_N \leq \mu_S + C_1$."

This hypothesis is vastly different from the two sided hypothesis as the two sided hypothesis merely tests the significant difference between the test and reference drugs. When the two sided analysis show a statistically significant difference between the test and reference formulation, it may be indicative of an important difference or of a trivially small difference (Westlake, 1979). The ABE hypothesis tests the practical equivalence (Berger & Hsu, 1996) of the two drugs. Further Westlake (1979) notes that the two sided hypothesis tests the wrong hypothesis. He stated that

"Since two formulations can hardly be expected to be identical, hypothesis testing of identity is simply directed at the wrong problem. The real question should really be: is the new formulation sufficiently similar to the standard in all important respects to suggest that it is therapeutically equivalent or is it sufficiently dissimilar to imply doubt as to therapeutic equivalence?"

We now recognize that we are not trying to prove that the test (T) and reference (R) drugs are equal. By estimating the difference between T and R and calculating the confidence interval of this difference (Westlake, 1979)., clinical judgment is exercised on arriving at the decision concerning bioequivalence. This is the logic behind using two one-sided hypothesis.

2.1.1 Use of confidence limits of (0.8,1.25) and log transformation

The modern concept of bioequivalence is based on a survey of physicians carried out by Westlake (1976) which concluded that a 20% difference (Westlake, 1979) in dose between two formulations would have no clinical significance for most drugs. Hence bioequivalence limits were set at 80% - 120%. But these limits are not symmetric since the pharamcokinetic (PK) parameters were tested after a log transformation. The FDA (2001) justifies the necessity to log transform AUC and Cmax with two reasons:

1. Clinical Rationale: The FDA Generic Drugs Advisory Committee recommended in 1991 that the primary comparison of interest in a BE study is the ratio, rather than the difference, between average parameter data from the T and R formulations. Using logarithmic transformation, the general linear statistical model employed in the analysis of BE data allows inferences about the difference between the two means on the log scale, which can then be re transformed into inferences about the ratio of the two averages on the original scale. Logarithmic transformation thus achieves a

general comparison based on the ratio rather than the differences.

2. Pharmacokinetic Rationale: Westlake observed that a multiplicative model is postulated for pharmacokinetic measures in BA and BE studies (i.e., AUC and Cmax). We calculate AUC and Cmax as $AUC = \frac{FD}{CL}$ and $Cmax = \frac{FD}{V}e^{-k_e T_{max}}$ where F is the fraction absorbed, D is the administered dose, and FD is the amount of drug absorbed and CL is the clearance of a given subject that is the product of the apparent volume(V) of distribution and the elimination rate(k_e).

Westlake (1976) proposed a procedure to resolve this issue of asymmetric confidence interval (CI). He set

$$C_2 \leq \mu_T - \mu_R \leq C_1,$$
$$k_2 SE - (\overline{X_T} - \overline{X_R}) \leq -(\mu_T - \mu_R) \leq k_1 SE - (\overline{X_T} - \overline{X_R}).$$

Since the decision of equivalence between T and R will be made on the basis of the largest of the absolute values of C_1 and C_2, the $max(|\log(0.8)|, |\log(1.20)|)$ is justified for the limits (Westlake, 1976). Conventionally, $k_1 + k_2 = 0$ but by choosing k_1 and k_2 such that $(k_1 + k_2)SE = 2(\overline{X_T} - \overline{X_R})$. We see that

$$k_1 SE - (\overline{X_T} - \overline{X_R}) = (\overline{X_T} - \overline{X_R}) - k_2 SE,$$
$$k_2 SE - (\overline{X_T} - \overline{X_R}) \leq -(\mu_T - \mu_R) \leq -[k_2 SE - (\overline{X_T} - \overline{X_R})]$$

to get symmetric CI about $\mu_T - \mu_R$. The hypothesis based on untransformed pharmacokinetic (PK) parameters AUC and Cmax is

$$H_{01} : \frac{\prod Test}{\prod Reference} \leq 0.8 \quad or \quad H_{02} : \frac{\prod Test}{\prod Reference} \geq 1.25.$$

Hence the bioequivalence limits of 80% - 125% or ± 0.2231436 on the natural log scale came to be in use.

2.1.2 Type I error: Level α of the test

Often a new test formulation has certain advantages over the reference formulation, such as fewer side effects or no pharmaco-kinetic interactions. For these cases, to prove overall superiority, it may be sufficient to show that for the primary endpoint, the test formulation is not relevantly inferior to the reference. Such studies are called non inferiority trials. This hypothesis can be expressed as

$$H_0 : \mu_T - \mu_R \leq \delta \quad vs. \quad H_1 : \mu_T - \mu_R > \delta$$

and tested with significance level α. It has been shown in Lehmann & Romano (2005) that the two one-sided hypothesis test at level α can be decomposed into two non-inferiority hypothesis tests each of level α. This is shown in figure 2. This can be seen by noting that the two one-sided hypothesis (H_0 and H_1) can be split into two hypotheses of the form

$$\begin{aligned} H_{01} : \mu_T - \mu_R \leq \ln 0.8 \quad or \quad H_{02} : \mu_T - \mu_R \geq \ln 1.25, \\ H_{11} : \mu_T - \mu_R > \ln 0.8 \quad \& \quad H_{12} : \mu_T - \mu_R < \ln 1.25. \end{aligned} \tag{2.2}$$

The null hypothesis H_{01} and its corresponding alternative, H_{11} is shown as a one side non-inferiority test in figure 2. Similarly we see that H_02 is a non-inferiority hypothesis as seen in section 1, Schrirmann's (1987) two one-sided t-test can be written as $H_0 = H_{01} \cup H_{02}$ vs $H_A = H_{A1} \cap H_{A2}$, where each are tested with a significance level α. Confidence sets

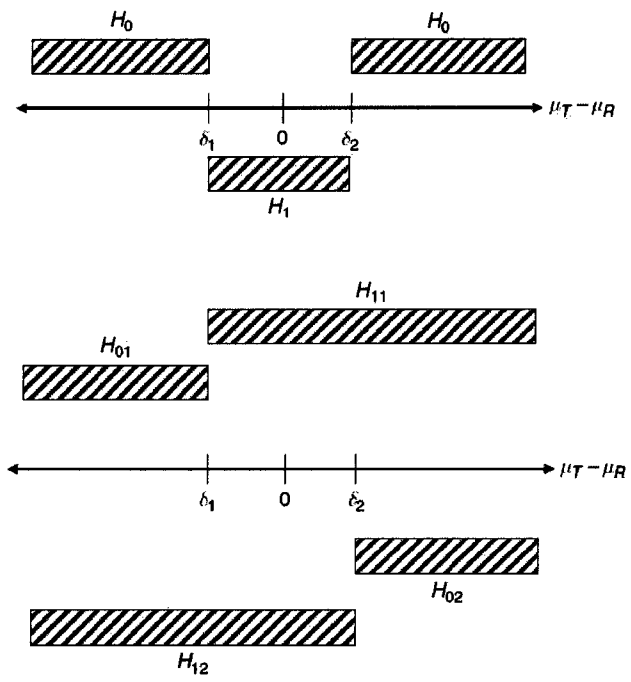


Figure 2: Decomposition of the two one-sided problem

for ratios (Von Luxburg & Franz, 2004) are

$$H_{01} : \mu_T - \mu_R \leq -\theta \quad or \quad H_{02} : \mu_T - \mu_R \geq \theta.$$

The rejection region for H_{01} and H_{02} can be written as

$$T_1 = \frac{\overline{Y_T} - \overline{Y_R} + \theta}{SE} > t_{1-\alpha,\nu} \quad and \quad T_2 = \frac{\overline{Y_T} - \overline{Y_R} - \theta}{SE} < -t_{1-\alpha,\nu}.$$

The probability of type I error is $P_{H_0}(RejectH_0)$. This probability is $P(Reject\ H_0|H_0 = True)$=$P_{H_0}(Reject\ H_0) = P_{H_0}(Reject\ H_{01} \cap Reject\ H_{02})$. The type I errors are

$$P_{H_{01}}(Reject\ H_{01}) = P_{H_{01}}(T_1 : \frac{\overline{Y_T} - \overline{Y_R} + \theta}{SE} > t_{1-\alpha,\nu}) = \alpha,$$
$$P_{H_{02}}(Reject\ H_{02}) = P_{H_{02}}(T_2 : \frac{\overline{Y_T} - \overline{Y_R} - \theta}{SE} < -t_{1-\alpha,\nu}) = \alpha.$$

Since both of the above two cases have monotonic power functions and the maximum are the boundary, the intersection of their rejection regions has asymptotic size bounded by α. In Lehmann et al. (2005) we see a proof of this generalized for any distribution. The FDA (2001) further stated

> "The general approach is to construct a 90% confidence interval for the quantity $\mu_T - \mu_R$ and to reach a conclusion of average BE if this confidence interval is contained in the interval $[-\theta_A, \theta_A]$. Due to the nature of normal-theory confidence intervals, this is equivalent to carrying out two one-sided tests of hypothesis at the 5% level of significance (Schuirmann1987)."

2.1.3 Power : 1-β of the test

Crossover designs are preferred by the FDA over parallel designs for the analysis of ABE. As noted by Chow & Wang (2001), this preference is due to

> "Intra subject variability could be eliminated if we could repeat the experiment many times (in practice, this just means the average of a large number of times) on the same subject under the same experimental condition. The reason is that intra subject variability tends to cancel out each other on average in a large scale."

Additionally the FDA (2001) explained the necessity to test the hypothesis under the assumption of the log-transformed data. It is usually desirable to sufficiently power the test with at least 80% power (i.e with type II error rate of $\beta = 0.2$). Now we look at the details of testing for PBE.

2.2 Population bioequivalence (PBE)

The FDA (FDA, 2001) noted the following as the preferred estimate for PBE or IBE:

$$\theta = E(Y_R - Y_T) - E(Y_R - \acute{Y_R}) \tag{2.3}$$

where Y_R, Y_T are the Reference and Test Formulation results respectively and $\acute{Y_R}$ is the replicated result. Replicated results are the subject's response to the same drug under the same dosage but at a different time period. A scaling reference downplays the amount of deviations in the Test and Reference estimates. In Hauschke (2007), the reason to use the replicated design is stated as

> "It is not possible to estimate the within-subject and between subject variances, each under test and reference formulation separately. This requires a replicate design where, in contrast to the standard crossover study, each study subject receives at least the reference formulation in two periods to enable the estimation of the corresponding within-subject variances. Of the various replicate designs that can be thought of, the FDA recommended in their 1997 and 1999 draft guidances (FDA,1997, 1999b) a four-period, two-sequence design, where the study subjects are randomly allocated to two treatment sequences."

2.2.1 Hypothesis test of PBE

The PBE hypothesis test is conducted with the following scaled moment-based aggregate criteria suggested by the FDA (2001)

$$H_0 : \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_0^2, \sigma_R^2)} \geq \frac{(\ln 1.25)^2 + 0.02}{0.04},$$
$$H_1 : \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_0^2, \sigma_R^2)} < \frac{(\ln 1.25)^2 + 0.02}{0.04} \tag{2.4}$$

where σ_0^2 is set by the FDA. The procedure is design specific and can be generalized. The FDA considered a completely randomized, two sequence, four period replicate design where each patient was administered to either a test or a reference drug formulation based on a randomization scheme.

2.2.2 Model design

The design is modeled as

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl} \tag{2.5}$$

where i=1,...,s indicates the number of sequences, j=1,...,n_i indicates the subjects within each sequence, k=R,T indicates the treatments, l=1,...,p_{ik} indicate replicates on treatment k for subjects within sequence i.

The response is Y_{ijkl} for replicate l on treatment k for subject j in sequence i and γ_{ikl} is the fixed effect of replicate l on treatment k in sequence i. The random effect is δ_{ijk} for subject j in sequence i on treatment k and ϵ_{ijkl} is the random error. It is assumed that

ϵ_{ijkl} are mutually independent and iid with

$$\begin{pmatrix} \epsilon_{ijTl} \\ \epsilon_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{WithinT} & 0 \\ 0 & \sigma^2_{WithinR} \end{pmatrix}\right] \tag{2.6}$$

such that the errors are independently distributed. Also, the random effect δ_{ijk} is

$$\begin{pmatrix} \delta_{ijT} \\ \delta_{ijR} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{BetweenT} & \rho\sigma_{BetweenR}\sigma_{BetweenT} \\ \rho\sigma_{BetweenR}\sigma_{BetweenT} & \sigma^2_{BetweenR} \end{pmatrix}\right] \tag{2.7}$$

The leads to overall response Y_{ijkl} to be distributed as

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}\right] \tag{2.8}$$

In order to calculate the overall Test and Reference variance, we set

$$\begin{aligned} \sigma^2_T &= \sigma^2_{BT} + \sigma^2_{WT}, \\ \sigma^2_R &= \sigma^2_{BR} + \sigma^2_{WR}. \end{aligned} \tag{2.9}$$

For the following example, a two sequence, four period balanced design will be used. Set the first sequence of the formulation randomization as TRTR and the second sequence as RTRT.

Table 1: Two sequence, four period balanced design

Subject	Sequence	Period1	Period2	Period3	Period4
1	1	Y_{1jT1}	Y_{1jR1}	Y_{1jT2}	Y_{1jR2}
2	1	.	.	.	.
.	1	.	.	.	.
m+1	2	Y_{2jR1}	Y_{2jT1}	Y_{2jR2}	Y_{2jT2}
.	2	.	.	.	.
.	2	.	.	.	.
j	2	.	.	.	.

2.2.3 Steps in population bioequivalence analysis

The population bioequivalence estimator involves the calculation of Θ and comparing it to the maximum acceptable limit of θ_P. Θ is defined as

$$\Theta = \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_R^2, \sigma_0^2)} \tag{2.10}$$

where as seen previously by FDA convention, $\Theta \leq \theta_P$. The value of θ_P is set using the calculation $\theta_P = \frac{ln(1.25)^2+0.02}{0.04} = 1.744826$. Linearizing this equation, we get

$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - max(\sigma_R^2, \sigma_0^2) * \theta_P \leq 0. \tag{2.11}$$

The FDA guidance directs that PBE is attained if the upper confidence interval of $\eta_{95\%}$ is less than 0. Thus the following are the steps for the analysis of PBE:

1. Determine the differences in the averages of the replicates of Test and Reference. Define I_{ij} as

$$\begin{aligned} I_{1j} &= \frac{(Y_{1jT1} + Y_{1jT2})}{2} - \frac{(Y_{1jR1} + Y_{1jR2})}{2}, \\ I_{2j} &= \frac{(Y_{2jT1} + Y_{2jT2})}{2} - \frac{(Y_{2jR1} + Y_{2jR2})}{2}. \end{aligned} \tag{2.12}$$

for each of the sequences i=1, 2 and each subject j in sequence i.

Define U_{ijT} as the average of the replicates on Test and U_{ijR} as the average of the replicates on Reference. Calculate them as

$$U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}. \tag{2.13}$$

Here U_{1jT} and U_{2jT} are independent as the subjects differ in the two sequences.

Define V_{ijT} as the difference of replicates on test and V_{ijR} as the difference of replicates on reference drugs. Estimate V_{ijk} with

$$V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}},$$
$$V_{2jT} = \frac{(Y_{2jT1} - Y_{2jT2})}{\sqrt{2}}. \tag{2.14}$$

Here V_{1jT} and V_{2jT} are again independent as the subjects differ in the two sequences.

2. Calculate the mean and the variances of I_{ij}, U_{ijk} and V_{ijk} respectively by sequence using equation (2.8).

$$E(I_{1j}) = \frac{(\mu_T + \mu_T)}{2} - \frac{(\mu_R + \mu_R)}{2}$$

I_{1j} and I_{2j} are independent as they are estimates from two different independent samples. The variance of U_{ijk} and V_{ijk} are

$$Var(U_{1jT}) = Var\left[\frac{(Y_{1jT1} + Y_{1jT2})}{2}\right]$$
$$Var(V_{1jT}) = Var\left[\frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}}\right]$$

Without loss of generality, we set Σ_1 and Σ_2. Thus,

$$\begin{pmatrix} Y_{1jT1} \\ Y_{1jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_1 \\ \Sigma_1 & \sigma_T^2 \end{pmatrix}\right]. \tag{2.15}$$

Thus, we see that $Var(U_{1jT}) = \frac{\sigma_T^2+\Sigma_1}{2}$ and $Var(V_{1jT}) = \sigma_T^2 - \Sigma_1$. Also,

$$\begin{pmatrix} Y_{2jT1} \\ Y_{2jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_2 \\ \Sigma_2 & \sigma_T^2 \end{pmatrix}\right] \tag{2.16}$$

3. Identify the estimates for the variance using the aggregate measures for the two sequences as

$$\sigma_{UT}^2 = \frac{1}{2}(\sigma_{UT_{seq1}}^2 + \sigma_{UT_{seq2}}^2)$$
$$\sigma_{VT}^2 = \frac{1}{2}(\sigma_{VT_{seq1}}^2 + \sigma_{VT_{seq2}}^2)$$

From σ_{UT}^2 and σ_{VT}^2, we can see that

$$\sigma_{UT}^2 = \frac{1}{2}(\frac{\sigma_T^2+\Sigma_1}{2} + \frac{\sigma_T^2+\Sigma_2}{2}) = \frac{1}{2}(\sigma_T^2 + \frac{\Sigma_1+\Sigma_2}{2})$$
$$\sigma_{VT}^2 = \frac{1}{2}(\sigma_T^2 - \Sigma_1 + \sigma_T^2 - \Sigma_2) = \sigma_T^2 - \frac{\Sigma_1+\Sigma_2}{2} \tag{2.17}$$

We now have variance estimators using equation (2.17) and location difference using $\widehat{(\mu_T - \mu_R)} = \frac{(\overline{I_1}+\overline{I_2})}{2}$.

4. Obtain η and calculate the upper confidence interval for η using Cornish-Fisher's

expansion. We estimate $\widehat{\eta}$ as

$$\widehat{\eta} = \left(\frac{\overline{I_1} + \overline{I_2}}{2}\right)^2 + \left(\widehat{\sigma^2_{UT}} + \frac{\widehat{\sigma^2_{VT}}}{2}\right) - \left(\widehat{\sigma^2_{UR}} + \frac{\widehat{\sigma^2_{VR}}}{2}\right) - \max\left(\widehat{\sigma^2_R}, \sigma^2_0\right)\theta_P < 0$$

Refer to Chapter 3 for the calculations of Cornish Fisher's expansion. The upper 95^{th} confidence interval is calculated by

$$H = \sum P_q + \left(\sum B_q\right)^{\frac{1}{2}}.$$

If H < 0 then Population bioequivalence is concluded.

2.2.4 Cornish Fisher's expansion

The principle behind the Cornish-Fisher's expansion is that close to exact confidence intervals for a parameter are more accurate when higher-order approximation in the expansions for the quantiles are used. The previous section described the need to find the upper confidence interval of η to conclude PBE.

> "For constructing asymptotically correct confidence intervals for a parameter on the basis of an asymptotically normal statistic, the first-order approximation to the quantiles of the statistic comes from using the central limit theorem. The higher-order expansions for the quantiles produce more accurate approximations than does just the normal quantile. (DasGupta, 2008)"

The Cornish-Fisher expansions are higher-order expansions for quantiles and are essentially obtained from recursively inverted Edgeworth expansions, starting with the normal quantile as the initial approximation. In (Cornish & Fisher, 1938), we first see that the density functions are based on the cumulants of a distribution. If we are interested in the percentiles of the sum of two random variables Z=X+Y, from (Cornish & Fisher, 1938),

one gets

$$P[Z \leq \mu_X + \mu_Y + R_\beta(\sigma_X^2 + \sigma_Y^2)^{\frac{1}{2}} + (R_\beta^2 - 1)(\mu_{3X} + \mu_{3Y})/6(\sigma_X^2 + \sigma_Y^2) + ...] = \beta$$

where μ_X, μ_Y, σ_X^2, σ_Y^2, μ_{3X}, μ_{3Y} are the first, second and third central moments respectively of X and Y and β is the desired exact percentile. The Cornish-Fisher expansion is based on the principle of power series

$$M(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx$$

$$M(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \int_{-\infty}^{\infty} x^r f(x)dx = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \mu_r^{'} \tag{2.18}$$

where the function is differentiable and continuous at all points. Further $\mu_r^{'}$ is the r^{th} moment of the distribution of x about the origin. In our situation however, we have more than two random variables which leads to the approximation (Howe, 1974)

$$P\left\{\sum_{i=1}^{p} X_i \leq \sum_{i=1}^{p} \mu_i + \left[\sum_{i=1}^{p} (X_{i\beta} - \mu_i)^2\right]^{\frac{1}{2}}\right\} \cong \beta \tag{2.19}$$

where the X_i are distributed independently with means μ_i and β percentile of $X_{i\beta}$. Now $X_{i\beta}$ can be derived from the Cornish-Fisher's expansion of the cumulants and estimating the constants such that β is approached as close as possible. Since we need to find the upper 95% probability of capturing η, the FDA (2001) suggested the use of $H = \sum P_q + (\sum B_q)^{\frac{1}{2}}$.

CHAPTER III

BOOTSTRAP POPULATION BIOEQUIVALENCE

Analysis of population bioequivalence focuses on estimation of the mean difference and the total variance of the log transformed BA measures from the two drug formulations. Unbiased estimators using the method of moments (Chinchilli & Esinhart, 1996) estimate these parameters.

Following the estimation of the mean difference and the variances, a 95% upper confidence bound for a linearized form of the population BE criterion is obtained. Population BE is established for a log-transformed BA measure if the upper 95% confidence bound for this linearized criterion is less than or equal to zero (FDA, 2001).

3.1 Distributional assumptions of metrics in BE trials

Before performing a statistical analysis in BE trials, AUC and Cmax are generally log transformed. The three most commonly cited reasons for log transforming AUC and Cmax are

- AUC is non-negative
- Distribution of AUC is highly skewed
- PK models are multiplicative

The drug concentration at each time point is a function of many random processes. They are absorption, distribution, metabolism and elimination that act proportionally to the amount of the drug present in the body. Thus the resulting distribution is log normal (Midha & Gavalas, 1993).

3.2 Design

In a BE trial, the test (T) and the reference (R) drug formulations are administered to healthy volunteers and the drug concentrations are measured over time. Frequently cross-over designs as shown in table 2 are employed, although parallel group designs are used as well. Cross-over designs are generally preferred because of their ability to compare the test

Table 2: Two sequence, four period balanced design

Subject	Sequence	Period1	Period2	Period3	Period4
1	1	Y_{1jT1}	Y_{1jR1}	Y_{1jT2}	Y_{1jR2}
.	1	.	.	.	.
.	1	.	.	.	.
m+1	2	Y_{2jR1}	Y_{2jT1}	Y_{2jR2}	Y_{2jT2}
.	2	.	.	.	.
.	2	.	.	.	.
j	2	.	.	.	.

and reference formulations within a subject. We focus on BE trials using a (2x4) cross-over design i.e a two sequence, four period replicated balanced design as suggested by the FDA (2001).

The first sequence has a test, reference, test and reference (TRTR) schedule while the second sequence has a reference, test, reference and test (RTRT) schedule. The response is Y_{ijkl} for replicate l on treatment k and subject j in sequence i. The fixed effect is γ_{ikl} and the random effect is δ_{ijk} with random error ϵ_{ijkl}. The design is as follows

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl} \tag{3.1}$$

where i=1,...,s indicates the number of sequences, j=1,...,n_i indicates the subjects within each sequence, k=R,T indicates the treatments and l=1,...,p_{ik} the replicates on treatment k.

ϵ_{ijkl} and δ_{ijk} are mutually independent and distributed as shown below:

$$\begin{pmatrix} \epsilon_{ijTl} \\ \epsilon_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{WithinT} & 0 \\ 0 & \sigma^2_{WithinR} \end{pmatrix}\right], \tag{3.2}$$

$$\begin{pmatrix} \delta_{ijT} \\ \delta_{ijR} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BR} \end{pmatrix}\right]. \tag{3.3}$$

From the design, we get a bivariate response of the form

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}\right] \tag{3.4}$$

The next section introduces the hypothesis to test PBE.

3.2.1 Hypothesis

The proposed null and alternative hypothesis based on the FDA regulations (2001) are

$$\begin{aligned} H_0 &: \frac{(\mu_T - \mu_R)^2 + \sigma^2_T - \sigma^2_R}{max(\sigma^2_0, \sigma^2_R)} \geq \theta_P \\ H_1 &: \frac{(\mu_T - \mu_R)^2 + \sigma^2_T - \sigma^2_R}{max(\sigma^2_0, \sigma^2_R)} < \theta_P \end{aligned} \tag{3.5}$$

where $\sigma^2_T = \sigma^2_{WT} + \sigma^2_{BT}$ and $\sigma^2_R = \sigma^2_{WR} + \sigma^2_{BR}$ are the total variances of the test and the reference drugs. The constants σ^2_0 and θ_P are fixed regulatory standards.

As seen above, the FDA guidance currently adopts an aggregate approach, using an aggregated test statistic for evaluating both means and variance components simultaneously. In contrast, several disaggregate approaches have been suggested where tests for

each component are performed separately. For example, Liu and Chow (1996) proposed a disaggregate approach for evaluating IBE where three components (intra subject variability, subject-by-formulation interaction, and average) are separately tested multiple times with intersection-union tests. However, as the dimensions (p) of tests increases, the power of the $(1-2\alpha)$ confidence set (Leena Choi, 2008) based approach could decrease sharply for dimensions greater than one as shown in Hwang (1996).

The aggregated test statistic is linearized as follows:

$$H_0 : (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2) \geq 0,$$
$$H_1 : (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2) < 0. \tag{3.6}$$

Here, $\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2)$ and the null hypothesis reduces to a one sided problem defined by a linear combination. The FDA fixed 0.02 as the maximum difference for the variance under the test and reference formulations. Usually θ = log 1.25 = -log 0.80 = 0.223. These values (FDA, 2001) originated from the notion that the ratio of the population means in the original scale (the mean of the test is 80 - 125% of that of the reference) are considered to be sufficiently close for drugs having an average therapeutic window. For PBE, the FDA sets θ_P = 1.744826 and σ_0^2 = 0.04. The linearized hypothesis is of the form

$$H_0 : \eta \geq 0,$$
$$H_1 : H_0 : \eta < 0.$$

If the null is rejected, population bioequivalence (the two drugs are similar across population groups) is inferred. Otherwise, the two drugs are significantly different across the populations. The next section describes the present procedure of testing PBE hypothesis.

3.2.2 Least squares Cornish Fisher's procedure (LSCF)

The present procedure tests PBE using Cornish Fisher's (CF) (1938) expansion. In LSCF, η is calculated as $\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2)$. The procedures in estimating μ_i and σ_i^2 are described below. If the upper confidence interval $\eta_{95\%}$ is less than zero, population bioequivalence is concluded.

Following are the steps in computing the least squares Cornish Fisher's (LSCF) expansion:

1. From table 2, the response Y_{ijkl} is distributed as

$$\begin{pmatrix} Y_{ijRl} \\ Y_{ijTl} \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 + \sigma_{WR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 + \sigma_{WT}^2 \end{pmatrix} \right]$$

   where each subject j has two observations for one of the two treatments. Each subject belongs to only one sequence. The data has 'N' subjects partitioned into two sequences with $\frac{N}{2}$ subjects in each sequence. In this example, a balanced design is used. The variances σ_B^2 and σ_W^2 are the between and within variances. For the first sequence the patients have a TRTR schedule and the second sequence subjects have an RTRT schedule.

2. Define I as the difference in test and reference drug replicate averages. Compute this difference I_{ij} as

$$I_{1j} = \frac{(Y_{1jT1} + Y_{1jT2})}{2} - \frac{(Y_{1jR1} + Y_{1jR2})}{2},$$
$$I_{2j} = \frac{(Y_{2jT1} + Y_{2jT2})}{2} - \frac{(Y_{2jR1} + Y_{2jR2})}{2}$$

   for each of the sequence i=1, 2.

3. Calculate U_{ijT} as the average of the test drug replicates and U_{ijR} as the average of the reference drug replicates. This average is

$$U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}.$$

U_{1jT} and U_{2jT} are independent as they are estimates from two different independent samples

4. Define V_{ijT} as the difference of the replicates of the test and V_{ijR} as the difference of the replicates of the reference drug. V_{ijk} is calculated as

$$V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}},$$
$$V_{1jR} = \frac{(Y_{1jR1} - Y_{1jR2})}{\sqrt{2}}.$$

5. Calculate the variance of the variables U_{ijk}, V_{ijk} for each of the two sequences. Estimate the variance of test drug σ_T^2 as $\sigma_{BT}^2 + \sigma_{WT}^2$ and reference drug σ_R^2 as $\sigma_{BR}^2 + \sigma_{WR}^2$. For the first sequence, the variance is estimated with

$$Var(U_{1jT}) = \frac{Var(Y_{1jT1}) + Var(Y_{1jT2}) + 2Cov(Y_{1jT1}, Y_{1jT2})}{4},$$
$$Var(V_{1jT}) = \frac{Var(Y_{1jT1}) + Var(Y_{1jT2}) - 2Cov(Y_{1jT1}, Y_{1jT2})}{2}.$$

Without loss of generality, set the covariance (Σ_1) for the first sequence and the two test drug periods. The resulting distribution of the test drug in the first sequence is

$$\begin{pmatrix} Y_{1jT1} \\ Y_{1jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_1 \\ \Sigma_1 & \sigma_T^2 \end{pmatrix} \right]. \tag{3.7}$$

Similarly, the distribution of the test drug in the second sequence is

$$\begin{pmatrix} Y_{2jT1} \\ Y_{2jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_2 \\ \Sigma_2 & \sigma_T^2 \end{pmatrix}\right]. \tag{3.8}$$

It can be proved that σ_T^2 is a linear combination of the variances σ_{UT}^2 and σ_{VT}^2. To prove $\sigma_T^2 = \sigma_{UT}^2 + \frac{\sigma_{VT}^2}{2}$, consider the following

$$\begin{aligned}
\sigma_T^2 &= \overline{\sigma_{UT}^2} + \overline{\frac{\sigma_{VT}^2}{2}} \\
\sigma_T^2 &= \tfrac{1}{2}\left(\sigma_{UT_{seq1}}^2 + \sigma_{UT_{seq2}}^2\right) + \tfrac{1}{2}\left(\tfrac{1}{2}\left\{\sigma_{VT_{seq1}}^2 + \sigma_{VT_{seq2}}^2\right\}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left(\sigma_{UT_{seq1}}^2 + \frac{\sigma_{VT_{seq1}}^2}{2}\right) + \tfrac{1}{2}\left(\sigma_{UT_{seq2}}^2 + \frac{\sigma_{VT_{seq2}}^2}{2}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left(Var(U_{1jT}) + \frac{Var(V_{1jT})}{2}\right) + \tfrac{1}{2}\left(Var(U_{2jT}) + \frac{Var(V_{2jT})}{2}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left[\left(\tfrac{\sigma_T^2+\Sigma_1}{2}\right) + \left(\tfrac{\sigma_T^2-\Sigma_1}{2}\right)\right] + \tfrac{1}{2}\left[\left(\tfrac{\sigma_T^2+\Sigma_2}{2}\right) + \left(\tfrac{\sigma_T^2-\Sigma_2}{2}\right)\right].
\end{aligned}$$

By expanding the above equation, it is concluded that

$$\sigma_{UT}^2 + \frac{\sigma_{VT}^2}{2} = \sigma_T^2.$$

Similarly, for the reference drug, $\left(\sigma_{UR}^2 + \frac{\sigma_{VR}^2}{2}\right) = \sigma_R^2$.

6. The expected values of the difference for the test and reference drugs from the two sequences across the four periods or two replicates using equation 3.7 are

$$\begin{aligned}
E\left(I_{1j}\right) &= E\left[\frac{\left(Y_{1jT1} + Y_{1jT2}\right)}{2} - \frac{\left(Y_{1jR1} + Y_{1jR2}\right)}{2}\right] = \frac{2\mu_T - 2\mu_R}{2}, \\
E\left(I_{2j}\right) &= E\left[\frac{\left(Y_{2jT1} + Y_{2jT2}\right)}{2} - \frac{\left(Y_{2jR1} + Y_{2jR2}\right)}{2}\right] = \frac{2\mu_T - 2\mu_R}{2}.
\end{aligned}$$

Thus from the average of the two sequences, $\frac{E(I_{1j})+E(I_{2j})}{2} = \mu_T - \mu_R$.

7. Estimate the aggregate statistic η using the linear combinations of means and variances as

$$\widehat{\eta} = \left(\widehat{\mu_T - \mu_R}\right)^2 + \widehat{\sigma_T^2} - (1 + \theta_P)\, max(\widehat{\sigma_R^2}, 0.04).$$

Calculate the upper confidence interval of $\widehat{\eta}$ using the Cornish Fisher's expansion. To illustrate CF's expansion, consider H as the upper bound in the equation

$$H = \sum P_q + \left(\sum B_q\right)^{\frac{1}{2}}$$

where P_q represents the point estimates i.e mean, variances and B_q represents the upper bound of these point estimates (95%).

8. Table 3 outlines the various point estimates and their respective upper bounds.

Table 3: Point estimates and their distributions

P_q=Point Estimate	C=Confidence Bound	B_q=Upper α limit
P_1=$(\widehat{\mu_T - \mu_R})^2$	$U_1 = \left(\widehat{P_1} + t_{1-\alpha,N-s}\left(\sum_{i=1}^{s} n_i^{-1} s_I^2\right)^{\frac{1}{2}}\right)^2$	B_1=$\left(U_1 - \widehat{P_1}\right)^2$
P_2=$\widehat{\sigma_{Uk}^2}$	$U_2 = \widehat{\sigma_{Uk}^2}\frac{(N-2)}{\chi^2_{\alpha,N-2}}$	B_2=$\left(U_2 - \widehat{\sigma_{Uk}^2}\right)^2$
P_3=$\frac{1}{2}\widehat{\sigma_{Vk}^2}$	$U_3 = \frac{1}{2}\widehat{\sigma_{Vk}^2}\frac{(N-2)}{\chi^2_{\alpha,N-2}}$	B_3=$\left(U_3 - \frac{1}{2}\widehat{\sigma_{Vk}^2}\right)^2$

Thus, calculate the upper CI of $\widehat{\eta}$ using Cornish Fisher's expansion. The upper 95%

confidence value of $\widehat{\eta}$ is calculated as

$$\hat{\eta} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_T^2} - (1+\theta_P)\max\left(\widehat{\sigma_R^2}, \sigma_0^2\right),$$

$$\hat{\eta} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_{UT}^2} + \tfrac{1}{2}\widehat{\sigma_{VT}^2} - (1+\theta)\max\left(\widehat{\sigma_{UR}^2} + \tfrac{1}{2}\widehat{\sigma_{VR}^2}, \sigma_0^2\right),$$

$$\widehat{\eta}_{1-\alpha} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_{UT}^2} + \tfrac{1}{2}\widehat{\sigma_{VT}^2} - (1+\theta)\max\left(\widehat{\sigma_{UR}^2} + \tfrac{1}{2}\widehat{\sigma_{VR}^2}, \sigma_0^2\right)$$
$$+\Bigg\{\left[\left(|\widehat{\mu_T - \mu_R}| + t_{\alpha,N-2}\sqrt{\tfrac{s_j^2}{n_1+n_2-2}}\right)^2 - \widehat{(\mu_T - \mu_R)}^2\right]^2 + \left[\tfrac{(N-2)\widehat{\sigma_{UT}^2}}{\chi^2_{\alpha,N-2}} - \widehat{\sigma_{UT}^2}\right]^2$$
$$+\left[\tfrac{1}{2}\tfrac{(N-2)\widehat{\sigma_{VT}^2}}{\chi^2_{\alpha,N-2}} - \tfrac{1}{2}\widehat{\sigma_{VT}^2}\right]^2 + \left[\tfrac{-(1+\theta_P)(N-2)\widehat{\sigma_{UR}^2}}{\chi^2_{\alpha,N-2}} + (1+\theta_P)\widehat{\sigma_{UR}^2}\right]^2$$
$$+\left[\tfrac{-(1+\theta_P)(N-2)\frac{1}{2}\widehat{\sigma_{VR}^2}}{\chi^2_{\alpha,N-2}} + (1+\theta_P)\tfrac{1}{2}\widehat{\sigma_{VR}^2}\right]^2\Bigg\}^{1/2}.$$

Once $\widehat{\eta}_{95\%}$ is computed, conclude PBE if $\widehat{\eta}_{95\%}$ is less than zero. When H_0 is rejected, PBE is concluded. The following section proposes the robust bootstrap procedure as an alternative to the LSCF procedure.

3.2.3 Robust bootstrap procedure

The robust analog of least squares Cornish Fisher (LSCF) involves calculation of the robust bootstrap estimates of $\widehat{\eta_{95\%}}$. Separate the data from table 2 based on the two sequences and conduct bootstrap (Efron & Tibshirani, 1993) analysis. This is done to maintain the covariance structure.

Use median as the robust location estimate. For the variance estimates, use MAD, S_n, Q_n (Ola Hssjer & Croux, 1996), IQR and Gini. Calculate $\widehat{\eta}$ and also the upper 95^{th} percentile of $\widehat{\eta}$ which is the $95^{th}\widehat{\eta}$ of the bootstrapped data sorted in an ascending order. Steps in robust population bioequivalence using bootstraps are as follows:

1. Start with the data as in table 2. Each subject j has two observations for one of the two treatments. The N subjects are partitioned into two sequences with $\frac{N}{2}$ subjects

in each sequence i.e a balanced design as seen in the table. The variances σ_B^2 and σ_W^2 are the between and the within variances. From this setup, for the first sequence, we have a TRTR schedule and for the second sequence, an RTRT schedule. The response is distributed as

$$\begin{pmatrix} Y_{ijRl} \\ Y_{ijTl} \end{pmatrix} \sim G\left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 + \sigma_{WR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 + \sigma_{WT}^2 \end{pmatrix}\right].$$

2. For each of the two sequences, generate a simple random sample with replacement of the response Y_{ijkl}. If $Y_{ijk} = (y_{i1k}, ...y_{iNk})$ then generate $Y_{1jk} = (y_{1jk}, ...y_{1jk})$ and $Y_{2j'k} = (y_{2j'k}, ...y_{2j'k})$ for each of the two sequences. Bootstrap each sequence separately as it maintains the consistent covariance structure.

   This gives 2M datasets each of which have $\frac{N}{2}$ subjects and only one sequence with four periods. Combine Y_{1jk} and $Y_{2j'k}$ to obtain M datasets and estimate M $\eta's$.

3. Define I_{ijk} as the averages of the replicates of the test and reference drugs. Calculate I_{ijk} as

$$I_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$I_{1jR} = \frac{(Y_{1jR1} + Y_{1jR2})}{2},$$
$$I_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2},$$
$$I_{2jR} = \frac{(Y_{2jR1} + Y_{2jR2})}{2}$$

   for each of the sequences i=1, 2. Using I_{ijk}, the location estimate for the test and the reference drugs can be estimated.

4. Define U_{ijT} as the average of the replicates for the test drug and U_{ijR} as the average

of the replicates for the reference drug. Compute U_{ijk} as

$$U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}.$$

Here U_{1jT} and U_{2jT} are independent as they are estimates from two different independent samples.

5. Define V_{ijT} as the difference of replicates of the test and V_{ijR} as the difference of replicates of the reference drugs. Calculate V_{ijk} with

$$V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}}$$
$$V_{2jR} = \frac{(Y_{2jR1} - Y_{2jR2})}{\sqrt{2}}$$

6. Obtain the robust difference in location between the test and reference drugs as the median of difference of I_{1Tj}, I_{1Rj}, I_{2Tj} and I_{2Rj} for each of the M datasets. Estimate the location difference as

$$\xi_T - \xi_R = \frac{\text{Median}_{I_{1Tj}} + \text{Median}_{I_{2Tj}}}{2} - \frac{\text{Median}_{I_{1Rj}} + \text{Median}_{I_{2Rj}}}{2}.$$

7. Without loss of generality, from LSCF, the variance is estimated as $\left(\widehat{\sigma^2_{UT}} + \frac{\widehat{\sigma^2_{VT}}}{2}\right) = \widehat{\sigma^2_T}$ and for the reference drug the variance is estimated by $\left(\widehat{\sigma^2_{UR}} + \frac{\widehat{\sigma^2_{VR}}}{2}\right) = \widehat{\sigma^2_R}$. If $\widehat{Var_T}$ is the robust expression of $\widehat{\sigma^2_T}$ then

$$\left(Var_{UT} + \frac{Var_{VT}}{2}\right) = \widehat{Var_T}$$
$$\left(Var_{UR} + \frac{Var_{VR}}{2}\right) = \widehat{Var_R}$$

represents the robust estimates of spread. From the asymptotic theory, estimate spread using different spread estimators.

8. These estimators of spread are MAD, Gini, IQR, S_n and Q_n. A parallel can be drawn between the LSCF and the robust procedure based on asymptotic theory as in table 4. Here $\tilde{\xi}_1^*$ represents the median of I_1^* for the M bootstrap samples. MAD which is

Table 4: LSCF and robust location, scale of each bootstrap sample

Parameter	Least Squares	R method	Gini method
$\widehat{\left(\frac{\mu_T-\mu_R}{2}\right)^2}$	$\left(\frac{\overline{I_1}+\overline{I_2}}{2}\right)^2$	$\left(\frac{\tilde{\xi}_1^*+\tilde{\xi}_2^*}{2}\right)^2$	$\left(\frac{\tilde{\xi}_1^*+\tilde{\xi}_2^*}{2}\right)^2$
$\widehat{\sigma_{Uk}^2}$	$\widehat{\sigma_{Uk}^2}$	$(1.482 \cdot MAD_{Uk}*)^2$	$\left(G_{Uk} * \frac{\sqrt{\pi}}{2}\right)^2$
$\widehat{\frac{1}{2}\sigma_{Vk}^2}$	$\frac{1}{2}\widehat{\sigma_{Vk}^2}$	$\frac{1}{2}(1.4826 \cdot MAD_{Vk}*)^2$	$\frac{1}{2}\left(G_{Vk} * \frac{\sqrt{\pi}}{2}\right)^2$

median absolute deviation is calculated as $MAD_x = \text{median}_i(|x_i - \text{median}_j(x_j)|)$. Gini's mean difference is calculated as $Gini = \sum_{i<j} |x_i - x_j| \Big/ \begin{pmatrix} n \\ 2 \end{pmatrix}$. For a normal distribution, $1.4826 \cdot MAD$ and $\frac{\sqrt{\pi}}{2}G$ are unbiased estimators of the standard deviation. MAD has low efficiency for normal distributions, and it may not always be appropriate for symmetric distributions.

The two statistics that Rousseeuw and Croux (1993) proposed as alternatives to MAD are S_n and Q_n. S_n is calculated with

$$S_n = 1.1926 \cdot \text{med}_i \left(\text{med}_j \left(|x_i - x_j|\right)\right)$$

where the outer median (taken over i) is the median of the n medians of $|x_i - x_j|$, j = 1, 2, ... , n. To reduce small-sample bias, $c_{sn}S_n$ is used to estimate σ where c_{sn} is the correction factor (1992b). The second statistic is Q_n (1992a) estimated as

$$Q_n = 2.219\{|x_i - x_j|\,; i < j\}_{(k)}$$

where

$$k = \begin{pmatrix} n \\ 2 \end{pmatrix}$$

and h = [n/2] + 1. In other words, Qn is 2.219 times the kth order statistic of the C_2^n distances between the data points. The bias-corrected statistic $c_{qn}Q_n$ is used to estimate σ,where c_{qn} is a correction factor (Rousseeuw & Croux, 1992c).

The interquartile range (IQR) is the difference between the upper and lower quartiles. For a normal population, IQR/1.34898 (DasGupta & Haff, 2006) is an unbiased estimator of the standard deviation.

9. Calculate $\widehat{\eta}$ for the M datasets by using the above estimators. Now pool the M datasets and estimate the upper 95% confidence interval of η by selecting the 95^{th} $\widehat{\eta}$ sorted in ascending order. With this step, $\widehat{\eta}$ and $\widehat{\eta_{95}}$ are estimated using each of the spreads MAD, Gini, IQR, S_n and Q_n.

Now, compare the proposed robust procedures to the LSCF's procedure. In order to find the procedure most resistant to outliers, run sensitivity analysis on an example shown below.

3.3 Analysis of an example

Apply the present and proposed procedures on a dataset. This dataset was procured from the FDA website (2003b) which was created on August 18, 2003 and updated on June 20, 2005. Introduction to the dataset used is as follows:

> "In reference to the Federal Register notice on "Preliminary Draft Guidance for Industry on In Vivo bioequivalence Studies Based on Population and Individual bioequivalence Approaches: Availability", vol. 62, No. 249, Dec. 30, 1997,

Table 5: Example to illustrate the PBE procedure

SUBJECT	PER	SEQ	TRT	AUC
1	1	RTTR	R	5.696
1	2	RTTR	T	5.445
1	3	RTTR	T	8.481
1	4	RTTR	R	6.774
.	.	.	.	.
.	.	.	.	.
104	1	TRRT	T	2.9
104	2	TRRT	R	4.05
104	3	TRRT	R	4.287
104	4	TRRT	T	2.85

the Food and Drug Administration is announcing the availability of data that were used by the Agency in support of the proposal and the detailed description of statistical methods for individual and population approaches."

The dataset used for the analysis is 'DRUG 3*'(including 3b - 3d used as an illustration) from the above source. It is a combination of the three datasets which are modified to fit the RTTR and TRRT schedule. This data is a two sequence, four period replicate design with 104 subjects who are randomized into one of the two sequences. The subjects in the first sequence start with a RTTR schedule (reference-test-test-reference) while sequence two have a TRRT schedule. There is a sufficient washout period between the test and reference drugs to avoid carryover effects. Table 5 illustrates this dataset. Re-order the data by transposing on the period.

The response is Y_{ijkl} for replicate l on treatment k for subject j in sequence i. The fixed effect is γ_{ikl} and the random effect is δ_{ijk} with random error ϵ_{ijkl}. The design used is

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl} \tag{3.9}$$

Table 6: Transformed two sequence, four period balanced design

Subject	Sequence	Period1	Period2	Period3	Period4
1	1	log(5.696)	log(5.445)	log(8.481)	log(6.774)
2	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
104	2	log(2.9)	log(4.05)	log(4.287)	log(2.85)

where i=1,2 indicates the number of sequences, j=1,...,104 indicates the subjects within each sequence, k=R,T indicates the treatments, l=1,2 indicate replicates on treatment k for subjects within sequence i. Due to the balanced design, there are 52 subjects in the first sequence and 52 subjects in the second sequence.

Steps in LSCF PBE are as follows:

1. Calculate the difference between the test and reference drugs averages $I_{1j} = \frac{(Y_{1jT1}+Y_{1jT2})}{2} - \frac{(Y_{1jR1}+Y_{1jR2})}{2}$ and $I_{2j} = \frac{(Y_{2jT1}+Y_{2jT2})}{2} - \frac{(Y_{2jR1}+Y_{2jR2})}{2}$ for each of the sequences i=1, 2. Their average is the location estimate for the difference in test and reference drugs.

2. Calculate U_{ijk} and V_{ijk} as explained in the LSCF procedure. With these, the between and within variances are estimated for the aggregate test statistic.

3. Calculate the test and reference drug variances as

$$\sigma_T^2 = \frac{Var(U_{1T}) + Var(U_{2T})}{2} + \frac{1}{2}\frac{Var(V_{1T}) + Var(V_{2T})}{2},$$

$$\sigma_R^2 = \frac{Var(U_{1R}) + Var(U_{2R})}{2} + \frac{1}{2}\frac{Var(V_{1R}) + Var(V_{2R})}{2}$$

and the difference in test and reference drug location with $\delta = \frac{I_{1j}+I_{2j}}{2}$.

4. Estimate the aggregate measure η, as shown below

$$\hat{\eta} = \hat{\delta}^2 + \widehat{\sigma_T^2} - (1 + 1.744826) \max\left(\widehat{\sigma_R^2}, 0.04\right).$$

5. Add outliers to 5% of the data i.e on six subjects. Rerun the above procedure calculate $\widehat{\eta}$. Increment these outliers with $\pm$ 1,2,3,4,5,6 σ.

Thus the LSCF estimate of η with or without outliers is calculated. Now, to estimate robust η, use the five proposed procedures for the cases of with or without outliers.

Steps in robust bootstrap PBE are as follows:

1. Start with the data as in table 6. Using the log transformed response (Y_{ijkl}), calculate the difference between the Test and Reference drug averages with $I_{1Tj} = \frac{(Y_{1jT1}+Y_{1jT2})}{2}$, $I_{1Rj} = \frac{(Y_{1jR1}+Y_{1jR2})}{2}$, $I_{2Tj} = \frac{(Y_{2jT1}+Y_{2jT2})}{2}$ and $I_{2Rj} = \frac{(Y_{2jR1}+Y_{2jR2})}{2}$ for each of the sequences i=1, 2. Calculate the difference in location of the test and reference drugs for the two sequences explained above as $\xi_T - \xi_R = \frac{\text{Median}_{I_{1Tj}}+\text{Median}_{I_{2Tj}}}{2} - \frac{\text{Median}_{I_{1Rj}}+\text{Median}_{I_{2Rj}}}{2}$.

2. Calculate U_{ijk} and V_{ijk} as explained in the LSCF procedure. By calculating them, estimate the between and within spread used in estimating the aggregate test statistic.

3. Calculate the test and reference drug spreads $\sigma_T^2 = \frac{Var(U_{1T})+Var(U_{2T})}{2} + \frac{1}{2}\frac{Var(V_{1T})+Var(V_{2T})}{2}$ and $\sigma_R^2 = \frac{Var(U_{1R})+Var(U_{2R})}{2} + \frac{1}{2}\frac{Var(V_{1R})+Var(V_{2R})}{2}$ and $\delta = \frac{I_{1j}+I_{2j}}{2}$.

   Here δ (the robust location) is the difference $\xi_T - \xi_R$ and Var are the variances estimated in each case by Gini, MAD, IQR, S_n and Q_n as described below. The unbiased estimators of the variance in each of these cases are:

   Gini : $\sigma^2 = \left(G\frac{\sqrt{\pi}}{2}\right)^2$

$$\text{MAD}: \sigma^2 = (1.4826 \cdot MAD)^2$$

$$\text{IQR}: \sigma^2 = \left(\frac{IQR}{1.34898}\right)^2$$

$$S_n: \sigma^2 = (1.1926 \cdot \text{med}_i \left(\text{med}_j \left(|x_i - x_j|\right)\right))^2$$

$$Q_n: \sigma^2 = \left(2.219 \left\{|x_i - x_j|; i < j\right\}_{(k)}\right)^2.$$

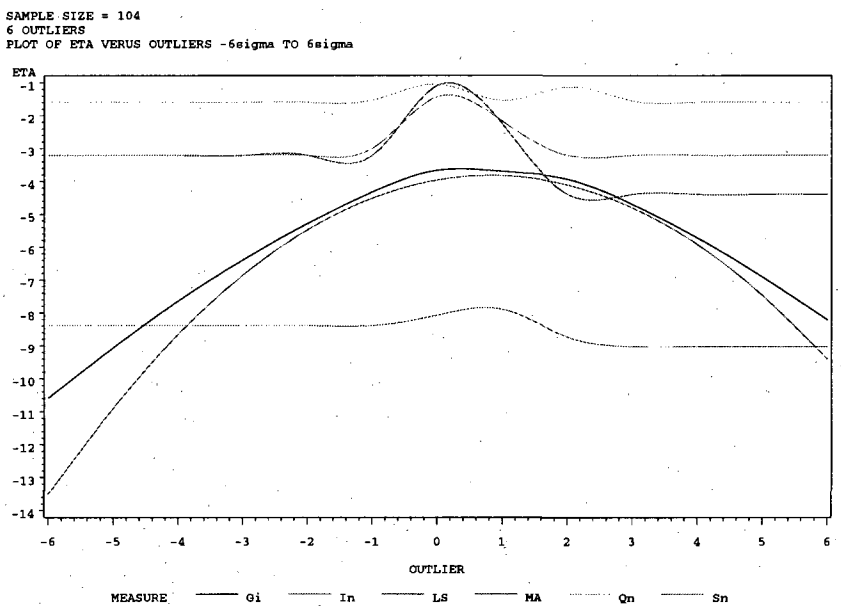
4. Estimate η for the five procedures using the robust location and variance estimates as

$$\hat{\eta} = \widehat{\delta^2} + \widehat{\sigma_T^2} - (1 + 1.744826) \max\left(\widehat{\sigma_R^2}, 0.04\right).$$

5. Add outliers to 5% of the data i.e six subjects. Rerun the above procedure and calculate $\widehat{\eta}$. Increment the outliers with $\pm$ 1,2,3,4,5,6 σ.

Compare the results of sensitivity analysis of LSCF to the proposed five procedures. The plot of $\widehat{\eta}$ versus the incremental outliers from -6σ to 6σ is shown in figure 3. By increasing

Figure 3: Large sample PBE sensitivity analysis



outliers, the LS procedures i.e LSCF and Gini are most affected. The robust procedures

are very stable and Q_n is the most stable robust procedure. Gini is marginally better than LSCF since median was used in location estimation. As the outliers are increased on either side to $\pm 6\sigma$, $\widehat{\eta}$ varied from -4 to -13 for LSCF while Q_n varied from -1.2 to -1.5.

Rousseeuw and Croux proposed the Q_n estimate of scale as an alternative to MAD. It shares desirable robustness properties with MAD (50% breakdown point, bounded influence function). In addition, it has significantly better normal efficiency (82%) and it does not depend on symmetry. Qn is the most stable procedure to estimate η in the presence of outliers. A simulation study comparing the validity and power of the LSCF with the proposed bootstrap procedures is conducted. The next section discusses this comparison.

3.4 PBE comparison of level and power

In the simulation analysis, generate data as in table 2. By controlling the input parameters, η is fixed. These parameters include the various between and within variances and the means of the test and the reference drugs.

By setting the true value of η at the boundary i.e zero, calculate the significance level as the probability of falsely rejecting the null. By setting the true vale of η at the rejection region, calculate the power as a function of the probability of falsely accepting the null. Further on the basis of MSE, the better procedure is identified.

3.4.1 Validity

To test for validity, set the hypothesis at the boundary condition. The hypothesis of interest is

$$H_0 : \eta \geq 0 : \text{(Non Population Bioequivalent)},$$

$$H_1 : \eta < 0 : \text{(Population Bioequivalent)}.$$

The definition of type I error is P_{H_0}(Reject the Null hypothesis)=α. At the boundary, the value of $\eta = 0$ and the probability of the type I error is maximum.

1. Set the true value of $\eta = 0$ as shown below

$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, \sigma_0^2)\,\theta_P = 0$$
$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, 0.04)\,1.744826 = 0.$$

   One of the possible boundary condition could be setup by $\mu_T = \mu_R$ and $\sigma_T^2 = \sigma_R^2 + \max(\sigma_R^2, 0.04)\,1.744826$. As an example let the mean differences be set to zero ($\mu_T - \mu_R = 0$), the variances set to σ_R^2=0.3 and σ_T^2=0.8234478. Such a setup has true $\eta = 0$.

2. After specifying the input parameters, generate two hundred datasets having a bivariate normal distribution of the form

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 \end{pmatrix}\right].$$

   For each of the datasets, calculate $\widehat{\eta}$ and $\widehat{\eta_{95\%}}$ for the LSCF and the five proposed procedures. For each of the robust bootstrap procedure, conduct two thousand bootstraps on each of the two hundred datasets to obtain two hundred $\widehat{\eta_{95\%}}$.

3. Calculate the proportion of cases when the null is rejected. This proportion represents the empirical probability : P_{H_0} (Reject H_0) = α. Compare this empirical α from LSCF, Gini, MAD, Q_n, S_n and IQR. Calculate the mean squared errors (MSE) with the two hundred datasets for each of the procedure as:

$$MSE = \sqrt{\sum_{i=1}^{i=P} \frac{(\widehat{\eta_i} - \eta_{True})^2}{(P-1)}}.$$

Thus, the empirical level and MSE of the LSCF and the five proposed procedures are computed. Next, compute the empirical power of the six procedures.

3.4.2 Power

To compute the empirical power, set the true value of η in the alternative condition. The hypothesis is

$$H_0 : \eta \geq 0 : (NonBioequivalent)$$
$$H_1 : \eta < 0 : (Bioequivalent).$$

Definition of type II error is P_{H_A}(Fail to Reject the Null hypothesis) and power = 1 - P(Type II error).

1. Set the true value of η less than zero as shown below

$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, \sigma_0^2)\,\theta_P = -0.80,$$
$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, 0.04)\,1.744826 = -0.80.$$

   For example one of the possible boundary condition setup could be μ_T - μ_R = -0.2, the variances σ_T^2 = 0.34 and σ_R^2 = 0.43. Since η_{True} = -0.80, the null should be rejected.

2. After specifying the input parameters, generate two hundred datasets that are distributed as bivariate normal of the form

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 \end{pmatrix} \right].$$

   For each of the datasets, calculate $\widehat{\eta}$ and $\widehat{\eta_{95\%}}$ for the LSCF and the five proposed procedures. For each of the robust bootstrap procedure, conduct two thousand bootstraps

on each of the two hundred datasets to obtain two hundred $\widehat{\eta_{95\%}}$.

3. Calculate the proportion of cases the null is accepted. This proportion represents the empirical probability of P_{H_A} (Fail to reject H_0) = P(Type II error). The empirical power is 1 - P(Type II error) for LSCF, Gini, MAD, Q_n, S_n and IQR. Calculate MSE using the two hundred datasets for each procedure as

$$MSE = \sqrt{\sum_{i=1}^{i=P} \frac{(\widehat{\eta_i} - \eta_{True})^2}{(P-1)}}.$$

The next section discusses the findings of the simulation study comparing validity and power of the present LSCF with the five proposed procedures.

3.5 Examples comparing validity and power

For simulation, the between and within variances were set based upon the FDA (2001) guidelines and from Chow et al (2002). The possible values of the variance σ_T^2 and σ_R^2 vary from a range of 0.15 to 0.5.

Define small outliers as 3σ outliers and large outliers as 6σ outliers. These outliers are set based upon the criteria that at least 5% of the data may possess outliers. AUC_∞ and Cmax contain outliers due to prolonged excretion rate of the drug or the absorption rate depending upon the subject. Outliers are added to five subjects in the data. The outliers are in two main categories. Outliers in the test drug or outliers in the reference drug.

In a simulation study of two thousand bootstraps on samples of size fifteen to twenty five, the bootstrap was found to be inconsistent. This may be attributed to the inconsistent covariance structure during bootstraps. However, consistent results were found for samples of size sixty or above. Hence, samples of size hundred, hundred and fifty and two hundred are used.

For each of the cases, validity and power is computed. From this, a graph is plotted that displays the differences. Results of the simulation procedure are summarized below.

3.5.1 Type I error (α) and power (γ) with small test outliers

Graph A.1 plots power and α which are calculated for small test outliers. The graphical summary is obtained from the type I error table B.4 and power from the table B.3.

For the case of small variability ($\sigma^2 = 0.15$), the LSCF procedure performed better than the remaining procedures in both level and power. The next best procedure comparable to LSCF is Gini. Both LSCF and Gini are comparable in their MSE.

With larger variability($\sigma^2 = 0.5$), it is noted that the LSCF procedure is not the best. IQR, S_n seems a lot more efficient than before with smaller MSE. However, Gini is better than LSCF in both power and level. LSCF and Gini worked best with smaller test drug variance and smaller outliers.

3.5.2 Type I error (α) and power (γ) with small reference outliers

Graph A.2 plots power and α which are calculated for small reference outliers. The graphical summary is obtained from the type I error table B.6 and power from table B.5.

For the case of small variability ($\sigma^2 = 0.15$), LSCF procedure and Gini have higher significance level (15%). With such a level, power has little meaning and thus the LS procedures failed. Q_n is better among the various robust procedures.

With larger variability ($\sigma^2 = 0.5$), the LS procedures, LSCF and Gini have large significance level and all the robust procedures MAD, IQR, S_n and Q_n performed better. So, with outliers in the reference drug, it is clear that the validity of the LS procedure is severely affected.

3.5.3 Type I error (α) and power (γ) with large test outliers

Graph A.3 plots power and α which are calculated for large Test outliers. The graphical summary is obtained from the type I error table B.8 and the power from table B.7.

For the case of small variability ($\sigma^2 = 0.15$), the LS procedures compromised with the significance level of the test. IQR, S_n are more conservative tests and the robust procedures are better overall and have higher power.

With larger variability ($\sigma^2 = 0.5$), the LS procedures are worse for both validity and power. All the robust procedures work well and are more efficient with smaller MSE. Robust procedures work best with larger Test outliers.

3.5.4 Type I error (α) and power (γ) with large reference outliers

Graph A.4 plots the power and α which are calculated for large reference outliers. The graphical summary is obtained from the type I error table B.10 and power from table B.9.

LSCF and Gini, the two LS procedures are compromised due to outliers and this is seen by their level. In both small and large variances of the data, Q_n is the most conservative with significance level and has high power. Overall, the robust procedures perform better when there are more than 3σ outliers.

3.6 Small sample study

As seen in these simulations, consistent results for samples of size sixty or above are obtained. However such samples are available only on phase II of the drug development. So, it becomes necessary to address the cases of clinical trials where samples of size twenty are quite commonly used. In Leena et al. (2008), typical BE tests are conducted on subjects of size twelve to thirty. For small samples, bootstrap procedures may be of suspect because the covariance structure may breakdown and also the outliers may have a greater effect at

such small sample sizes. In the next chapter, the small sample analysis of PBE is addressed.

CHAPTER IV

SMALL SAMPLE POPULATION BIOEQUIVALENCE

PBE analyzed in phase I of a clinical trial have small sample sizes. The FDA (2001), Hyslop et al. (2000) and Patterson et al (2002) have used small sample sizes for PBE analysis in their papers. Small sample sizes refer to samples of size N=18, 22, etc. With small samples, the bootstrap procedure previously developed does not give consistent results.

In this chapter, the theory developed by Chinchilli et al (1996), Cornish et al (1938), Stefan (2001) and Anirban et al (2008) is used to calculate the Cornish-Fisher confidence interval using closed forms of Gini and IQR. Gini and IQR have a readily available closed form distribution. Estimate the mean difference, variances and the population bioequivalence criterion. Population BE is established for a particular log-transformed BA measure if the 95% upper confidence bound for the linearized criterion is less than or equal to zero (FDA, 2001).

4.1 Distributional assumptions of metrics in BE trials

Before performing a statistical analysis in BE trials, AUC and Cmax are generally log transformed. The three most commonly cited reasons for log transforming AUC and Cmax are

- AUC is non-negative
- Distribution of AUC is highly skewed
- PK models are multiplicative

The drug concentration at each time point is a function of many random processes. They are absorption, distribution, metabolism and elimination that act proportionally to the amount of the drug present in the body. Thus the resulting distribution is log normal (Midha & Gavalas, 1993).

4.2 Design

The test (T) and the reference (R) formulations are administered to healthy volunteers and the drug concentrations are measured over time. A cross-over design is setup to compare the test and reference drug formulation's effect on a subject. For PBE, a 2x4 cross-over design i.e a two sequence, four period replicated balanced design (FDA, 2001) as explained above is considered.

The data in table 2 for PBE of large samples is also used here. Apart from this sample size, the rest of the parameters are reused for the setup. For the first sequence, subjects have a TRTR schedule and for the second sequence a RTRT schedule. The design is as follows

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl} \tag{4.1}$$

where i=1,...,s indicates the number of sequences, j=1,...,n_i indicates the subjects within each sequence, k=R,T indicates the treatments, l=1,...,p_{ik} indicate replicates on treatment k for subjects within sequence i.

The response is Y_{ijkl} for replicate l on treatment k for subject j in sequence i and γ_{ikl} is the fixed effect while the random effect is δ_{ijk} for subject j with a random error ϵ_{ijkl}. The random errors ϵ_{ijkl} are mutually independent and identically distributed as

$$\begin{pmatrix} \epsilon_{ijTl} \\ \epsilon_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{WithinT} & 0 \\ 0 & \sigma^2_{WithinR} \end{pmatrix}\right]. \tag{4.2}$$

Also, the random subject interaction effect is distributed as shown below

$$\begin{pmatrix} \delta_{ijT} \\ \delta_{ijR} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BR} \end{pmatrix}\right]. \tag{4.3}$$

The resulting response is distributed as

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}\right]. \tag{4.4}$$

The next section introduces the hypothesis to test PBE.

4.2.1 Hypothesis

The proposed null and alternative hypothesis based on the FDA regulations (2001) are

$$\begin{aligned} H_0 &: \frac{(\mu_T - \mu_R)^2 + \sigma^2_T - \sigma^2_R}{max(\sigma^2_0, \sigma^2_R)} \geq \theta_P \\ H_1 &: \frac{(\mu_T - \mu_R)^2 + \sigma^2_T - \sigma^2_R}{max(\sigma^2_0, \sigma^2_R)} < \theta_P \end{aligned} \tag{4.5}$$

where $\sigma^2_T = \sigma^2_{WT} + \sigma^2_{BT}$ and $\sigma^2_R = \sigma^2_{WR} + \sigma^2_{BR}$ are the total variances of the test and the reference drugs. The constants σ^2_0 and θ_P are fixed regulatory standards.

As seen above, the FDA guidance currently adopts an aggregate approach, using an aggregated test statistic for evaluating both means and variance components simultaneously. In contrast, several disaggregate approaches have been suggested where tests for each component are performed separately. For example, Liu and Chow (1996) proposed a disaggregate approach for evaluating IBE where three components (intra subject variability, subject-by-formulation interaction, and average) are separately tested multiple times

with intersection-union tests. However, as the dimensions (p) of tests increases, the power of the $(1-2\alpha)$ confidence set (Leena Choi, 2008) based approach could decrease sharply for dimensions greater than one as shown in Hwang (1996).

The aggregated test statistic is linearized as follows:

$$H_0 : (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2) \geq 0,$$
$$H_1 : (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2) < 0. \quad (4.6)$$

Here, $\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2)$ and the null hypothesis reduces to a one sided problem defined by a linear combination. The FDA fixed 0.02 as the maximum difference for the variance under the test and reference formulations. Usually θ = log 1.25 = -log 0.80 = 0.223. These values (FDA, 2001) originated from the notion that the ratio of the population means in the original scale (the mean of the test is 80 - 125% of that of the reference) are considered to be sufficiently close for drugs having an average therapeutic window. For PBE, the FDA sets θ_P = 1.744826 and σ_0^2 = 0.04. The linearized hypothesis is of the form

$$H_0 : \eta \geq 0,$$
$$H_1 : H_0 : \eta < 0.$$

If the null is rejected, population bioequivalence (the two drugs are similar across population groups) is inferred. Otherwise, the two drugs are significantly different across the populations. The next section describes the present procedure of testing PBE hypothesis.

4.2.2 Least squares Cornish Fisher's procedure (LSCF)

The present procedure tests PBE using Cornish Fisher's (CF) (1938) expansion. In LSCF, η is calculated as $\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_P * max(\sigma_0^2, \sigma_R^2)$. The procedures in estimating μ_i and σ_i^2 are described below. If the upper confidence interval $\eta_{95\%}$ is less than

zero, population bioequivalence is concluded.

Following are the steps in computing the least squares Cornish Fisher's (LSCF) expansion:

1. From table 2, the response Y_{ijkl} is distributed as

$$\begin{pmatrix} Y_{ijRl} \\ Y_{ijTl} \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma^2_{BR} + \sigma^2_{WR} & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma^2_{BT} + \sigma^2_{WT} \end{pmatrix} \right]$$

   where each subject j has two observations for one of the two treatments. Each subject belongs to only one sequence. The data has 'N' subjects partitioned into two sequences with $\frac{N}{2}$ subjects in each sequence. In this example, a balanced design is used. The variances σ^2_B and σ^2_W are the between and within variances. For the first sequence the patients have a TRTR schedule and the second sequence subjects have an RTRT schedule.

2. Define I as the difference in test and reference drug replicate averages. Compute this difference I_{ij} as

$$I_{1j} = \frac{(Y_{1jT1} + Y_{1jT2})}{2} - \frac{(Y_{1jR1} + Y_{1jR2})}{2},$$
$$I_{2j} = \frac{(Y_{2jT1} + Y_{2jT2})}{2} - \frac{(Y_{2jR1} + Y_{2jR2})}{2}$$

   for each of the sequence i=1, 2.

3. Calculate U_{ijT} as the average of the test drug replicates and U_{ijR} as the average of

the reference drug replicates. This average is

$$U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}.$$

U_{1jT} and U_{2jT} are independent as they are estimates from two different independent samples.

4. Define V_{ijT} as the difference of the replicates of the test and V_{ijR} as the difference of the replicates of the reference drug. V_{ijk} is calculated as

$$V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}},$$
$$V_{1jR} = \frac{(Y_{1jR1} - Y_{1jR2})}{\sqrt{2}}.$$

5. Calculate the variance of the variables U_{ijk}, V_{ijk} for each of the two sequences. Estimate the variance of test drug σ_T^2 as $\sigma_{BT}^2 + \sigma_{WT}^2$ and reference drug σ_R^2 as $\sigma_{BR}^2 + \sigma_{WR}^2$. For the first sequence, the variance is estimated with

$$Var(U_{1jT}) = \frac{Var(Y_{1jT1}) + Var(Y_{1jT2}) + 2Cov(Y_{1jT1}, Y_{1jT2})}{4},$$
$$Var(V_{1jT}) = \frac{Var(Y_{1jT1}) + Var(Y_{1jT2}) - 2Cov(Y_{1jT1}, Y_{1jT2})}{2}.$$

Without loss of generality, set the covariance (Σ_1) for the first sequence and the two test drug periods. The resulting distribution of the test drug in the first sequence is

$$\begin{pmatrix} Y_{1jT1} \\ Y_{1jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_1 \\ \Sigma_1 & \sigma_T^2 \end{pmatrix} \right]. \tag{4.7}$$

Similarly, the distribution of the test drug in the second sequence is

$$\begin{pmatrix} Y_{2jT1} \\ Y_{2jT2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \Sigma_2 \\ \Sigma_2 & \sigma_T^2 \end{pmatrix}\right]. \tag{4.8}$$

It can be proved that σ_T^2 is a linear combination of the variances σ_{UT}^2 and σ_{VT}^2. To prove $\sigma_T^2 = \sigma_{UT}^2 + \frac{\sigma_{VT}^2}{2}$, consider the following

$$\begin{aligned}
\sigma_T^2 &= \overline{\sigma_{UT}^2} + \overline{\frac{\sigma_{VT}^2}{2}} \\
\sigma_T^2 &= \tfrac{1}{2}\left(\sigma_{UT_{seq1}}^2 + \sigma_{UT_{seq2}}^2\right) + \tfrac{1}{2}\left(\tfrac{1}{2}\left\{\sigma_{VT_{seq1}}^2 + \sigma_{VT_{seq2}}^2\right\}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left(\sigma_{UT_{seq1}}^2 + \frac{\sigma_{VT_{seq1}}^2}{2}\right) + \tfrac{1}{2}\left(\sigma_{UT_{seq2}}^2 + \frac{\sigma_{VT_{seq2}}^2}{2}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left(Var(U_{1jT}) + \frac{Var(V_{1jT})}{2}\right) + \tfrac{1}{2}\left(Var(U_{2jT}) + \frac{Var(V_{2jT})}{2}\right) \\
\sigma_T^2 &= \tfrac{1}{2}\left[\left(\tfrac{\sigma_T^2+\Sigma_1}{2}\right) + \left(\tfrac{\sigma_T^2-\Sigma_1}{2}\right)\right] + \tfrac{1}{2}\left[\left(\tfrac{\sigma_T^2+\Sigma_2}{2}\right) + \left(\tfrac{\sigma_T^2-\Sigma_2}{2}\right)\right].
\end{aligned}$$

By expanding the above equation, it is concluded that

$$\sigma_{UT}^2 + \frac{\sigma_{VT}^2}{2} = \sigma_T^2.$$

Similarly, for the reference drug, $\left(\sigma_{UR}^2 + \frac{\sigma_{VR}^2}{2}\right) = \sigma_R^2$.

6. The expected values of the difference for the test and reference drugs from the two sequences across the four periods or two replicates using equation 4.7 are

$$\begin{aligned}
E\left(I_{1j}\right) &= E\left[\frac{\left(Y_{1jT1} + Y_{1jT2}\right)}{2} - \frac{\left(Y_{1jR1} + Y_{1jR2}\right)}{2}\right] = \frac{2\mu_T - 2\mu_R}{2}, \\
E\left(I_{2j}\right) &= E\left[\frac{\left(Y_{2jT1} + Y_{2jT2}\right)}{2} - \frac{\left(Y_{2jR1} + Y_{2jR2}\right)}{2}\right] = \frac{2\mu_T - 2\mu_R}{2}.
\end{aligned}$$

Thus from the average of the two sequences, $\frac{E(I_{1j})+E(I_{2j})}{2} = \mu_T - \mu_R$.

7. Estimate the aggregate statistic η using the linear combinations of means and variances as

$$\widehat{\eta} = \left(\widehat{\mu_T - \mu_R}\right)^2 + \widehat{\sigma_T^2} - (1+\theta_P)\, max(\widehat{\sigma_R^2}, 0.04).$$

Calculate the upper confidence interval of $\widehat{\eta}$ using the Cornish Fisher's expansion. To illustrate CF's expansion, consider H as the upper bound in the equation

$$H = \sum P_q + \left(\sum B_q\right)^{\frac{1}{2}}$$

where P_q represents the point estimates i.e mean, variances and B_q represents the upper bound of these point estimates (95%).

8. Table 7 outlines the various point estimates and their respective upper bounds.

Table 7: Point estimates and their distributions

P_q=Point Estimate	C=Confidence Bound	B_q=Upper α limit
$P_1=\widehat{(\mu_T - \mu_R)^2}$	$U_1 = \left(\widehat{P_1} + t_{1-\alpha,N-s}\left(\sum_{i=1}^{s} n_i^{-1} s_I^2\right)^{\frac{1}{2}}\right)^2$	$B_1=\left(U_1 - \widehat{P_1}\right)^2$
$P_2=\widehat{\sigma_{Uk}^2}$	$U_2 = \widehat{\sigma_{Uk}^2}\frac{(N-2)}{\chi^2_{\alpha,N-2}}$	$B_2=\left(U_2 - \widehat{\sigma_{Uk}^2}\right)^2$
$P_3=\frac{1}{2}\widehat{\sigma_{Vk}^2}$	$U_3 = \frac{1}{2}\widehat{\sigma_{Vk}^2}\frac{(N-2)}{\chi^2_{\alpha,N-2}}$	$B_3=\left(U_3 - \frac{1}{2}\widehat{\sigma_{Vk}^2}\right)^2$

Thus, calculate the upper CI of $\widehat{\eta}$ using Cornish Fisher's expansion. The upper 95%

confidence value of $\widehat{\eta}$ is calculated as

$$\hat{\eta} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_T^2} - (1+\theta_P)\max\left(\widehat{\sigma_R^2}, \sigma_0^2\right),$$

$$\hat{\eta} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_{UT}^2} + \tfrac{1}{2}\widehat{\sigma_{VT}^2} - (1+\theta)\max\left(\widehat{\sigma_{UR}^2} + \tfrac{1}{2}\widehat{\sigma_{VR}^2}, \sigma_0^2\right),$$

$$\widehat{\eta}_{1-\alpha} = \widehat{(\mu_T - \mu_R)}^2 + \widehat{\sigma_{UT}^2} + \tfrac{1}{2}\widehat{\sigma_{VT}^2} - (1+\theta)\max\left(\widehat{\sigma_{UR}^2} + \tfrac{1}{2}\widehat{\sigma_{VR}^2}, \sigma_0^2\right)$$
$$+\Bigg\{\left[\left(\left|\widehat{\mu_T - \mu_R}\right| + t_{\alpha,N-2}\sqrt{\tfrac{s_l^2}{n_1+n_2-2}}\right)^2 - \widehat{(\mu_T - \mu_R)}^2\right]^2 + \left[\tfrac{(N-2)\widehat{\sigma_{UT}^2}}{\chi^2_{\alpha,N-2}} - \widehat{\sigma_{UT}^2}\right]^2$$
$$+\left[\tfrac{1}{2}\tfrac{(N-2)\widehat{\sigma_{VT}^2}}{\chi^2_{\alpha,N-2}} - \tfrac{1}{2}\widehat{\sigma_{VT}^2}\right]^2 + \left[\tfrac{-(1+\theta_P)(N-2)\widehat{\sigma_{UR}^2}}{\chi^2_{\alpha,N-2}} + (1+\theta_P)\,\widehat{\sigma_{UR}^2}\right]^2$$
$$+\left[\tfrac{-(1+\theta_P)(N-2)\frac{1}{2}\widehat{\sigma_{VR}^2}}{\chi^2_{\alpha,N-2}} + (1+\theta_P)\,\tfrac{1}{2}\widehat{\sigma_{VR}^2}\right]^2\Bigg\}^{1/2}.$$

Once $\widehat{\eta}_{95\%}$ is computed, conclude PBE if $\widehat{\eta}_{95\%}$ is less than zero. When H_0 is rejected, PBE is concluded. The following section proposes the robust bootstrap procedure as an alternative to the LSCF procedure.

4.2.3 Proposed small sample procedures

The robust procedure is identical to the LSCF procedure in terms of data manipulation and the grouping to calculate I_{ijk}, U_T, U_R. A closed form distributions of Gini and IQR is suggested for CF expansion. Steps for small sample PBE analysis are as follows:

1. Start with the data in table 2 where each subject j has two observations for one of the two treatments. The subjects belong to only one sequence. The data has N subjects partitioned into two sequences with $\frac{N}{2}$ subjects in each sequence i.e a balanced design. σ_B^2 and σ_W^2 are the between and within variances for sequence i and replicate(period) l. From the setup, the first sequence has a TRTR schedule and the second sequence has a RTRT schedule.

2. Define I as the averages of the replicates of test and reference drugs. Calculate I_{ijk}

Table 8: Two sequence, four period balanced design

Subject	Sequence	Period1	Period2	Period3	Period4
1	1	Y_{1jT1}	Y_{1jR1}	Y_{1jT2}	Y_{1jR2}
2	1	.	.	.	.
.	1	.	.	.	.
.	1	.	.	.	.
m	1	.	.	.	.
m+1	2	Y_{2jR1}	Y_{2jT1}	Y_{2jR2}	Y_{2jT2}
m+2	2	.	.	.	.
.	2	.	.	.	.
.	2	.	.	.	.
j	2	.	.	.	.

as

$$I_{1jT} = \frac{(Y_{1jT1}+Y_{1jT2})}{2},$$
$$I_{1jR} = \frac{(Y_{1jR1}+Y_{1jR2})}{2},$$
$$I_{2jT} = \frac{(Y_{2jT1}+Y_{2jT2})}{2},$$
$$I_{2jR} = \frac{(Y_{2jR1}+Y_{2jR2})}{2}.$$

for each of the sequences i=1, 2. This gives the average effects of the test and the reference drugs for the two sequences.

3. Define U_{ijT} as the averages of the replicates of test and U_{ijR} as the averages of the replicates of the reference drugs. Calculate them as

$$U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2},$$
$$U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}.$$

Here U_{1jT} and U_{2jT} are independent as they are estimates from two different independent samples.

4. Define V_{1jT} as the difference of replicates of test drugs for the first sequence and V_{2jT}

as the difference of the replicates of test drugs for the second sequence. Calculate them as

$$V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}},$$
$$V_{2jT} = \frac{(Y_{2jT1} - Y_{2jT2})}{\sqrt{2}}.$$

Here V_{1jT} and V_{2jT} are independent as they are estimates from two different independent samples.

5. Obtain the robust location i.e the median of I_{1Tj}, I_{1Rj}, I_{2Tj} and I_{2Rj}. Using this, calculate the robust estimate of location difference as

$$\xi_T - \xi_R = \frac{\text{Median}_{I_{1Tj}} + \text{Median}_{I_{2Tj}}}{2} - \frac{\text{Median}_{I_{1Rj}} + \text{Median}_{I_{2Rj}}}{2}.$$

6. Use Gini and IQR as variance estimators. The standard errors of these variance estimators are readily available as shown below.

   - **IQR**: Based on the large sample assumption, the robust variance estimate of IQR is calculated. For estimating the scale parameter, σ of a location scale density $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, an estimate based on the interquartile range IQR=$X_{\left[\frac{3n}{4}:n\right]} - X_{\left[\frac{1n}{4}:n\right]}$ is used. Use of such an estimate is quite common when normality is suspect (DasGupta, 2008). IQR is distributed as

$$\sqrt{n}\left(IQR - \left(\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}}\right)\right) \rightarrow N\left(0, \frac{3}{f^2\left(\xi_{\frac{3}{4}}\right)} + \frac{3}{f^2\left(\xi_{\frac{1}{4}}\right)} - \frac{2}{f\left(\xi_{\frac{3}{4}}\right) f\left(\xi_{\frac{1}{4}}\right)}\right).$$

In particular, if $X_1, ..., X_n$ are iid $N(\mu, \sigma^2)$, then

$$\sqrt{n}\left(IQR - 1.35\sigma\right) \rightarrow N\left(0, 2.48\sigma^2\right). \tag{4.9}$$

Consequently, for normal data, $\frac{IQR}{1.35}$ is a consistent estimator of σ (DasGupta & Haff, 2006).

- **Gini Mean Difference :** Gini's mean difference is often used as an alternative to the standard deviation as a measure of spread (Nair, 1936).

  "The Mean Difference introduced by Prof. Corrado Gini as a measure of variation is defined as: If $x_1, x_2, ...x_n$ are n observed values of a variate x, the mean difference is defined as

  $$g = \frac{2}{n(n-1)} \sum_{i \neq j} |x_i - x_j|.$$

  "

The standard error of Gini's mean difference (g) was presented by U.S. Nair (1936) and further explained by Lomnicki (1952). If X's are normal $N(\mu, \sigma^2)$, the unbiased estimator of σ is $\sqrt{\pi}g/2$. To obtain the above proof, use the approximation theorems (Serfling, 2001), proofs by Nair (1936) and David (1968). The sketch of the theory is

$$g = \frac{2}{n(n-1)} \sum_{i \neq j} |x_i - x_j| = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |x_{ni} - x_{nj}|$$
$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (x_{ni} - x_{nj})$$

where x_{ni} and x_{nj} are the order statistics. Now Gini written a linear combination of the order statistics. Nair generated normal convergence of Gini by

expanding the x's as follows:

$$
\begin{aligned}
g &= \tfrac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (x_{ni} - x_{nj}), \\
&= \tfrac{2}{n(n-1)} \sum_{i=1}^{n} [(x_i - x_1) + (x_i - x_2) + \ldots + (x_i - x_{i-1})], \\
&= \tfrac{2}{n(n-1)} \left[2 \sum_{i=1}^{n} i x_i - (n+1) \sum_{i=1}^{n} x_i \right], \\
&= \tfrac{2}{n(n-1)} \left[2U - (n+1)V \right].
\end{aligned}
$$

By estimating U, V, U^2, UV and V^2 and their expectations, estimate the mean and variance of g. Take Jacobian at every stage and add them up as

$$
\begin{aligned}
E(g) = \bar{g} &= \tfrac{2}{n(n-1)} \left[2\bar{U} - (n+1)\bar{V} \right], \\
g^2 &= \tfrac{4}{n^2(n-1)^2} \left[4U^2 - 4(n+1)UV + (n+1)^2 V^2 \right], \\
\overline{g^2} = E(g^2) &= \tfrac{4}{n^2(n-1)^2} \left[4\overline{U^2} - 4(n+1)\overline{UV} + (n+1)^2 \overline{V^2} \right], \\
\sigma_g^2 &= E(g^2) - \overline{E(g)}^2.
\end{aligned}
$$

When X's are normally distributed, the Jacobian are estimated for the mean of g and the variance of g. They are

$$
\begin{aligned}
\overline{g} &= \tfrac{2\sigma}{\sqrt{\pi}}, \\
\sigma_g &= \tfrac{2\sigma}{\sqrt{n(n-1)}} \left[\tfrac{n+1}{3} + \tfrac{2\sqrt{3}(n-2)}{\pi} - \tfrac{2(2n-3)}{\pi} \right]^{\frac{1}{2}}.
\end{aligned}
$$

For a sample of size 10, the efficiency of this estimate is 98.1% and reaches 99% for small increments of sample size (David, 1968).

> "Gini is also slightly less sensitive to the presence of outliers than either s or σ. Although necessarily entailing a considerable loss in efficiency under normality, a symmetrically censored version of σ^*

has been put forward as

$$\sqrt{n}\left(g - \frac{2\sigma}{\sqrt{\pi}}\right) \dot{\rightarrow} N\left(0, \sigma_g\right)" \tag{4.10}$$

From these derivations, a Cornish-Fisher's expansion using Gini and IQR is generated that tests for small sample PBE.

7. Estimate the upper 95% confidence interval of $\widehat{\eta}$ using the following procedure

   - For the location, use Moses (Hollander & Wolfe, 2001) distribution free confidence interval based on Wilcoxon's rank sum test. For the upper 95% confidence interval of the difference in Test and Reference location, calculate

$$C_\alpha = \frac{n(2m+n+1)}{2} + 1 - w_\alpha,$$
$$\Delta_U = U^{mn+1-C_\alpha}$$

     where m and n represent the sample sizes. U is a value that is estimated from Hollander et al (2001). Δ_U is the desired upper confidence interval of the location differences.

   - In order to estimate the upper confidence limit of the variance estimates of Gini mean difference, use equation 4.10.

   - To estimate the upper confidence limit of the variance estimate of IQR, use equation 4.9.

8. The upper 95% confidence level of $\widehat{\eta}$ is estimated for Gini mean difference as

$$\begin{aligned}
\eta_{1-\alpha} &= \widehat{(\xi_T - \xi_R)^2} + \widehat{\zeta_{UT}^2} + \tfrac{1}{2}\widehat{\zeta_{VT}^2} - (1+\theta)\max\left(\widehat{\zeta_{UR}^2} + \tfrac{1}{2}\widehat{\zeta_{VR}^2}, \sigma_0^2\right) \\
&+ \left\{ \left[\Delta_U^2 - \widehat{(\xi_T - \xi_R)^2}\right]^2 + \left[\left(\widehat{\zeta_{UT}^2} + Z_\alpha * \widehat{\phi_{UT}}\right) - \widehat{\zeta_{UT}^2}\right]^2 \right. \\
&+ \left[\tfrac{1}{2}\left(\widehat{\zeta_{VT}^2} + Z_\alpha * \widehat{\phi_{VT}}\right) - \widehat{\tfrac{1}{2}\zeta_{VT}^2}\right]^2 \\
&+ \left[-(1+\theta_P)\left(\widehat{\zeta_{UR}^2} + Z_\alpha * \widehat{\phi_{UR}}\right) + (1+\theta_P)\widehat{\zeta_{UR}^2}\right]^2 \\
&+ \left.\left[-(1+\theta)\left(\widehat{\tfrac{1}{2}\zeta_{UR}^2} + Z_\alpha * \widehat{\tfrac{1}{2}\phi_{UR}}\right) + (1+\theta)\widehat{\tfrac{1}{2}\zeta_{UR}^2}\right]^2 \right\}^{1/2}
\end{aligned}$$

where $\widehat{\zeta_i^2}$ is the asymptotically unbiased variance estimate and $\widehat{\phi} = \sqrt{\frac{\pi}{4}\frac{\sigma_g^2}{n}\{2^2\sigma^2\}}$ is the standard error obtained using equation 4.10 and the Delta method. This is shown by the following equations :

$$\begin{aligned}
&\sqrt{n}\left(\tfrac{\sqrt{\pi}}{2}g - \sigma\right) \rightharpoonup N\left(0, \sigma_g^2\tfrac{\pi}{4}\right), \\
&\sqrt{n}\left(\left(\tfrac{\sqrt{\pi}}{2}g\right)^2 - \sigma^2\right) \rightharpoonup N\left(0, \sigma_g^2\tfrac{\pi}{4}\left(\tfrac{d\sigma^2}{d\sigma}\right)^2\right), \\
&\sqrt{n}\left(\left(\tfrac{\sqrt{\pi}}{2}g\right)^2 - \sigma^2\right) \rightharpoonup N\left(0, \sigma_g^2\tfrac{\pi}{4}(2\sigma)^2\right).
\end{aligned}$$

Similarly, for IQR calculate $\widehat{\eta}_{95\%}$ as

$$\begin{aligned}
\eta_{1-\alpha} &= \widehat{(\xi_T - \xi_R)^2} + \widehat{\tau_{UT}^2} + \tfrac{1}{2}\widehat{\tau_{VT}^2} - (1+\theta)\max\left(\widehat{\tau_{UR}^2} + \tfrac{1}{2}\widehat{\tau_{VR}^2}, \sigma_0^2\right) \\
&+ \left\{ \left[\Delta_U^2 - \widehat{(\xi_T - \xi_R)^2}\right]^2 + \left[\left(\widehat{\tau_{UT}^2} + Z_\alpha * \widehat{\varphi_{UT}}\right) - \widehat{\tau_{UT}^2}\right]^2 \right. \\
&+ \left[\tfrac{1}{2}\left(\widehat{\tau_{VT}^2} + Z_\alpha * \widehat{\varphi_{VT}}\right) - \widehat{\tfrac{1}{2}\tau_{VT}^2}\right]^2 \\
&+ \left[-(1+\theta_P)\left(\widehat{\tau_{UR}^2} + Z_\alpha * \widehat{\varphi_{UR}}\right) + (1+\theta_P)\widehat{\tau_{UR}^2}\right]^2 \\
&+ \left.\left[-(1+\theta)\left(\widehat{\tfrac{1}{2}\tau_{UR}^2} + Z_\alpha * \widehat{\tfrac{1}{2}\varphi_{UR}}\right) + (1+\theta)\widehat{\tfrac{1}{2}\tau_{UR}^2}\right]^2 \right\}^{1/2}
\end{aligned}$$

where $\widehat{\tau_i^2}$ is the asymptotically unbiased variance estimate and

$\widehat{\varphi} = \sqrt{\left(\frac{2.48}{1.35^2}\right)\frac{\sigma^2}{n}(2\sigma)^2}$ is the standard error obtained using the Delta method. This is proved by

$$\sqrt{n}\left(\frac{IQR}{1.35} - \sigma\right) \dashrightarrow N\left[0, \frac{2.48}{1.35^2}\sigma^2\right]$$
$$\sqrt{n}\left(\left(\frac{IQR}{1.35}\right)^2 - \sigma^2\right) \dashrightarrow N\left[0, \frac{2.48}{1.35^2}\sigma^2\left(\frac{d\sigma^2}{d\sigma}\right)^2\right]$$
$$\sqrt{n}\left(\left(\frac{IQR}{1.35}\right)^2 - \sigma^2\right) \dashrightarrow N\left[0, \frac{2.48}{1.35^2}\sigma^2(2\sigma)^2\right]$$

With the above two proposed procedures, compare small sample PBE using LSCF with small sample CF Gini and IQR. In the next section, sensitivity analysis is conducted on an example with these three procedures.

4.3 Sensitivity analysis of an example

Apply the LSCF, Gini and IQR procedures on a dataset. This dataset was procured from a FDA website which was created on August 18, 2003. Introduction to the dataset used is

> "In reference to the Federal Register notice on "Preliminary Draft Guidance for Industry on In Vivo bioequivalence Studies Based on Population and Individual bioequivalence Approaches: Availability", vol. 62, No. 249, Dec. 30, 1997, the Food and Drug Administration (FDA) is announcing the availability of data that were used by the Agency in support of the proposal and the detailed description of statistical methods for individual and population approaches."

The dataset in table 9 is 'DRUG 17A' from the above source. It is a two sequence, four period replicate design with thirty six subjects who are randomized into one of the two sequences. The subjects in the first sequence start with a RTTR (Reference-Test-Test-Reference) schedule while the second sequence have a TRRT schedule. There is a sufficient washout period between the test and reference drugs to avoid carryover effect. AUC_∞ is the parameter of interest. Reorder the data by transposing the data on periods. The design

Table 9: Example to illustrate the PBE procedure

SUBJECT	PER	SEQ	TRT	AUC	AUCINF	CMAX
1	1	RTTR	2	1020.65	1020.65	109
1	2	RTTR	1	1321.23	1321.23	145
1	3	RTTR	1	900.42	900.42	106
1	4	RTTR	2	1173.61	1173.61	146
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
36	1	TRRT	1	2212.39	2212.39	226
36	2	TRRT	2	1438.48	1438.48	137
36	3	TRRT	2	1984.76	1984.76	237
36	4	TRRT	1	2640.43	2640.43	237

used is similar to the design from the large sample PBE procedure. This design is

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \epsilon_{ijkl} \tag{4.11}$$

where i=1,2 indicates the number of sequences, j=1,...,36 indicates the subjects within each sequence, k=R,T indicates the treatments, l=1,2 indicates replicates on treatment k for subjects within sequence i. Due to a balanced design, there are eighteen subjects in the first sequence and eighteen subjects in the second sequence.

The response is Y_{ijkl} for replicate l on treatment k for subject j in sequence i. The fixed effect is γ_{ikl} of replicate l on treatment k in sequence i. The random effect is δ_{ijk} for subject j in sequence i on treatment k and ϵ_{ijkl} is a random error.

- Steps in small sample LSCF are as follows:

  1. For the above dataset, calculate the difference between the test and reference drug averages $I_{1j} = \frac{(Y_{1jT1}+Y_{1jT2})}{2} - \frac{(Y_{1jR1}+Y_{1jR2})}{2}$ and $I_{2j} = \frac{(Y_{2jT1}+Y_{2jT2})}{2} - \frac{(Y_{2jR1}+Y_{2jR2})}{2}$ for each of the two sequences i=1, 2. Also calculate U_{ijk} and V_{ijk} with $U_{1jT} = \frac{(Y_{1jT1}+Y_{1jT2})}{2}$, $U_{2jT} = \frac{(Y_{2jT1}+Y_{2jT2})}{2}$ and $V_{1jT} = \frac{(Y_{1jT1}-Y_{1jT2})}{\sqrt{2}}$,

$V_{2jT} = \frac{(Y_{2jT1} - Y_{2jT2})}{\sqrt{2}}$.

2. Calculate $\widehat{\delta}$, $\widehat{\sigma_T^2}$ and $\widehat{\sigma_R^2}$ as explained in the previous section. Estimate η as $\hat{\eta} = \widehat{\delta^2} + \widehat{\sigma_T^2} - (1 + 1.744826) \max\left(\widehat{\sigma_R^2}, 0.04\right)$.

3. Apply outliers to 5% of the data. After adding outliers to three subjects, rerun the above procedure and calculate $\widehat{\eta}$. These outliers are $\pm$ 1,2,3,4,5,6 σ outliers.

- Steps in small sample robust procedure using Gini and IQR are as follows

1. Calculate the test and reference drugs averages with $I_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2}$, $I_{1jR} = \frac{(Y_{1jR1} + Y_{1jR2})}{2}$, $I_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}$ and $I_{2jR} = \frac{(Y_{2jR1} + Y_{2jR2})}{2}$ for each of the sequences *i*=1, 2. The robust difference in location is $\xi_T - \xi_R = \frac{\text{Median}_{I_{1Tj}} + \text{Median}_{I_{2Tj}}}{2} - \frac{\text{Median}_{I_{1Rj}} + \text{Median}_{I_{2Rj}}}{2}$.

2. Calculate U_{ijk} and V_{ijk} with $U_{1jT} = \frac{(Y_{1jT1} + Y_{1jT2})}{2}$, $U_{2jT} = \frac{(Y_{2jT1} + Y_{2jT2})}{2}$ and $V_{1jT} = \frac{(Y_{1jT1} - Y_{1jT2})}{\sqrt{2}}$ and $V_{2jT} = \frac{(Y_{2jT1} - Y_{2jT2})}{\sqrt{2}}$.

3. Calculate the averages for the two sequences
$\sigma_T^2 = \frac{Var(U_{1T}) + Var(U_{2T})}{2} + \frac{1}{2}\frac{Var(V_{1T}) + Var(V_{2T})}{2}$ and
$\sigma_R^2 = \frac{Var(U_{1R}) + Var(U_{2R})}{2} + \frac{1}{2}\frac{Var(V_{1R}) + Var(V_{2R})}{2}$ and $\delta = \frac{I_{1j} + I_{2j}}{2}$.
For Gini, $\sigma^2 = \left(G\frac{\sqrt{\pi}}{2}\right)^2$ and for IQR, $\sigma^2 = \left(\frac{IQR}{1.34898}\right)^2$. Estimate η for the data above as $\hat{\eta} = \widehat{\delta^2} + \widehat{\sigma_T^2} - (1 + 1.744826) \max\left(\widehat{\sigma_R^2}, 0.04\right)$. and $\widehat{\eta_{95}}$ by the following procedure. Upper 95% confidence level of η is estimated for Gini

mean difference as

$$\eta_{1-\alpha} = \widehat{(\xi_T - \xi_R)}^2 + \widehat{\zeta_{UT}^2} + \tfrac{1}{2}\widehat{\zeta_{VT}^2} - (1+\theta)\max\left(\widehat{\zeta_{UR}^2} + \tfrac{1}{2}\widehat{\zeta_{VR}^2}, \sigma_0^2\right)$$
$$+ \left\{ \left[\Delta_U^2 - \widehat{(\xi_T - \xi_R)}^2\right]^2 + \left[\left(\widehat{\zeta_{UT}^2} + Z_\alpha * \widehat{\phi_{UT}}\right) - \widehat{\zeta_{UT}^2}\right]^2 \right.$$
$$+ \left[\tfrac{1}{2}\left(\widehat{\zeta_{VT}^2} + Z_\alpha * \widehat{\phi_{VT}}\right) - \widehat{\tfrac{1}{2}\zeta_{VT}^2}\right]^2$$
$$+ \left[-(1+\theta_P)\left(\widehat{\zeta_{UR}^2} + Z_\alpha * \widehat{\phi_{UR}}\right) + (1+\theta_P)\widehat{\zeta_{UR}^2}\right]^2$$
$$\left. + \left[-(1+\theta)\left(\widehat{\tfrac{1}{2}\zeta_{UR}^2} + Z_\alpha * \widehat{\tfrac{1}{2}\phi_{UR}}\right) + (1+\theta)\widehat{\tfrac{1}{2}\zeta_{UR}^2}\right]^2 \right\}^{1/2}$$

where $\widehat{\zeta_i^2}$ is the asymptotically unbiased variance estimate and $\widehat{\phi} = \sqrt{\frac{\pi}{4}\frac{\sigma_g^2}{n}\{2^2\sigma^2\}}$ the standard error obtained using equations 4.10 and delta method. Similarly, for IQR

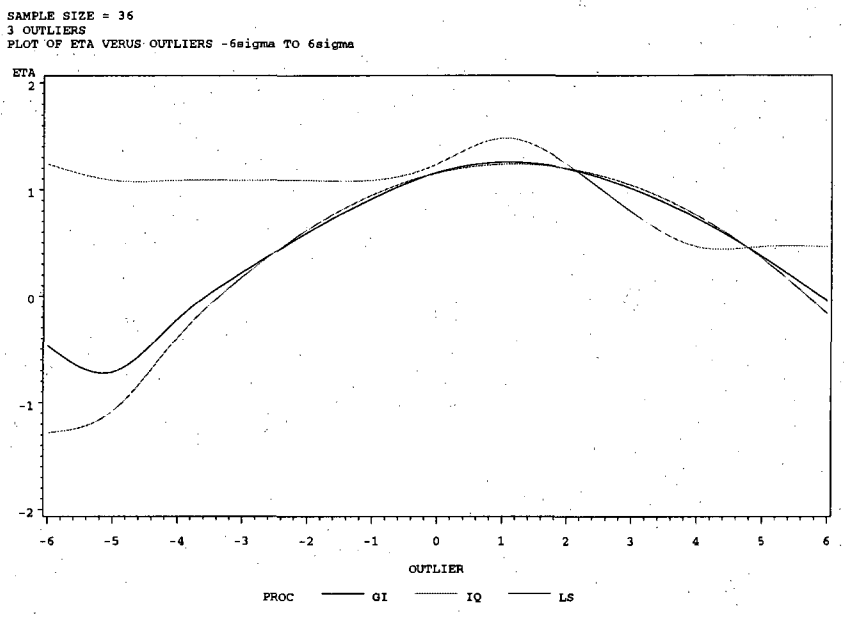
$$\eta_{1-\alpha} = \widehat{(\xi_T - \xi_R)}^2 + \widehat{\tau_{UT}^2} + \tfrac{1}{2}\widehat{\tau_{VT}^2} - (1+\theta)\max\left(\widehat{\tau_{UR}^2} + \tfrac{1}{2}\widehat{\tau_{VR}^2}, \sigma_0^2\right)$$
$$+ \left\{ \left[\Delta_U^2 - \widehat{(\xi_T - \xi_R)}^2\right]^2 + \left[\left(\widehat{\tau_{UT}^2} + Z_\alpha * \widehat{\varphi_{UT}}\right) - \widehat{\tau_{UT}^2}\right]^2 \right.$$
$$+ \left[\tfrac{1}{2}\left(\widehat{\tau_{VT}^2} + Z_\alpha * \widehat{\varphi_{VT}}\right) - \widehat{\tfrac{1}{2}\tau_{VT}^2}\right]^2$$
$$+ \left[-(1+\theta_P)\left(\widehat{\tau_{UR}^2} + Z_\alpha * \widehat{\varphi_{UR}}\right) + (1+\theta_P)\widehat{\tau_{UR}^2}\right]^2$$
$$\left. + \left[-(1+\theta)\left(\widehat{\tfrac{1}{2}\tau_{UR}^2} + Z_\alpha * \widehat{\tfrac{1}{2}\varphi_{UR}}\right) + (1+\theta)\widehat{\tfrac{1}{2}\tau_{UR}^2}\right]^2 \right\}^{1/2}$$

where α=0.05, N=36, n1=n2=18, θ_P=1.744826, θ=0.04, $\widehat{\tau_i^2}$ is the asymptotically unbiased variance estimate and $\widehat{\varphi} = \sqrt{\left(\frac{2.48}{1.35^2}\right)\frac{\sigma^2}{n}(2\sigma)^2}$ the standard error obtained using delta method.

4. Apply outliers to 5% of the data. After adding outliers to three subjects, rerun the above procedure and calculate $\widehat{\eta}$. These outliers are $\pm$ 1,2,3,4,5,6 σ.

The section below describes the comparison of the small sample LSCF procedure with the two proposed procedures. With the analysis of this procedure, the results are summa-

Figure 4: Small sample PBE sensitivity analysis



rized in graph 4. As the outliers are increased in size, LSCF and Gini are most affected. Gini is marginally better than LSCF. With outliers ranging from -6σ to +6σ, $\widehat{\eta}$ from LSCF procedure varied from 1.2 to -1.2 while that of IQR varied from 1.2 to 0.3.

Such a variation in the test statistic changes the conclusion of the hypothesis due to outliers. Clearly IQR is more resistant to outliers than the LS procedures. In the next section, validity and power of the LSCF procedure is compared to the proposed procedures.

4.4 Small sample PBE comparison of level and power

Simulation analysis generated data as in table 2. By controlling the input parameters, η is fixed. These parameters include the various between and within variances and the means of the test and the reference drugs.

By setting the true value of η at the boundary i.e zero, calculate the significance level by the

probability of falsely rejecting the null. By setting the true vale of η at the rejection region, calculate the power as a function of the probability of falsely accepting the null. Further on the basis of MSE, determine the better procedure. The simulation analysis is run for the following cases:

1. Mild test drug formulation outliers which have 3σ outliers,
2. Mild reference drug formulation outliers which have 3σ outliers,
3. Mild outliers which have 3σ outliers for both test and reference drug formulations,
4. Large outliers which have 6σ outliers for both test and reference drug formulations.

The values of small and large variances are obtained from publications as seen in previous chapters.

4.4.1 Validity

To test for validity, set the hypothesis at the boundary condition. The hypothesis of interest is

$$H_0 : \eta \geq 0 : \text{(Non Population Bioequivalent)},$$
$$H_1 : \eta < 0 : \text{(Population Bioequivalent)}.$$

The definition of type I error is P_{H_0}(Reject the Null hypothesis)=α. At the boundary, the value of $\eta = 0$ and the probability of the type I error is maximum.

1. Set the true value of $\eta = 0$ as shown below

$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, \sigma_0^2)\,\theta_P = 0$$
$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, 0.04)\,1.744826 = 0.$$

One of the possible boundary condition could be setup by $\mu_T = \mu_R$ and $\sigma_T^2 = \sigma_R^2 + \max(\sigma_R^2, 0.04)\,1.744826$. As an example let the mean differences be set to zero

($\mu_T - \mu_R = 0$), the variances set to σ_R^2=0.3 and σ_T^2=0.8234478. Such a setup has true $\eta = 0$.

2. After specifying the input parameters, generate two thousand datasets having a bivariate normal distribution of the form

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 \end{pmatrix}\right].$$

   For each of the datasets, calculate $\widehat{\eta}$ and $\widehat{\eta_{95\%}}$ for the LSCF and the two proposed procedures.

3. Calculate the proportion of cases when the null is rejected. This proportion represents the empirical probability P_{H_0} (Reject H_0) = α. Calculate the mean squared errors (MSE) from two thousand $\widehat{\eta}$.

With these steps, the empirical significance level is computed for LSCF and the two proposed procedures.

4.4.2 Power

To compute the empirical power, set the true value of η in the alternative condition. The hypothesis is

$$H_0 : \eta \geq 0 : (NonBioequivalent)$$
$$H_1 : \eta < 0 : (Bioequivalent).$$

Definition of type II error is P_{H_A}(Fail to Reject the Null hypothesis) and power = 1 - P(Type II error).

1. Set the true value of η less than zero as shown below

$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, \sigma_0^2)\,\theta_P = -0.80,$$
$$\eta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \max(\sigma_R^2, 0.04)\,1.744826 = -0.80.$$

   For example one of the possible boundary condition setup could be μ_T - μ_R = -0.2, σ_T^2 = 0.34 and σ_R^2 = 0.43. Since η_{True} = -0.80, the null should be rejected.

2. After specifying the input parameters, generate two thousand datasets that are distributed as bivariate normal of the form

$$\begin{pmatrix} Y_{ijTl} \\ Y_{ijRl} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 \end{pmatrix}\right].$$

   For each of the datasets, calculate $\widehat{\eta}$ and $\widehat{\eta_{95\%}}$ for the LSCF and the two proposed procedures.

3. Calculate the proportion of cases when the null is accepted. This proportion represents the empirical probability of P_{H_A} (Fail to reject H_0) = P(Type II error). We now have empirical power as 1 - P(Type II error) for LSCF, Gini and IQR. Calculate MSE using the two thousand datasets.

The next section discusses the findings of the simulation study comparing validity and power of the present LSCF with the two proposed procedures.

4.5 Examples comparing validity and power

The small sample case for simulation based upon the suggested variances of between and within factors from the FDA guidelines FDA (2001) and from Chow et al (2002) is elaborated. The variances are broadly categorized into small and large variance and further with

small and large outliers.

Further the outliers are limited to only one or two subjects out of the twenty subjects. The outliers are then bifurcated into two main categories, outliers in the test drug or outliers in the reference drug. These outliers were set based upon the criteria that at least 5% of the data contains outliers. AUC_{∞} and Cmax are quite easily prone to outliers due to prolonged excretion rate of the drug or the absorption rate depending upon the subject. Calculate validity, power and MSE from the simulated datasets. With this, a comparative graph is plotted for the three procedures.

4.5.1 Power and level α with small outliers

Graph A.6 plots the power and level α which are calculated for one test drug outlier. A subject's test drug response was offset by 0σ to 4σ outliers. The graphical summary is obtained from the type I error table B.13 and power from the table B.14.

The results of LSCF and Gini are similar initially. But due to the robust location, Gini ended up being the better of the two procedures with outliers. IQR is not efficient with outliers and has high MSE. Since IQR is less conservative in level, LSCF and Gini are better procedures for data with modest outliers.

4.5.2 Power and level α with large outliers

Graph A.6 plots power and level α which are calculated for two test drug outliers. However, in this case, the outliers were both on the test and reference drug formulations.

IQR is more stable than the LS procedures with large outliers. However, IQR is a less conservative procedure with high significance level. Both LSCF and Gini are severely affected with outliers. LSCF and Gini have close results and both procedures failed their validity due to outliers. Further research is needed to resolve this effect of outliers.

CHAPTER V

AVERAGE BIOEQUIVALENCE

The two treatment, two period (2 x 2) crossover trial is routinely used to test average bioequivalence for two drugs. In this trial, subjects are randomly assigned to two groups, usually of equal size. Subjects in the first group receive treatment 'T' followed by treatment 'R' (TR schedule) and vice versa for the other group (RT schedule). A suitable washout period is imposed between treatments in order to eliminate potential carryover effects of the first treatment. After the administration of each treatment, blood samples are collected at fixed time points, and the concentration of the drug in the blood is quantified. The typical primary endpoint of interest is the area under the drug concentration versus time curve (AUC), which represents the bioavailability of the drug. The two treatments are declared bioequivalent if their true relative average bioavailability is estimated to be within prespecified 'bioequivalence limits' with high confidence (Stefanescu & Mehrotra, 2008).

The normality of log(AUC) and log(Cmax) are discussed in the previous chapters. A two one-sided hypothesis test is followed in the next section.

5.1 Distributional assumptions of metrics in BE trials

For the statistical analysis in BE trials, AUC and Cmax are generally log transformed. The three most commonly cited reasons for using the log transformed AUC are

- AUC is non-negative
- Distribution of AUC is highly skewed
- PK models are multiplicative

Since the drug concentration at each time is a function of many random processes (absorption, distribution, metabolism and elimination) that reasonably would act proportionally to the amount of drug present in the body, this suggests that the resulting distribution is log-normal (Midha & Gavalas, 1993).

5.2 Design

Table 10 presents a dataset with two sequences and two periods. There is a sufficient washout period between the two periods to prevent any carry over effect. The design sug-

Table 10: Two sequence, two period balanced design

Subject	Sequence	Period1	Period2
1	1	Y_{1jT}	Y_{1jR}
.	1	.	.
.	1	.	.
m+1	2	Y_{2jR}	Y_{2jT}
.	2	.	.
j	2	.	.

gested by the FDA (2001) and Devan et al. (2008) is of the form

$$y_{ijk} = \pi_i + \mu_k + s_{j(i)} + e_{ijk}. \tag{5.1}$$

The response y_{ijk} is the log transformed AUC or log transformed Cmax for treatment k and subject j within sequence i. $s_{j(i)}$ is the random effect and e_{ijk} the random error. Thus for

the two sequences and two periods, the responses are

$$
\begin{aligned}
y_{1j1} &= \pi_1 + \mu_1 + s_{j(1)} + e_{1j1}, \\
y_{1j2} &= \pi_1 + \mu_2 + s_{j(1)} + e_{1j2}, \\
y_{2j2} &= \pi_2 + \mu_2 + s_{j(2)} + e_{2j2}, \\
y_{2j1} &= \pi_2 + \mu_1 + s_{j(2)} + e_{2j1}.
\end{aligned}
$$

Assume the random subject effect $s_{j(i)}$ to be independently and identically distributed as N(0 , ϕ_1) and the random error e_{ijk}, also independently and identically distributed as N(0 , ϕ_0). $s_{j(i)}$ and e_{ijk} are mutually independent (Stefanescu & Mehrotra, 2008).

Take the difference between the test and reference drug responses as suggested in Stefanescu & Mehrotra (2008). This difference is seen as

$$
\begin{aligned}
y_{1j1} - y_{1j2} &= \mu_1 - \mu_2 + e_{1j1} - e_{1j2}, \\
y_{2j1} - y_{2j2} &= \mu_1 - \mu_2 + e_{2j1} - e_{2j2}.
\end{aligned}
$$

The random subject effect is eliminated. The response matrix is thus a multivariate matrix with two columns that are the log transformed AUC and Cmax differences. In the next section, the hypothesis to test for ABE is presented.

5.2.1 Hypothesis

The FDA (2003a) directs testing the difference in location effects using Schruimann's two one-sided hypothesis. The limits 0.8 and 1.25 are fixed by the FDA (2003a). The multi-

variate hypothesis is of the form

$$H_{01}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \leq \ln 0.8 \cup H_{02}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \geq \ln 1.25,$$

$$H_{A1}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} > \ln 0.8 \cap H_{A2}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} < \ln 1.25.$$

Set $\Delta\mu_{AUC}$ as the mean difference of the test and reference drugs for AUC and $\Delta\mu_{Cmax}$ as the mean difference of the test and reference drugs for Cmax.

5.2.2 LS procedure

For the LS procedure, the location estimate of the difference $\mu_T - \mu_R$ is obtained by the simple mean difference of the response for the two periods. This is calculated as shown

$$Z_{ij_{AUC}} = y_{ijT_{AUC}} - y_{ijR_{AUC}},$$
$$Z_{ij_{C\max}} = y_{ijT_{C\max}} - y_{ijR_{C\max}}.$$

The sample averages of the differences $Z_{ij_{AUC}}$ and $Z_{ij_{Cmax}}$ are $\bar{\Delta}\mu_{AUC}$ and $\bar{\Delta}\mu_{Cmax}$. These averages are distributed as

$$\underset{\sim}{\bar{Z}} = \begin{pmatrix} \bar{\Delta}_{AUC} \\ \bar{\Delta}_{C\max} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \Delta_{AUC} \\ \Delta_{C\max} \end{pmatrix}, \frac{1}{n}\Sigma \right)$$

where $\Delta_{AUC} = \mu_T - \mu_R$ for AUC and $\Delta_{Cmax} = \mu_T - \mu_R$ for Cmax. The covariance is

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$$

with $\sigma_{11}^2 = var(Z_{ij_{AUC}})$ and $\sigma_{22}^2 = var(Z_{ij_{Cmax}})$ and $\sigma_{12} = \sigma_{21} = covar(Z_{ij_{AUC}}, Z_{ij_{Cmax}})$.

5.2.3 Componentwise rank method

The Componentwise rank (CR) method (Hettmansperger & McKean, 1998) is used on the vector of Wilcoxon signed-rank statistics on each component. The procedure involves setting

$$S_4(\theta) = \begin{pmatrix} \sum \frac{R(|x_{i1}-\theta_1|)}{n+1} \operatorname{sgn}(x_{i1}-\theta_1) \\ \sum \frac{R(|x_{i2}-\theta_2|)}{n+1} \operatorname{sgn}(x_{i2}-\theta_2) \end{pmatrix}$$

and for $\theta = 0$,

$$S_4(0) = \begin{pmatrix} \sum F_1^+(|x_{i1}|) \operatorname{sgn}(x_{i1}) \\ \sum F_2^+(|x_{i2}|) \operatorname{sgn}(x_{i2}) \end{pmatrix} + o_p(1) = \begin{pmatrix} 2\sum F_1^+(x_{i1}) - \frac{1}{2} \\ 2\sum F_2^+(x_{i2}) - \frac{1}{2} \end{pmatrix} + o_p(1)$$

where $F_j{}^+$ is the marginal distribution of $|X_{1j}|$ for j=1,2 and F_j is the marginal distribution of X_{1j}. Symmetry of the marginal distributions is used in the computation of the projections. We now identify A and B for the purpose of constructing the quadratic form of the test statistic, the asymptotic distribution of the vector of estimates and the non centrality parameter.

Since the multivariate central limit theorem can be applied on the project,

- The components of $S(\theta)$ should be non-increasing functions of θ_1 and θ_2
- $\mathrm{E}_0(S(0)) = 0$
- $\frac{1}{\sqrt{n}} S(0) \xrightarrow{D} Z \sim N_2(0, A)$
- $\sup_{\|b\| \le B} \left| \frac{1}{\sqrt{n}} S\left(\frac{1}{\sqrt{n}} b\right) - \frac{1}{\sqrt{n}} S(0) + \grave{B} b \right| \xrightarrow{P} 0$

the first two conditions are satisfied. Since under the null-hypothesis $\theta = 0$, $F(X_{i1})$ has a uniform distribution on (0,1) and introducing θ and differentiating with respect to θ_1 and

θ_2, the A and B matrices are

$$\frac{1}{n}A = \begin{pmatrix} \frac{1}{3} & \delta \\ \delta & \frac{1}{3} \end{pmatrix}$$

and

$$B = \begin{pmatrix} 2\int f_1^2(t)\,dt & 0 \\ 0 & 2\int f_2^2(t)\,dt \end{pmatrix}$$

where $\delta = 4\int\int F_1(s)\,F_2(t)\,dF(s,t) - 1$. Hence, similar to the vector of sign statistics, the vector of Wilcoxon signed rank statistics also have a covariance that depends on the underlying bivariate distribution. A consistent estimate of δ in A is given by

$$\widehat{\delta} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_{it}R_{jt}}{(n+1)(n+1)} \operatorname{sgn}(X_{it}) \operatorname{sgn}(X_{jt})$$

where R_{it} is the rank of $|X_{it}|$ in the t^{th} component among $|X_{1t}|, ..., |X_{nt}|$. This estimate is the conditional covariance and can be used in estimating A in the construction of an asymptotically distribution free test. For estimating the asymptotic covariance matrix of $\widehat{\theta}$ center the data and then compute. From the programs in the website (McKean, 2009), the robust spread is estimated.

The estimator that solves $S_4(\theta)$ is the vector of Hodges-Lehmann (HL) estimates for the two components i.e the vector of medians of Walsh averages for each component. Like the vector of medians, the vector of HL estimates is not equivalent under the orthogonal transformations and the test is not invariant under these transformations. This will show up in the efficiency with respect to the L_2 methods which are an equivariant estimate and an invariant test. From the robust analog, the location and covariance structure for the multivariate setting is estimated.

The location estimate of the difference $\mu_T - \mu_R$ is obtained from the Hodges

Lehmann's estimate for the differences of the form

$$Z_{ij_{AUC}} = y_{ijT_{AUC}} - y_{ijR_{AUC}},$$
$$Z_{ij_{C\max}} = y_{ijT_{C\max}} - y_{ijR_{C\max}}.$$

The robust location estimate $\widehat{Z}$ is distributed as

$$\underset{\sim}{\widehat{Z}} = \begin{pmatrix} \widehat{\Delta}_{AUC} \\ \widehat{\Delta}_{C\max} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \Delta_{AUC} \\ \Delta_{C\max} \end{pmatrix}, \frac{1}{n}\Sigma \right)$$

where $\widehat{\Delta}_{AUC}$ is the Hodges Lehmann's estimate of the vector of differences $Z_{ij_{AUC}}$ for the AUC and $\widehat{\Delta}_{Cmax}$ is the Hodges Lehmann's estimate of the vector of differences $Z_{ij_{Cmax}}$ for Cmax. These estimates have a covariance structure of

$$\Sigma = \frac{1}{n} B^{-1} A B^{-1}$$

where A and B matrices are calculated from the procedures explained above in Componentwise rank method. The variance from the robust procedure is

$$\Sigma = \frac{1}{n} \begin{bmatrix} \frac{1}{12\left[\int f_1^2(t)dt\right]^2} & \frac{\delta}{2\int f_1^2(t)dt.2\int f_2^2(t)dt} \\ \frac{\delta}{2\int f_1^2(t)dt.2\int f_2^2(t)dt} & \frac{1}{12\left[\int f_2^2(t)dt\right]^2} \end{bmatrix}$$

$$\Sigma = \frac{1}{n} \begin{bmatrix} \tau_1^2 & 3\delta\tau_1\tau_2 \\ 3\delta\tau_1\tau_2 & \tau_2^2 \end{bmatrix}$$

where $\tau_i = \frac{1}{\sqrt{12}\int f_i^2(t)dt}$

$\widehat{\delta} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_{it}R_{jt}}{(n+1)(n+1)} \mathrm{sgn}\left(X_{it}\right) \mathrm{sgn}\left(X_{jt}\right)$ where R_{it} is the rank of $|X_{it}|$ in the t^{th} component among $|X_{1t}|, ..., |X_{nt}|$ and $\widehat{\tau}_i$ is estimated as in Koul *et al.* (1987)

5.2.4 Ellipse generation

To calculate the confidence region, estimate the location and covariance matrix Σ. The $100(1-\alpha)\%$ confidence region for the mean of a p-dimensional distribution is determined by μ such that

$$n\left(\overline{x}-\mu\right)^{1} S^{-1}\left(\overline{x}-\mu\right) \leq \frac{p\left(n-1\right)}{\left(n-p\right)} F_{p,n-p}\left(\alpha\right) \tag{5.2}$$

where $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, $S = \frac{1}{(n-1)}\sum_{j=1}^{n}\left(x_j-\overline{x}\right)\left(x_j-\overline{x}\right)^{1}$ and x_1, x_2, ..., x_n are the sample observations (Johnson & Wichern, 1992) (here, p=2).

To construct a confidence region as an ellipse, the center and the lengths of the major and minor axes are needed. The direction and lengths of the axes of

$$n\left(\bar{x}-\mu\right)^{1} S^{-1}\left(\bar{x}-\mu\right) \leq c^{2} = \frac{p\left(n-1\right)}{\left(n-p\right)} F_{p,n-p}\left(\alpha\right)$$

are determined by

$$\sqrt{\frac{\lambda_i}{n}}c = \sqrt{\lambda_i}\sqrt{\frac{p\left(n-1\right)F_{p,n-p}\left(\alpha\right)}{n\left(n-p\right)}}$$

units along the eigen vectors e_i. Beginning at the center $\overline{x}$ or Hodges Lehmann's (HL) estimate, the axes of the confidence region ellipse are

$$\pm\sqrt{\lambda_i}\sqrt{\frac{p\left(n-1\right)}{n\left(n-p\right)}F_{p,n-p}\left(\alpha\right)}\, e_i$$

where $Se_i=\lambda_i e_i$ and i=1,2,...,p. The ratios of the λ_i are the relative elongation along pairs of axes. Construct an ellipse for the multivariate LS procedure and the Componentwise rank method and study the effect of outliers on this ellipse. In the next section, conduct sensitivity analysis on an example comparing the LS procedure with the proposed robust procedure.

5.3 Example of ABE

The datasets used for the sensitivity analysis were procured from the FDA website which was created on August 18, 2003. An overview of these datasets is

> "In reference to the Federal Register notice on "Preliminary Draft Guidance for Industry on In Vivo bioequivalence Studies Based on Population and Individual bioequivalence Approaches: Availability", vol. 62, No. 249, Dec. 30, 1997, the Food and Drug Administration (FDA) is announcing the availability of data that were used by the Agency in support of the proposal and the detailed description of statistical methods for individual and population approaches. "

Table 11: Example to illustrate the ABE procedure

ID	Seq	Period	TMT	AUC	Cmax
1	2	2	1	0.605305	1.525045
.	.	.	.	.	.
.	.	.	.	.	.
24	2	2	1	0.20412	1.08636
1	2	1	2	0.60206	1.534026
.	.	.	.	.	.
.	.	.	.	.	.
24	2	1	2	0.225309	1.113943

The dataset in table 11 used in the example is 'DRUG 25A' from the above source. This example is a two sequence, two period replicate design with twenty four subjects who are randomized into one of the two sequences. The subjects in the first sequence start with TR schedule while the second sequence subjects have an RT schedule. There is a sufficient washout period between the test and reference drugs to avoid carryover effect. Log transformed AUC and Cmax are shown in the table.

Take the difference in test and reference drug responses as shown in Devan et al. (2008). These differences in the two periods of a sequence are

$$y_{1jT} - y_{1jR} = \mu_T - \mu_R + e_{1jT} - e_{1jR},$$
$$y_{2jT} - y_{2jR} = \mu_T - \mu_R + e_{2jT} - e_{2jR}.$$

Estimate Hotelling T^2 test statistic for the LS and the robust procedures. Add outliers to the data and rerun sensitivity analysis on it. These outliers are $\pm$1,2,3,4,5,6σ

5.3.1 Hotelling T^2 with ABE LS procedure

Start with the differences $y_{1jT} - y_{1jR}$ and $y_{2jT} - y_{2jR}$. With these differences, calculate the sample means and the sample variances. The sample averages of the differences are distributed as

$$\underset{\sim}{\bar{Z}} = \begin{pmatrix} \bar{\Delta}_{AUC} \\ \bar{\Delta}_{C\max} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \Delta_{AUC} \\ \Delta_{C\max} \end{pmatrix}, \frac{1}{n}\Sigma \right).$$

Hotelling T^2 test statistic for the LS procedure is calculated as

$$T^2 = n \begin{bmatrix} (\bar{\Delta}_{AUC} - \Delta_{AUC}) & (\bar{\Delta}_{C\max} - \Delta_{C\max}) \end{bmatrix} \hat{\Sigma}^{-1} \begin{bmatrix} (\bar{\Delta}_{AUC} - \Delta_{AUC}) \\ (\bar{\Delta}_{C\max} - \Delta_{C\max}) \end{bmatrix}$$

where $\widehat{\Sigma}$ is the sample variance covariance matrix. For this analysis, set Δ_{AUC} and Δ_{Cmax} to zero. To this data add outliers varying from -6σ to 6σ and rerun the above procedure and collect Hotelling T^2 estimates.

5.3.2 Hotelling T^2 with ABE CR method

Start with the differences $y_{1jT} - y_{1jR}$ and $y_{2jT} - y_{2jR}$. For these differences, calculate the Hodges Lehmann estimate as the location estimate. The robust variance covariance matrix is calculated using the Componentwise rank method explained above (Hettmansperger & McKean, 1998). The robust location estimates are distributed as

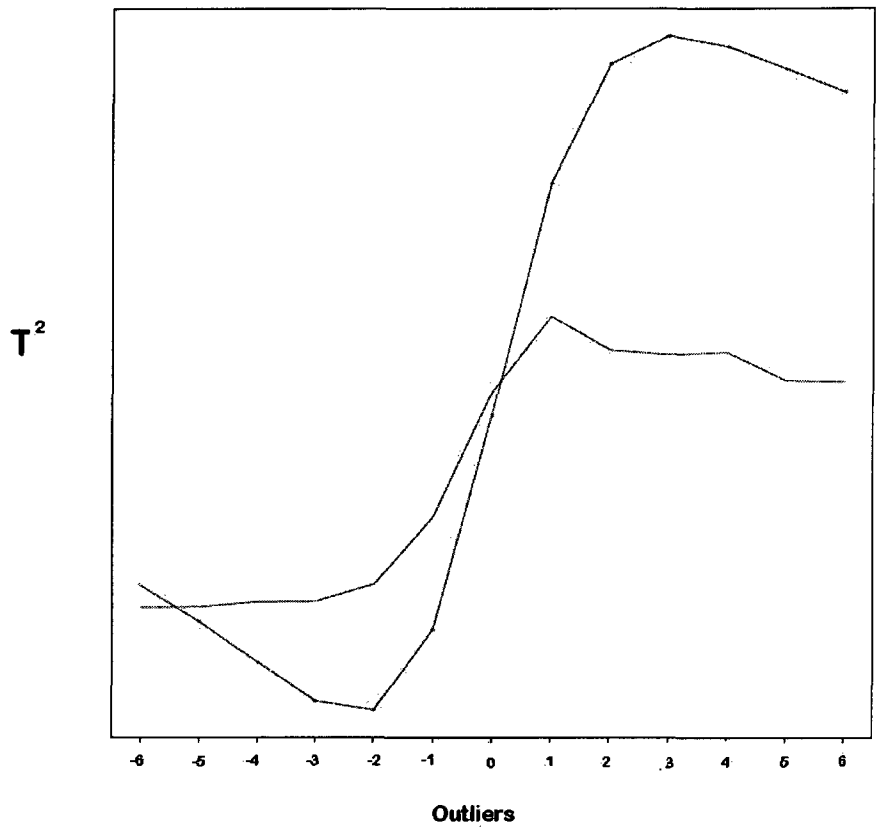
$$\underset{\sim}{\bar{Z}} = \begin{pmatrix} \widehat{\Delta}_{AUC} \\ \widehat{\Delta}_{C\max} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \Delta_{AUC} \\ \Delta_{C\max} \end{pmatrix}, \frac{1}{n}\Sigma \right)$$

where $\widehat{\Delta}_{AUC}$ is the Hodges Lehmann estimate of the vector of differences $Z_{ij_{AUC}}$ for AUC and $\widehat{\Delta}_{Cmax}$ is the Hodges Lehmann estimate of the vector of differences $Z_{ij_{Cmax}}$ for Cmax. The robust spread is $\widehat{\Sigma} = \frac{1}{n}B^{-1}AB^{-1}$ and is computed as explained in the above section. Estimate the Hotelling T^2 robust analog as

$$T^2_{analog} = n\left[\left(\widehat{\Delta}_{AUC} - \Delta_{AUC}\right) \quad \left(\widehat{\Delta}_{C\max} - \Delta_{C\max}\right) \right] \hat{\Sigma}^{-1} \begin{bmatrix} \left(\widehat{\Delta}_{AUC} - \Delta_{AUC}\right) \\ \left(\widehat{\Delta}_{C\max} - \Delta_{C\max}\right) \end{bmatrix}.$$

For this analysis, set Δ_{AUC} and Δ_{Cmax} to zero. To this data, add outliers that are -6σ to 6σ , rerun the above procedure and compute the Hotelling T^2 test statistic. The results of this procedure are summarized in graph 5. As the outliers increase in size, the LS procedure represented by the blue curve is severely affected. The robust procedure represented by the red curve is more stable and is more resistant to outliers. Such a varying T^2 statistic could result in an incorrect conclusion of the hypothesis due to outliers. Clearly, Componentwise rank method performed better as it is less susceptible to outliers. In the following section a simulation analysis comparing validity and power of the LS procedure and the Componentwise rank method is presented.

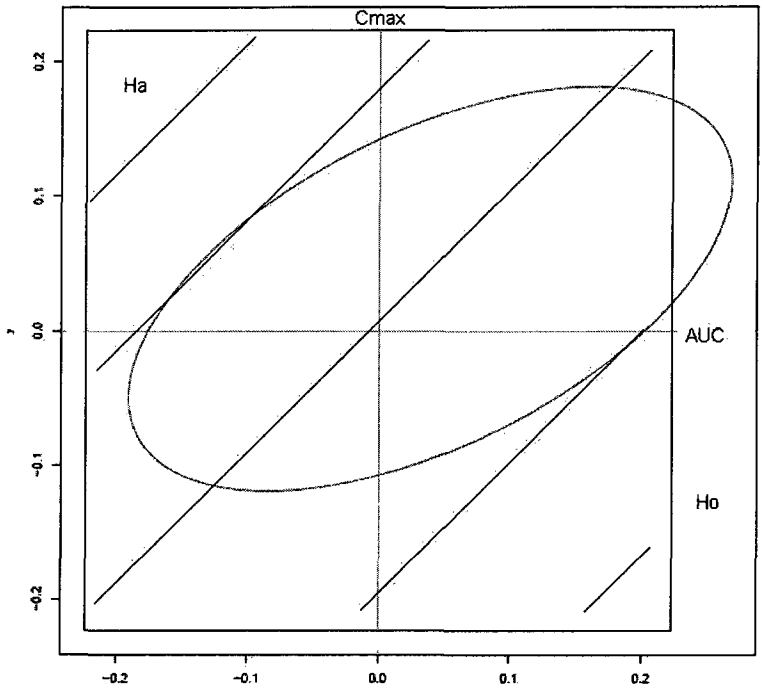
Figure 5: Sensitivity analysis of ABE Hotelling T^2 versus outliers



5.4 Average bioequivalence comparison of level and power

For the simulation study, compare the multivariate LS procedure with the multivariate Componentwise rank method by controlling the true means and variances. The confidence region is an ellipse that is constructed by these means and variances. With the ellipse constructed, count the number of cases where the ellipse falls inside the rejection region.

Figure 6: Plot of the null and alternative regions



5.4.1 Validity

The hypothesis of interest is shown below

$$H_{01}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \leq \ln 0.8 \cup H_{02}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} \geq \ln 1.25,$$

$$H_{A1}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} > \ln 0.8 \cap H_{A2}: \begin{bmatrix} \Delta\mu_{AUC} \\ \Delta\mu_{C\max} \end{bmatrix} < \ln 1.25.$$

Validity is tested at the boundary where the difference in means are either $\mu_T - \mu_R = log_e(0.8)$ or $\mu_T - \mu_R = log_e(1.25)$. It is at these locations that the type I error rate is the highest. Set μ_T = log(0.8) and μ_R = 0 for both AUC and Cmax. The steps for calculating empirical level are as follows :

1. Generate two thousand multivariate data sets of sample size n as shown in table 10. Let the true mean differences be 0.8 for AUC and Cmax. Calculate the difference in response $Y_{1jT} - Y_{1jR}$ such that

$$y_{1jT} - y_{1jR} = \mu_T - \mu_R + e_{1jT} - e_{1jR},$$
$$y_{2jT} - y_{2jR} = \mu_T - \mu_R + e_{2jT} - e_{2jR}.$$

The resulting difference matrix has n rows and two columns. Each column represents the difference in the response for a subject. Errors are the only remaining random effects.

Table 12: Response matrix

Subject	Sequence	AUC difference	Cmax Difference
1	1	$(Y_{1jT} - Y_{1jR})_{AUC}$	$(Y_{1jT} - Y_{1jR})_{CMAX}$
2	1	.	.
.	1	.	.
.	1	.	.
m+1	2	$(Y_{1jT} - Y_{1jR})_{AUC}$	$(Y_{1jT} - Y_{1jR})_{CMAX}$
.	2	.	.
.	2	.	.
j	2	.	.

2. Estimate the LS and the robust (R) estimates of location (one for AUC and the other for Cmax) and the variance covariance matrix from the procedure described in the above section. Construct the confidence region as an ellipse. The ellipse constructed for the LS procedure uses the normality assumption and ellipse constructed for the robust procedure uses the Componentwise rank method (Hettmansperger & McKean, 1998).

3. Sketch the boundary of the rejection region that is a rectangular space bounded by the co-ordinates (log_e(0.8), log_e(0.8)), (log_e(1.25),log_e(0.8)), (log_e(1.25),log_e(1.25))

and (log_e(0.8),log_e(1.25)). To interpret this space, the region inside the rectangle represents the rejection region as shown in the figure 6.

4. Probability of type I error is defined as probability of rejecting H_0 when H_0 is true. Empirical level is calculated by P(type I error) = $\alpha = P_0\ (-\Delta < \mu_T - \mu_R < \Delta)$. This level is estimated by the proportion of cases where the ellipse is contained completely inside the rectangle when in reality it exists at the boundary. Calculate mean squared error (MSE) for the LS and Componentwise rank methods as $MSE = \sqrt{\sum_{i=1}^{P} \frac{\left(\widehat{\Delta\mu_i} - \Delta\mu\right)^2}{P-1}}$.

5.4.2 Power

For calculating the empirical power, set the true mean differences to zero. Power is calculated as a function of the probability of type II error. Estimate the probability of type II error as P_{H_A}(fail to reject H_0). Following are the steps to compute empirical power

1. Generate two thousand multivariate data sets of sample size n as shown in table 10. Let the true mean differences be zero for AUC and Cmax. Calculate the difference in response $Y_{1jT} - Y_{1jR}$ such that

$$y_{1jT} - y_{1jR} = \mu_T - \mu_R + e_{1jT} - e_{1jR},$$
$$y_{2jT} - y_{2jR} = \mu_T - \mu_R + e_{2jT} - e_{2jR}.$$

   The resulting difference matrix has n rows and two columns where each column represents the difference in response for the subject. Errors are the only remaining random effects.

2. Estimate the LS and the robust (R) estimates of location (one for AUC and the other for Cmax) and the variance covariance matrix from the procedure described

in the above section. Construct the confidence region as an ellipse. The ellipse constructed from the LS procedure uses the normality assumption and ellipse constructed from the robust procedure uses the Componentwise rank method (Hettmansperger & McKean, 1998).

3. Sketch the boundary of the rejection region that is a rectangular space bounded by the co-ordinates (log_e(0.8), log_e(0.8)), (log_e(1.25),log_e(0.8)), (log_e(1.25),log_e(1.25)) and (log_e(0.8),log_e(1.25)). To interpret this space, the region inside the rectangle represents the rejection region as shown in the figure 6.

4. Probability of type II error is defined as the probability of failing to reject H_0 when H_A is true. Empirical power is calculated as 1-P(type II error)=1-P_{H_A} ($\mu_T - \mu_R \leq -\Delta$ or $\mu_T - \mu_R \geq \Delta$). The probability of type II error is calculated by the proportion of cases when any part of the ellipse falls outside the rectangle. Calculate MSE for the LS and Componentwise rank methods as $MSE = \sqrt{\sum_{i=1}^{P} \frac{(\widehat{\Delta\mu_i} - \Delta\mu)^2}{P-1}}$.

Results of this simulation procedure is discussed in the next section.

5.5 Comparison of level and power of LS and robust ABE

In each of these cases, with a sample size of twenty subjects, the simulations were run two thousand times. From these two thousand datasets, the level and power are estimated for the following cases.

5.5.1 LS and HL estimators plot with one 1.5σ outlier

Figure A.7 plots the graph when one outlier is added into the data. The first two plots are cases with no outliers and the bottom two plots show outliers. The robust procedure looks efficient with a mild outlier. The LS procedure performs fairly well and the two

procedures have comparable MSE. Both the procedures have similar significance level and power while the LS procedure has a mildly conservative level.

5.5.2 LS and HL estimators plot with two 1.5σ outliers

Figure A.8 plots the graph when two outliers are added into the data. The robust procedure is resistant to the outliers. However, the LS procedure is significantly affected by the outliers and the shape of the ellipse generated is different from LS procedure without outliers. The significance level of the robust procedure is very close to 5% unlike the LS procedure. Since the validity of the test of LS procedure is severely affected, the LS procedure produces incorrect conclusions in this scenario.

5.5.3 LS and HL estimators plot with two 3σ outliers

From the figure A.9, the robust procedure is moderately affected by the two 3σ outliers. However the LS procedure is now invalid as the significance level of the test is severely affected by outliers.

From the table B.15, it is seen that with no outliers, LS is the best procedure. But even with small outliers, the LS procedure is compromised and its validity is suspect. The robust procedure is more stable in the presence of outliers even with small sample sizes.

CHAPTER VI

CONCLUSIONS AND SCOPE FOR FURTHER RESEARCH

Bioequivalence analysis is used to compare the rate and extent of the drug absorbed by an NDA (test drug formulation) with an RLD (reference drug formulation). The FDA (2001) suggested AUC and Cmax as important pharmaco-kinetic parameters to be compared for equivalence analysis. Thus average, population and individual bioequivalence hypotheses procedures were proposed by FDA (2001).

6.1 Comparison of LS ABE with robust ABE

Average bioequivalence (ABE) was suggested to test the equivalence of the location of an NDA with an RLD using AUC and Cmax. A two one-sided hypothesis was directed (FDA, 2001) for ABE analysis. The reasons for the log-transformation of the pharmaco-kinetic parameters are explained in the introduction chapter. In ABE hypothesis, emphasis was laid on testing whether the difference in location of the test and the reference drugs were bound within the acceptable therapeutic difference ($\pm \log 1.25$). Least squared procedures tested the univariate log-transformed pharmacokinetic parameters. However, the test statistics using LS procedures were not resistant to outliers. Further, drugs which had high variability were not accounted for, in the hypothesis. Since small samples are generally used in phase I clinical trials, univariate ABE may be incomplete.

We suggested a multivariate two one-sided hypothesis using both AUC and Cmax for ABE analysis. In order to counter outliers, Componentwise rank method, a robust procedure was proposed. With the multivariate procedure, we constructed the confidence region shaped as an ellipse. The rectangular shaped rejection region (FDA, 2001) was also

examined. Sensitivity analyses were conducted on the two one-sided multivariate LS procedure and on the two one-sided multivariate Componentwise rank method. Hotelling T^2 test statistic was computed for both the LS and robust procedures for data with increasing outliers. Simulation analyses were performed to compare validity and power.

As the outliers increased in size, the sensitivity analyses indicated that the LS procedure was severely affected. The T^2 test statistic showed high variability in the presence of outliers that could lead to incorrect conclusions about the hypothesis. The Componentwise rank method was more robust and resistant to outliers and gave consistent T^2 statistic values. Our findings were summarized in table 13.

Table 13: Bioequivalence findings

Case	Variance	Outliers	Best	Worst
ABE	Small	$\leq 3\sigma$	LS	R
		$> 3\sigma$	R	LS
	Large	$\leq 3\sigma$	LS	R
		$> 3\sigma$	R	LS
PBE Large Sample	Small	$\leq 3\sigma$	LS, Gini	MAD, IQR
		$> 3\sigma$	S_n, Q_n	LS, Gini
	Large	$\leq 3\sigma$	LS, Gini, S_n, Q_n	
		$> 3\sigma$	Q_n	LS, Gini
PBE Small Sample	Small	$\leq 3\sigma$	LS, Gini	IQR
		$> 3\sigma$	IQR, Gini	LS
	Large	$\leq 3\sigma$	IQR	LS
		$> 3\sigma$	IQR	LS

The the simulation analyses of small sample multivariate ABE with no outliers showed that both the LS and robust procedures were comparable when testing at 5% significance. The LS procedure had a marginally higher power than the robust procedure. The MSE, however, was equivalent for the two.

The test of validity and power of the LS procedure when compared with the robust

procedure with mild outliers had a different result. In the presence of small outliers (1.5σ outliers), the validity of the LS procedure was severely affected. The level of the LS procedure was close to 10%. Since the level of the LS procedure was not conservative, the power of the test is inconclusive. Contrarily, the significance level of the robust procedure was close to 5%. Additionally, the MSE of the LS procedure was much higher than the robust procedure. These show that the robust procedure was more efficient in testing the hypothesis.

With 3σ outliers in the data, the LS procedure was severely affected. The LS procedure had a higher level while the robust procedure had a more conservative level. Also, the robust procedure had a much smaller MSE than the LS procedure.

The above leads to the conclusion that the Componentwise rank method on small sample ABE analysis is comparable to the LS procedure when the data has no outliers. Outliers severely affect the validity, power and MSE of the LS procedure while the robust procedure is much more conservative and resistant to the influence of outliers.

6.2 LSCF versus robust procedures for large sample PBE

PBE is assessed to prove bioequivalence of a to-be-marketed formulation when a major formulation change has been made prior to the approval of a new drug. It is tested on patients who would be taking the drug formulation for the first time. Population bioequivalence is considered only after average bioequivalence is approved. Chinchilli et al. (1996) proposed a two sequence, four period cross-over design which the FDA has recommended for PBE (and IBE) analysis.

Analysis of population bioequivalence focused on the estimation of the mean difference and the total variance of the log transformed BA measures of the two drug formulations. Unbiased estimators of these parameters were generated by the method of moments

(Chinchilli & Esinhart, 1996). Following the estimation of the mean difference and the variances, a 95% upper confidence bound for a linearized form of the population BE criterion was obtained. Population BE was established for a log-transformed BA measure when the 95% upper confidence bound for this linearized criterion was less than or equal to zero (FDA, 2001).

One of the issues discussed previously was the presence and impact of outliers. The independence criteria required for Cornish Fisher's expansion may be violated in the present procedure. To examine this, five bootstrap procedures that estimate the upper confidence bound of the linearized criterion was suggested. The bootstrap procedures were more discriminating when the sample size was larger than sixty. Thus, alternative procedures to large sample PBE analysis were proposed.

The robust procedure which used Qn to estimate the variance was least sensitive to outliers. As the outliers increased in size, the LS procedures (LSCF and Gini) were severely affected. The test statistic η showed high variability. The large sample PBE simulation results were summarized in table 13.

The bootstrap simulations showed that, with small outliers in the test drug and small variability in the data, the LS procedures (LSCF and Gini) had the largest power, smallest MSE and a significance level close to 5%. However, small outliers in the test drug and large variability in the data showed different results. In this context, the robust procedures were comparable to the LS procedures in significance level and power.

Alternately, with smaller outliers in the reference drug, the robust procedures performed much better than the LS procedures. The LS procedures were most compromised when the estimated significance level was 15%. The robust procedures however, had a significance level close to 5%. MSE of the LS procedure was also higher than the robust procedure.

With larger outliers in the reference drug, the LS procedures completely failed. The

significance level approached 20%. Such a large number renders meaningless power. The validity of the robust procedure with large outliers was approximately 5%. The robust procedures were also consistent with low MSE.

To conclude, for samples of size larger than sixty, smaller outliers in the test drug formulation do not severely affect the hypothesis. However, reference drug outliers significantly affect the overall result. Robust procedures handle outliers better and have consistent significance levels with comparable powers and lower MSE. Finally, the outlier occurrences in the test drug formulation gives differed results than outlier occurrences in the reference drug formulation.

6.3 LSCF versus robust procedures for small sample PBE

Phase I of a clinical trial typically used samples of size twenty or less. With such small sample sizes, the robust bootstrap PBE procedure did not give consistent results. It was proposed to use the CF expansion using closed forms of Gini and IQR to estimate the variance. For the robust location, we suggested the use of median. The sensitivity analysis clearly showed that the procedure using IQR for the variance estimate was more resistant to outliers. Since the median was used for the Gini procedure, Gini gave marginally better results than LSCF.

The two LS procedures, LSCF and Gini, were similar. However, due to it's robust location, Gini proved to be a better procedure with conservative level when the data had outliers. Since IQR was less conservative in significance level, LSCF and Gini were better procedures when the data had modest outliers ($< 3\sigma$).

However, for data with larger outliers ($> 3\sigma$), the LS procedures had a much larger significance level and a high MSE. Although IQR was stable, it was less conservative with low power. It was therefore concluded that all the three procedures failed when outliers

were larger than 3σ. Further research is needed to resolve the effect of outliers on small sample PBE.

6.4 Scope for further research

Given the above conclusions, there is a need to conduct additional research to address several issues. For the large sample population BE situation, robust bootstrap procedure was used. Investigation into why the robust procedures gave inconsistent results for small sample PBE analysis is needed.

All the results were based on normally distributed pharmaco-kinetic parameters. The implication of the present designs on non-normal unsymmetric data needs to be examined. The proposed bootstrap and the LS procedures should be tested against different distributions of the pharmacokinetic parameters.

The scope of multivariate analysis for PBE should be expanded. EMEA (2001) has already suggested the use of T_{max} using Wilcoxon scores to test differences in time to reach maximum concentration of drug in plasma. One can readily incorporate AUC, Cmax, Tmax into the proposed univariate model and with the definition of the underlying distribution (and covariance structure), test for PBE.

For small sample PBE, Gini and IQR were used as estimates of dispersion. Clearly the outlier analysis shows that these are not exhaustive and do not perform well in the presence of outliers. Additional research using MAD, Q_n, S_n and other robust estimators to compare the LS Cornish Fisher's procedure to the robust Cornish Fisher's procedures for small samples is needed.

Finally, for average bioequivalence, further work is needed to compare the effect of ABE on PBE. Multivariate procedures tend to have better power than the univariate procedure and the scope of such a usage should be reviewed for more than a bivariate case.

APPENDIX A

GRAPHS

The large sample PBE analysis is the plot of power versus the sample size ranging from 100 to 200 subjects. The bottom two graphs are the significance level (α) plotted against sample size to study the effect of outliers on the data.

With small sample PBE analysis, plots of level and power against the fixed samples but varying outliers are presented. This plot depicts the effect of test drug outliers compared to reference drug outliers. There is also an MSE plotted against the same horizontal axis.

For ABE analysis, four ellipses are plotted along with their rejection regions. The first two plots are the plots of the ellipse with LS and robust procedures. The bottom two plots of the ellipses depict the extent of change in the location, shape and size of the ellipse after adding outliers.

Figure A.1: γ and α with small Test outliers

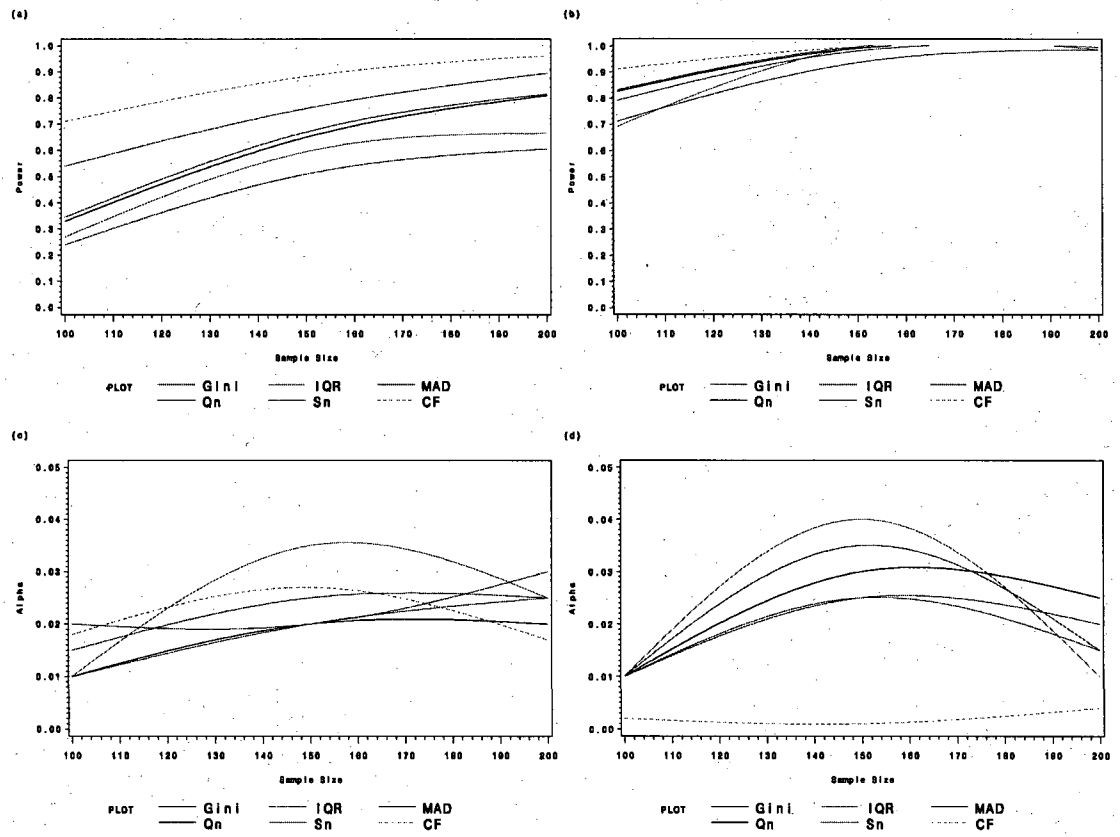


Figure A.2: γ and α with small Reference outliers

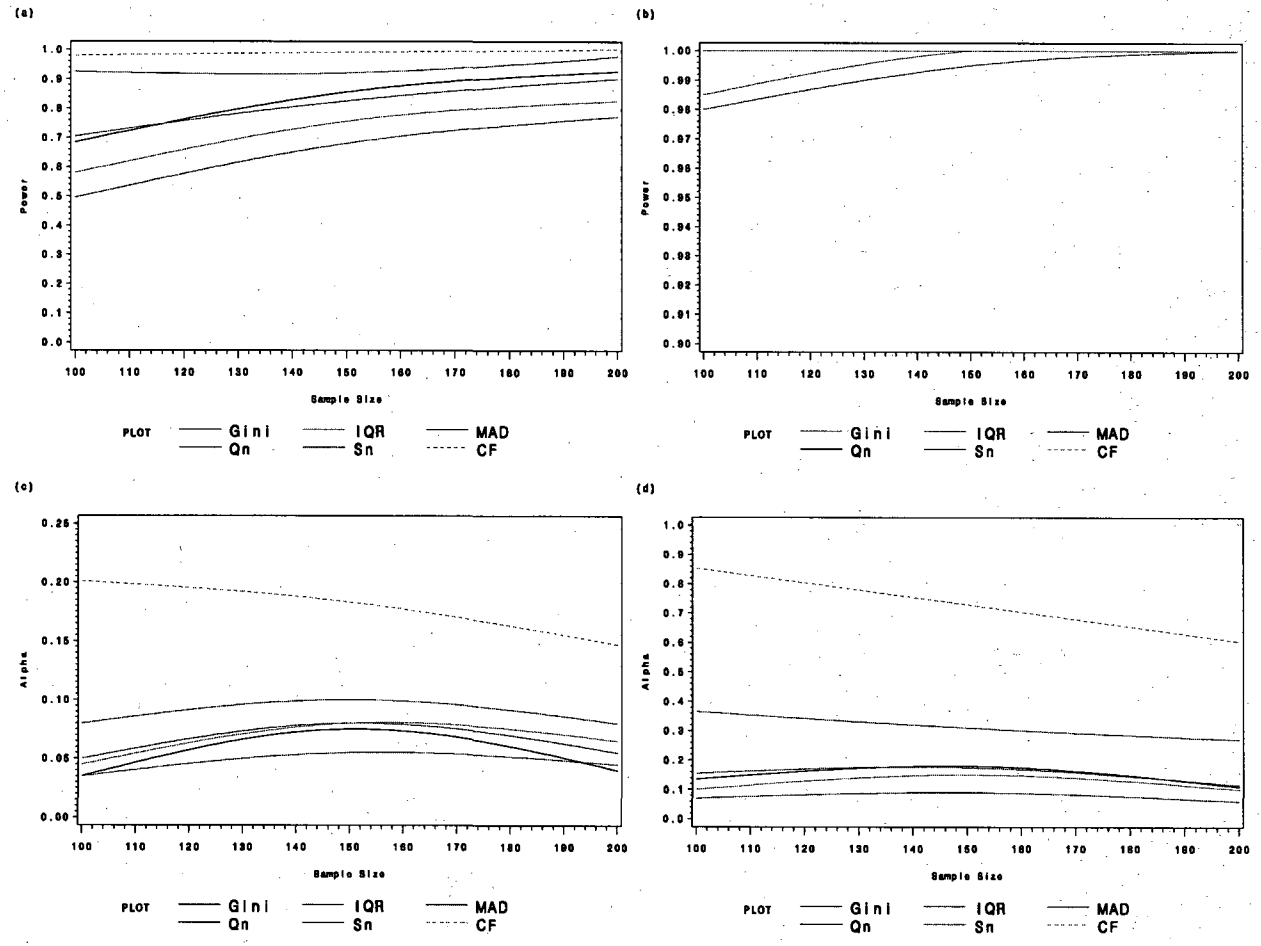


Figure A.3: γ and α with large Test outliers

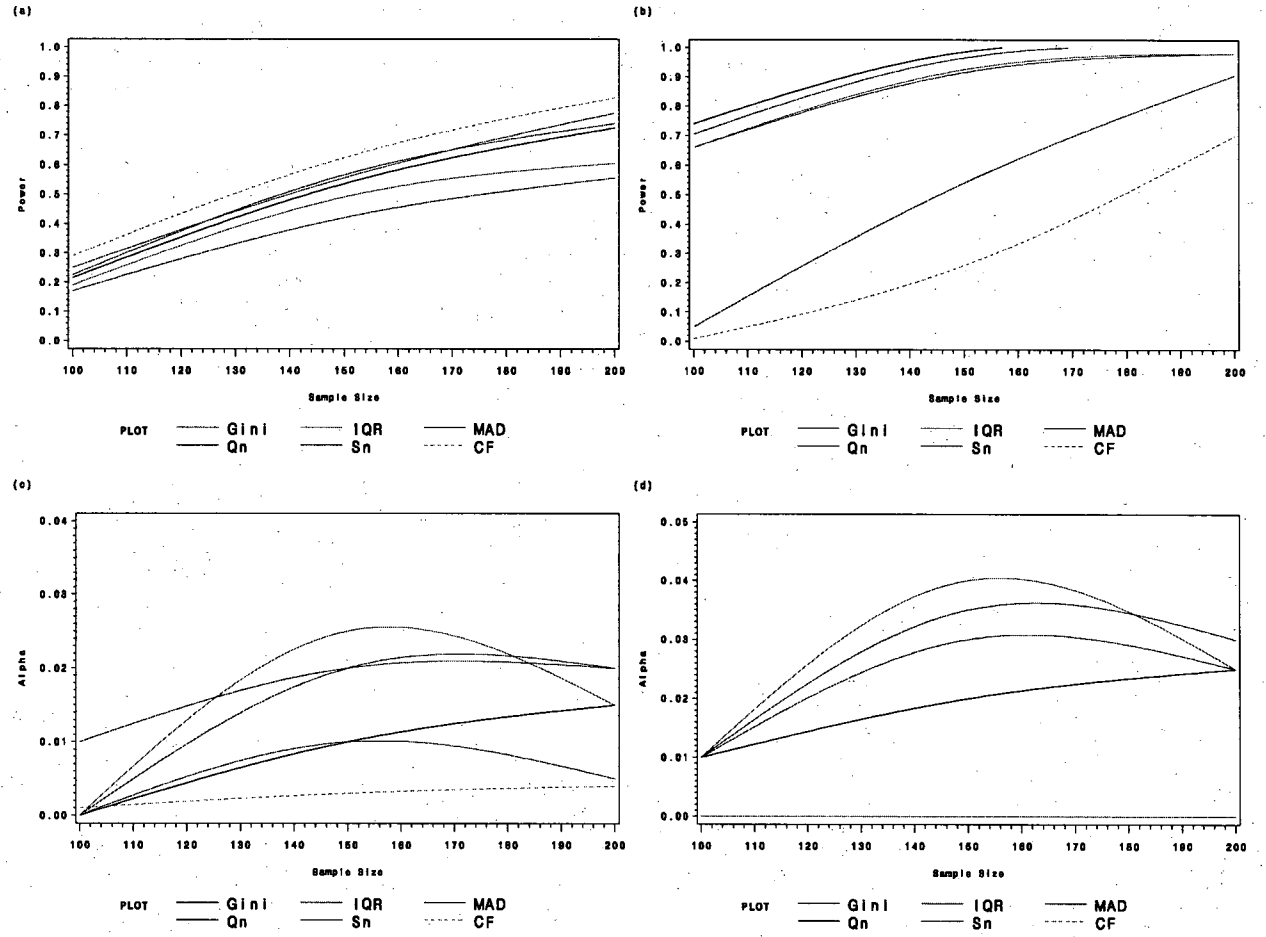


Figure A.4: γ and α with large Reference outliers

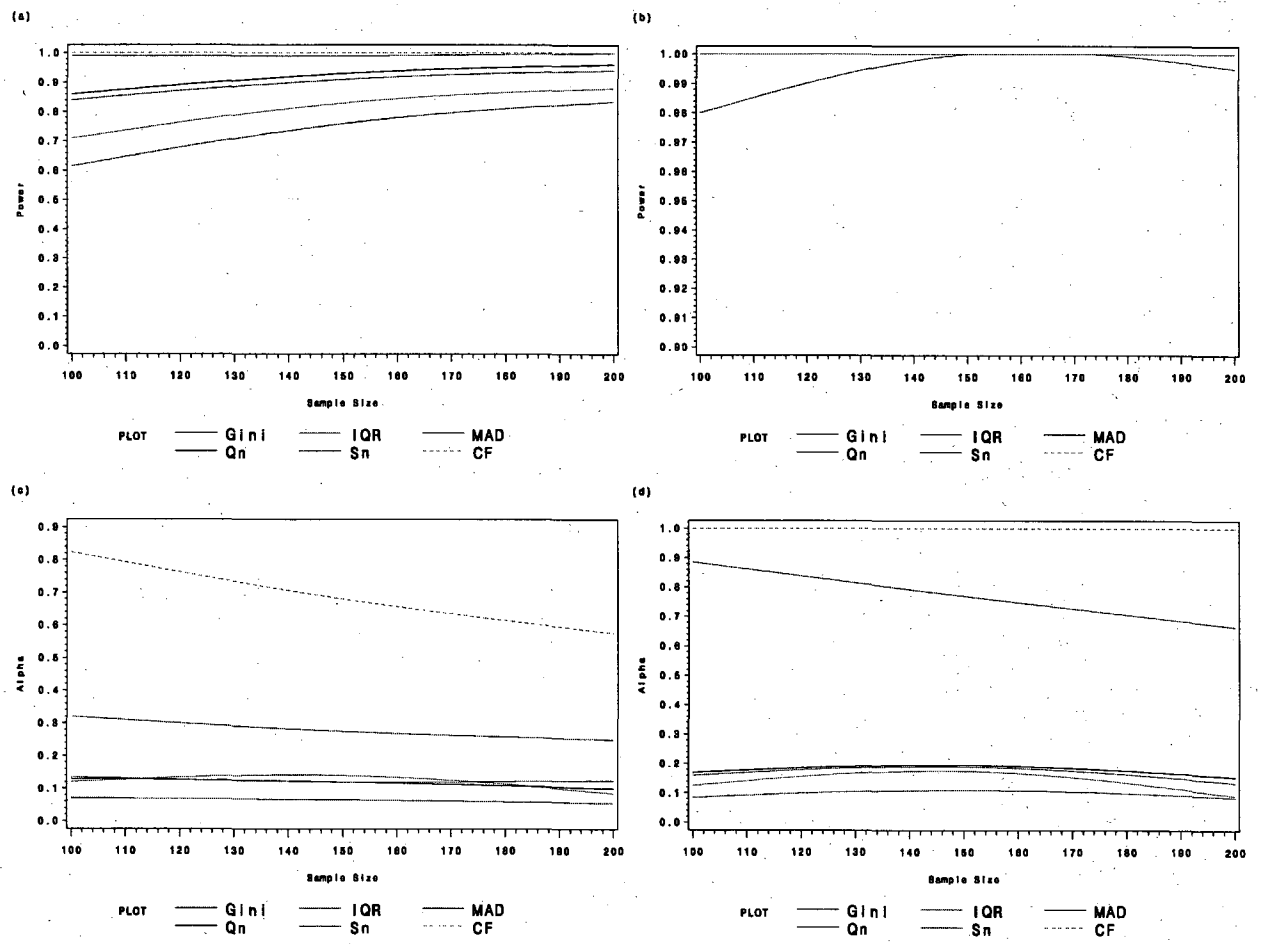


Figure A.5: Power and α with LS, Gini and IQR

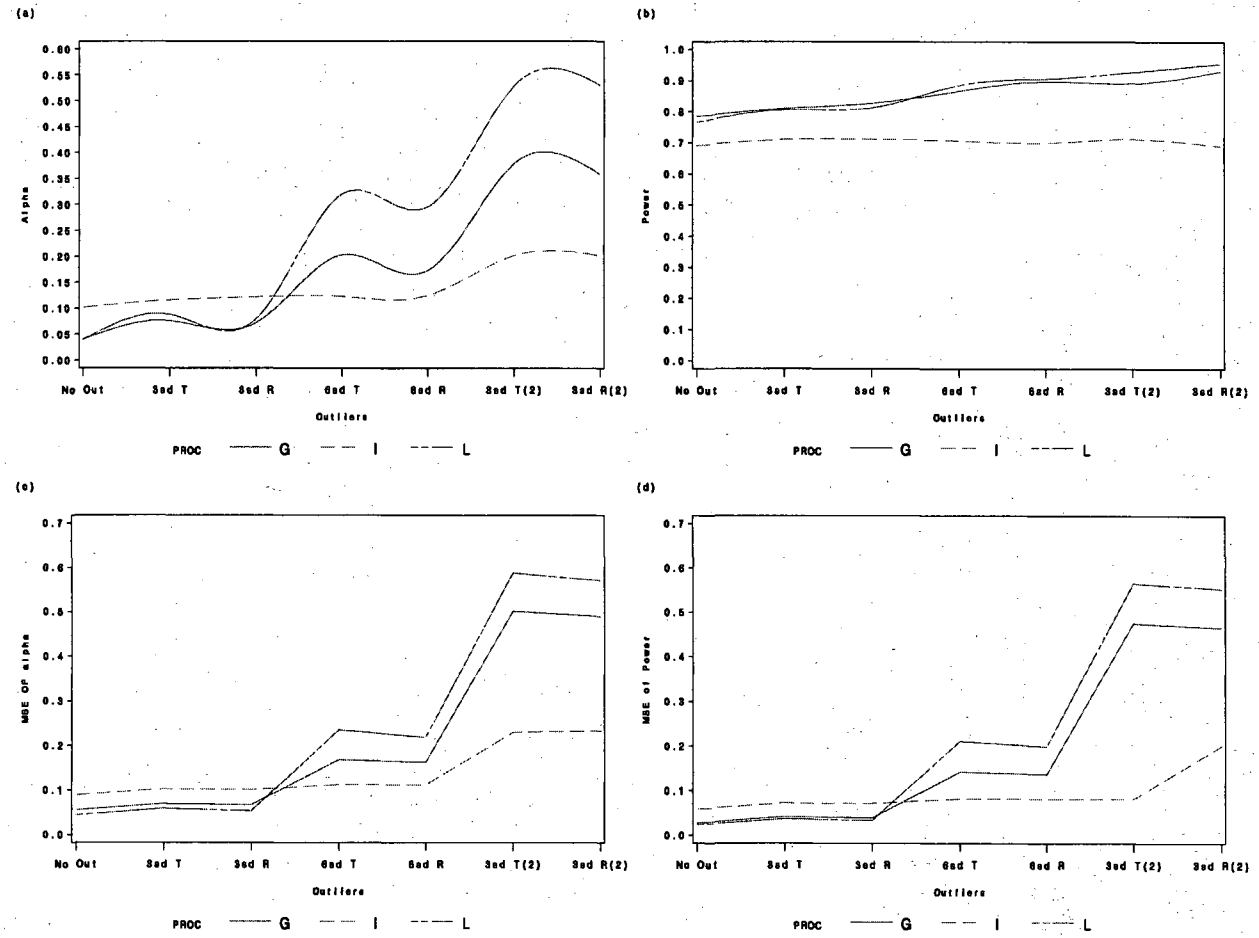


Figure A.6: Incremental outliers on Power and α with LS, Gini and IQR

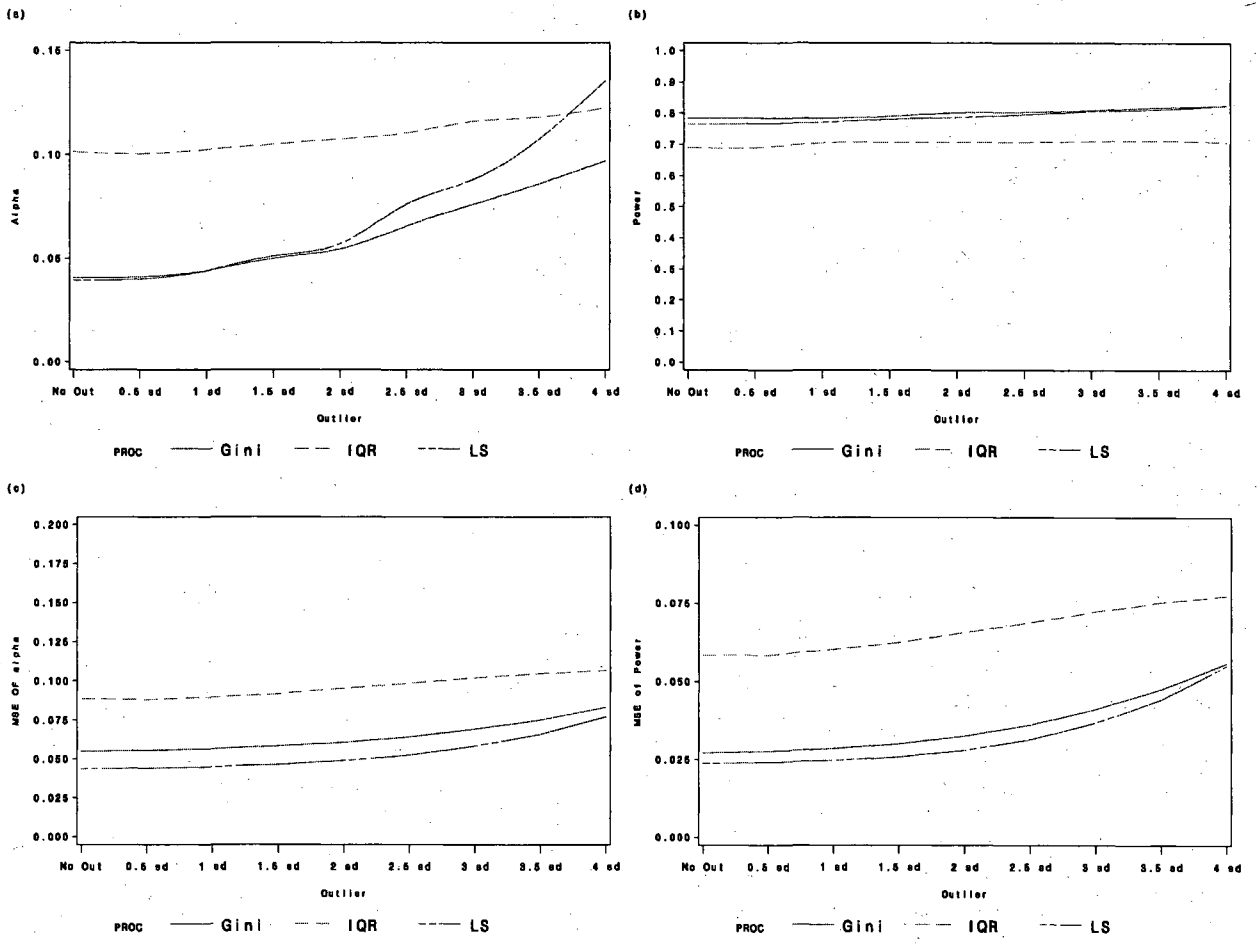


Figure A.7: LS and Hodges Lehmann estimators plot w/ one 1.5σ outlier

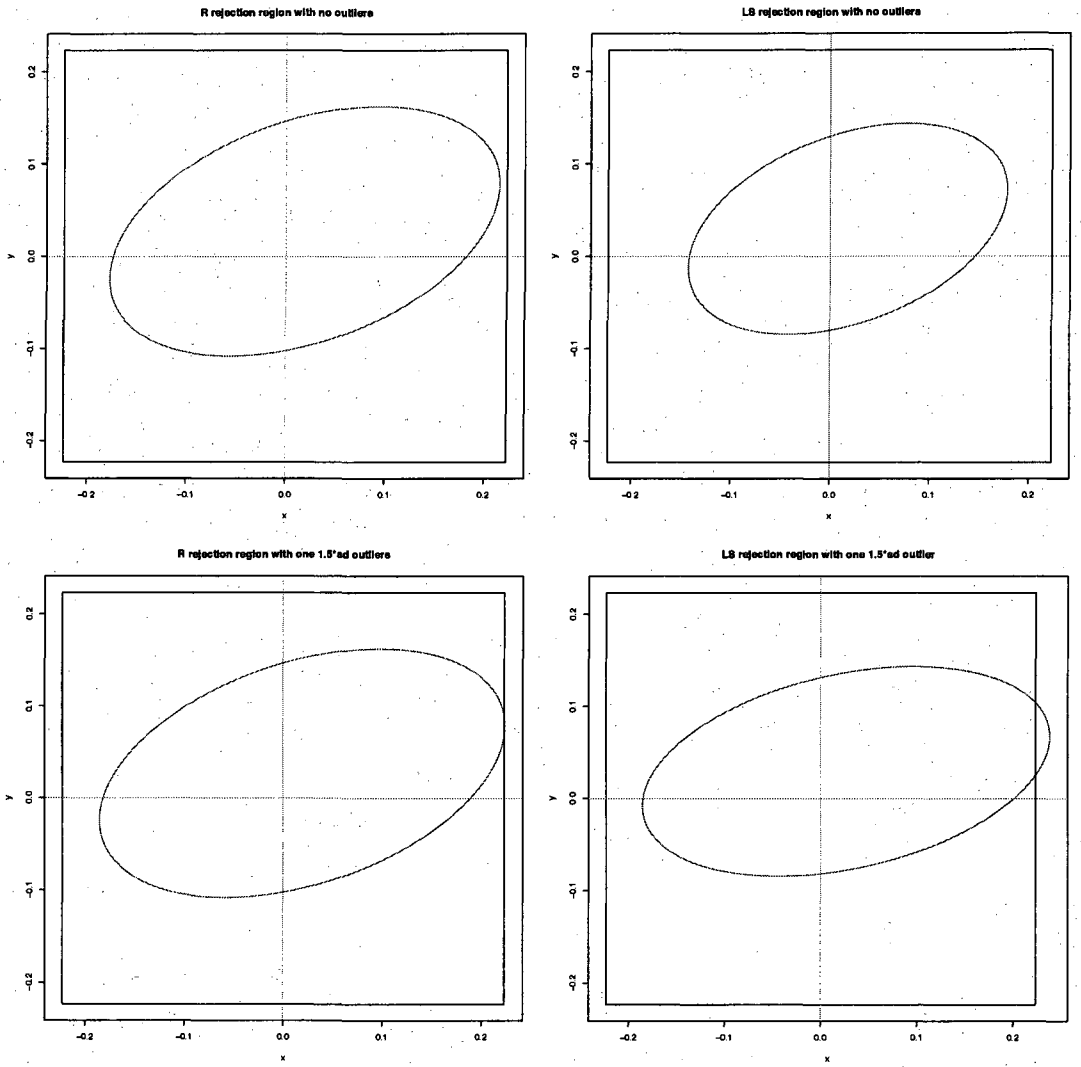


Figure A.8: LS and Hodges Lehmann estimators plot w/ two 1.5σ outliers

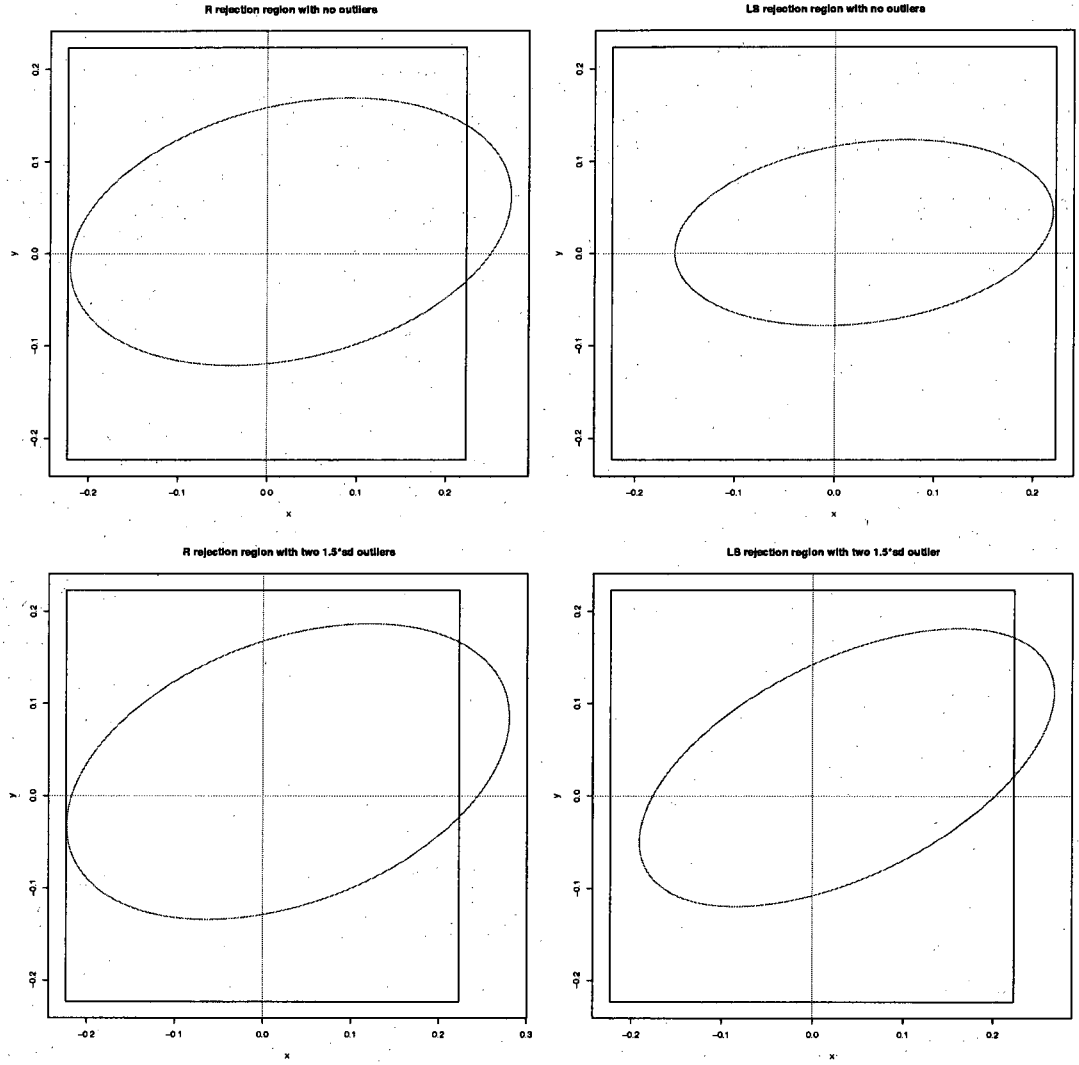
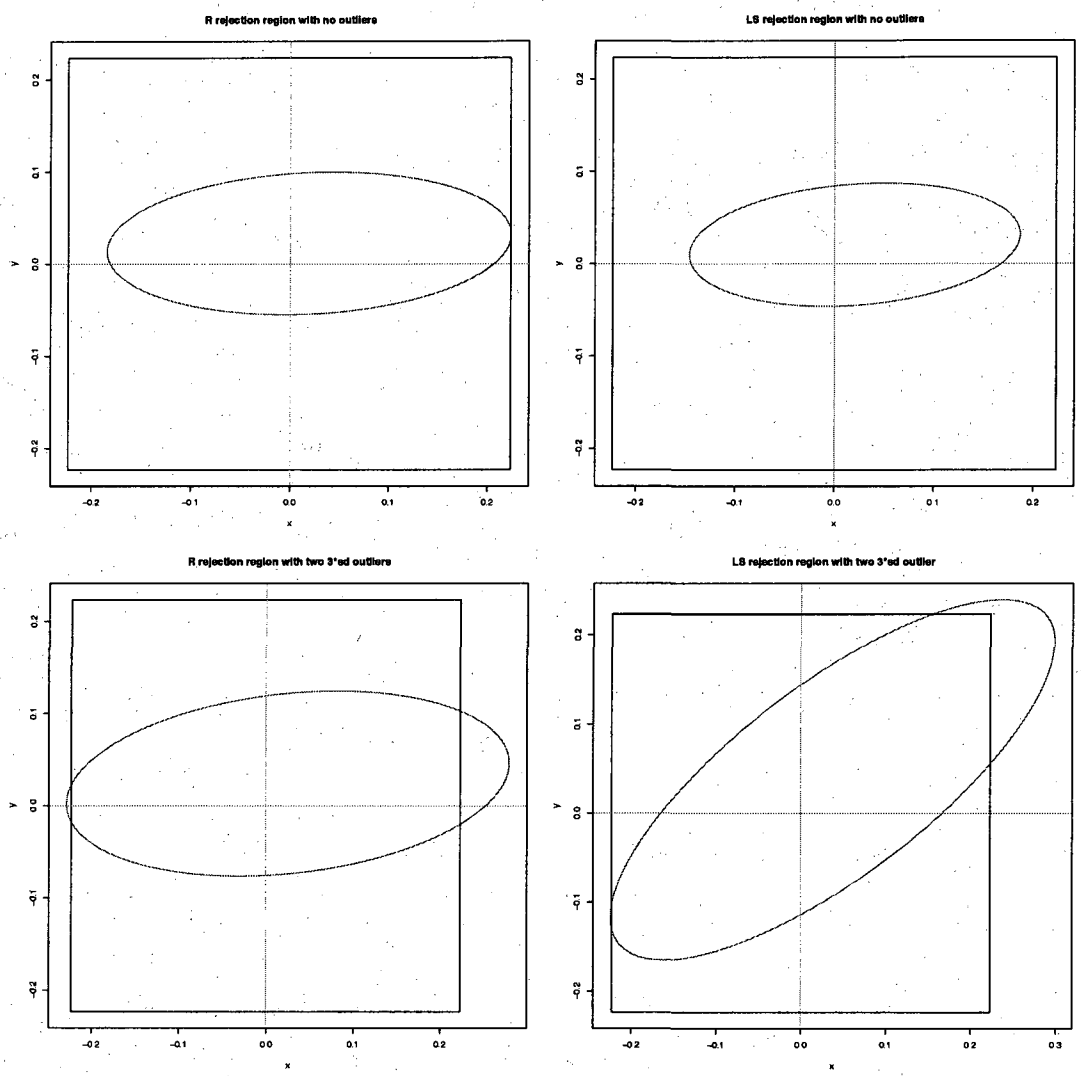


Figure A.9: LS and Hodges Lehmann estimators plot w/ two 3σ outliers



APPENDIX B

TABLES

2.1 PBE with no outliers

The below table compares the large and small sample PBE with no outliers for LSCF with Gini, IQR, MAD, Qn and Sn procedures.

Table B.1: Table of large sample PBE with no outliers

Sample Size	Method	η	η_{95}	Conclusion
100	Gini	-0.37467	-0.15984	Reject H_0
	Interquartile	-0.42121	-0.11836	Reject H_0
	LSCF	-0.30514	-0.1527	Reject H_0
	MAD	-0.50416	-0.13868	Reject H_0
	Qn	-0.34698	-0.08866	Reject H_0
	Sn	-0.45055	-0.17377	Reject H_0
150	Gini	-0.18168	-0.03521	Reject H_0
	Interquartile	-0.31484	-0.06096	Reject H_0
	LSCF	-0.11991	0.001625	Fail to Reject H_0
	MAD	-0.26597	-0.00973	Reject H_0
	Qn	-0.23495	-0.03431	Reject H_0
	Sn	-0.20858	-0.02522	Reject H_0
200	Gini	-0.35981	-0.18991	Reject H_0
	Interquartile	-0.40267	-0.17724	Reject H_0
	LSCF	-0.24655	-0.12594	Reject H_0
	MAD	-0.4557	-0.20459	Reject H_0
	Qn	-0.3844	-0.19065	Reject H_0
	Sn	-0.42334	-0.20866	Reject H_0

2.2 Large sample PBE power and level with outliers

- Case (a) small variance: To estimate power set $\sigma^2_{BT} = \sigma^2_{BR} = 0.15$ and $\sigma^2_{WT} = \sigma^2_{WR} = 0.15$. δ=0.5 which sets the true η to -0.27344. Small outliers are 3σ outliers

Table B.2: Table of small sample PBE with no outliers

Sample		η	$\eta_{95\%}$	Conclusion
20	LSCF	-0.08958	0.35296	Fail to Reject H_0
	IQR	-0.44122	-0.05353	Reject H_0
	Gini	-0.19356	0.18934	Fail to Reject H_0
16	LSCF	-0.3751	0.001043	Reject H_0
	IQR	-0.38012	-0.02661	Fail to Reject H_0
	Gini	-0.40715	-0.0308	Fail to Reject H_0

added to the test or reference drugs and large outliers are 6σ outliers added to the test or reference drugs.

- Case (b) large variance: To estimate power set $\sigma^2_{BT} = \sigma^2_{BR} = 0.25$ and $\sigma^2_{WT} = \sigma^2_{WR} = 0.25$. δ=0.5 which sets the true η to -0.6224 Small outliers are 3σ outliers added to the test or reference drugs and large outliers are 6σ outliers added to the test or reference drugs.

- Case (c) small variance: To estimate α set $\sigma^2_{BT} = \sigma^2_{BR} = 0.15$ and $\sigma^2_{WT} = \sigma^2_{WR} = 0.15$. δ=0.7234969 which sets the true η to 0. Small outliers are 3σ outliers added to the test or reference drugs and large outliers are 6σ outliers added to the test or reference drugs.

- Case (d) large variance: To estimate α set $\sigma^2_{BT} = \sigma^2_{BR} = 0.25$ and $\sigma^2_{WT} = \sigma^2_{WR} = 0.25$. δ=1.320984 which sets the true η to 0. Small outliers are 3σ outliers added to the test or reference drugs and large outliers are 6σ outliers added to the test or reference drugs.

Table B.3: Power with small test drug outliers

N	Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.54	0.27	0.24	0.33	0.345	0.71
150	0.76	0.595	0.51	0.65	0.67	0.883
200	0.895	0.665	0.605	0.81	0.815	0.962
b 100	0.83	0.69	0.71	0.825	0.79	0.909
150	0.995	0.995	0.935	0.99	0.98	0.995
200	1	0.985	0.985	1	0.995	1
N	MSE of Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.0171	0.02540	0.0261	0.0220	0.0204	0.0126
150	0.0118	0.01935	0.0185	0.0136	0.0136	0.0093
200	0.0077	0.0140	0.0143	0.0099	0.0107	0.0065
b 100	0.3331	0.2889	0.2952	0.3014	0.2499	0.3447
150	0.1862	0.2184	0.2013	0.1699	0.1537	0.1963
200	0.1262	0.172	0.1783	0.1312	0.1329	0.1324

Table B.4: Level α with small test outliers

N	Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.015	0.01	0.01	0.01	0.02	0.018
150	0.025	0.035	0.02	0.02	0.02	0.027
200	0.025	0.025	0.025	0.02	0.03	0.017
d 100	0.01	0.01	0.01	0.01	0.01	0.002
150	0.025	0.04	0.025	0.03	0.035	0.001
200	0.015	0.01	0.02	0.025	0.015	0.004
N	MSE of Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.0228	0.0301	0.0315	0.0284	0.0257	0.0164
150	0.0157	0.0230	0.0223	0.0177	0.0174	0.0118
200	0.0099	0.0164	0.0167	0.01229	0.0131	0.0084
d 100	0.5629	0.4575	0.4828	0.5211	0.4407	0.6096
150	0.315	0.3192	0.3127	0.29044	0.2643	0.3382
200	0.2024	0.252	0.2586	0.2089	0.2091	0.2252

Table B.5: Power with small reference drug outliers

N	Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.925	0.58	0.495	0.685	0.705	0.98
150	0.92	0.755	0.68	0.855	0.825	0.99
200	0.975	0.825	0.77	0.925	0.9	1
b 100	1	0.985	0.98	1	1	1
150	1	1	0.995	1	1	1
200	1	1	1	1	1	1
N	MSE of Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.0157	0.0308	0.0263	0.0150	0.0225	0.0198
150	0.0128	0.0275	0.0214	0.0122	0.0175	0.0123
200	0.0083	0.0150	0.0144	0.0085	0.0120	0.0076
b 100	0.8365	0.5734	0.4536	0.3467	0.4978	1.4128
150	0.4571	0.4298	0.2988	0.2235	0.3070	0.6942
200	0.2326	0.2035	0.1871	0.1185	0.1608	0.3882

Table B.6: Level α with reference drug outliers

N	Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.08	0.045	0.035	0.035	0.05	0.201
150	0.1	0.08	0.055	0.075	0.08	0.183
200	0.08	0.065	0.045	0.04	0.055	0.147
d 100	0.365	0.1	0.07	0.135	0.155	0.853
150	0.31	0.15	0.09	0.18	0.175	0.729
200	0.27	0.1	0.06	0.11	0.115	0.604
N	MSE of Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.0206	0.0362	0.0312	0.0196	0.0274	0.0244
150	0.0168	0.0323	0.0257	0.0160	0.0218	0.0152
200	0.0109	0.0181	0.0172	0.0109	0.0148	0.0095
d 100	1.0836	0.7816	0.6472	0.5397	0.7158	1.8724
150	0.6314	0.5879	0.4376	0.3591	0.4555	0.9219
200	0.3377	0.2901	0.2683	0.1993	0.2493	0.5272

Table B.7: Power with large test drug outliers

N	Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.25	0.19	0.17	0.215	0.225	0.29
150	0.555	0.49	0.42	0.535	0.565	0.623
200	0.775	0.605	0.555	0.725	0.74	0.827
b 100	0.05	0.66	0.66	0.74	0.705	0.009
150	0.54	0.925	0.915	0.985	0.965	0.26
200	0.905	0.98	0.98	1	1	0.7
N	MSE of Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.0302	0.0284	0.0300	0.0302	0.0251	0.0309
150	0.0171	0.0197	0.0201	0.0168	0.0154	0.0177
200	0.0105	0.0149	0.0156	0.0118	0.0119	0.0119
b 100	1.4039	0.3070	0.3120	0.3652	0.2864	3.0376
150	0.585	0.221	0.2075	0.1923	0.1652	1.437
200	0.349	0.1780	0.1832	0.1471	0.1426	0.8689

Table B.8: Level α with large test drug outliers

N	Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0	0	0	0	0.01	0.001
150	0.01	0.025	0.02	0.01	0.02	0.003
200	0.005	0.015	0.02	0.015	0.02	0.004
d 100	0	0.01	0.01	0.01	0.01	0
150	0	0.04	0.03	0.02	0.035	0
200	0	0.025	0.025	0.025	0.03	0
N	MSE of Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.0385	0.0346	0.0368	0.0384	0.0319	0.0404
150	0.0218	0.0235	0.0246	0.0215	0.0196	0.0228
200	0.0133	0.0180	0.0186	0.0148	0.0147	0.0152
d 100	1.805	0.4956	0.518	0.6118	0.5033	4.3558
150	0.7697	0.3308	0.3248	0.3249	0.2861	2.0568
200	0.4502	0.2607	0.2686	0.2686	0.2249	1.2338

Table B.9: Power with large reference drug outliers

N	Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.99	0.71	0.615	0.86	0.84	1
150	0.99	0.83	0.76	0.93	0.91	1
200	1	0.88	0.835	0.96	0.94	1
b 100	1	0.98	0.98	1	1	1
150	1	1	1	1	1	1
200	1	0.995	0.995	1	1	1
N	MSE of Power					
	Gini	IQR	MAD	Qn	Sn	LSCF
a 100	0.0639	0.0554	0.0436	0.0321	0.0455	0.1074
150	0.0360	0.0370	0.0276	0.0209	0.0284	0.0530
200	0.0207	0.0187	0.0176	0.0123	0.0169	0.0299
b 100	7.5060	0.7479	0.5372	0.6150	0.7340	18.0897
150	3.1595	0.4642	0.3216	0.3260	0.3957	8.5661
200	1.5612	0.2139	0.1953	0.1629	0.1924	4.8756

Table B.10: Level α with large reference drug outliers

N	Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.32	0.12	0.07	0.13	0.135	0.823
150	0.275	0.14	0.065	0.12	0.12	0.68
200	0.25	0.085	0.055	0.1	0.125	0.576
d 100	0.885	0.125	0.085	0.17	0.16	1
150	0.77	0.175	0.11	0.195	0.19	1
200	0.665	0.09	0.085	0.155	0.135	1
N	MSE of Alpha					
	Gini	IQR	MAD	Qn	Sn	LSCF
c 100	0.0734	0.0633	0.0511	0.03977	0.0538	0.1238
150	0.0419	0.0431	0.0330	0.0260	0.0338	0.0612
200	0.0250	0.0223	0.0209	0.0156	0.0205	0.03492
d 100	8.0807	0.9725	0.755	0.8637	1	21.021
150	3.4794	0.6296	0.4651	0.4751	0.5546	8.7312
200	1.7568	0.3093	0.2867	0.2569	0.2916	5.0315

2.3 Small sample PBE power and level

For the setting of a small sample PBE analysis with or without outliers, the outliers vary from zero to six sigma.

- No Out : implies that the data with no outliers were considered
- 3sigma(test) : implies that one subject's Test reading was having an outlier of 3 standard deviations
- 3sigma(Ref) : implies that one subject's Reference reading was having an outlier of 3 standard deviations. This was conducted to see if the location of the outliers affected the power or type I error of the test.
- 6sigma(test) : implies that one subject's Test reading was having an outlier of 6 standard deviations.
- 6sigma(Ref) : implies that one subject's Reference reading was having an outlier of 6 standard deviations.
- 2-3sigma(test): implies that two subject's Test reading were having outliers of 3 standard deviations.
- 2-3sigma(Ref) : implies that two subject's Reference reading were having outliers of 3 standard deviations.

Table B.11: Small sample α with LSCF, Gini and IQR

N=20	Outliers	Procedure	η	$\eta_{upperlimit}$	MSE_{ETA}	α
	None	LS	0.01324	0.377943	0.04378	0.0395
	None	IQR	-0.00516	0.382656	0.088684	0.1015
	None	Gini	-0.00019	0.384766	0.055147	0.0405
	3sigma(test)	LS	-0.06623	0.326196	0.058345	0.088
	3sigma(test)	IQR	-0.05446	0.363629	0.10223	0.116
	3sigma(test)	Gini	-0.07937	0.347397	0.069362	0.076
	3sigma(Ref)	LS	-0.06624	0.326326	0.053097	0.0795
	3sigma(Ref)	IQR	-0.05712	0.360945	0.101018	0.122
	3sigma(Ref)	Gini	-0.07923	0.347594	0.066361	0.072
	6sigma(test)	LS	-0.37928	0.145053	0.234938	0.32
	6sigma(test)	IQR	-0.06587	0.373889	0.112259	0.123
	6sigma(test)	Gini	-0.29514	0.24437	0.168275	0.2025
	6sigma(Ref)	LS	-0.37932	0.145421	0.218136	0.2955
	6sigma(Ref)	IQR	-0.06858	0.371139	0.111897	0.125
	6sigma(Ref)	Gini	-0.29495	0.244294	0.162423	0.173
	2-3sigma(test)	LS	-0.6839	-0.01233	0.587483	0.529
	2-3sigma(test)	IQR	-0.20877	0.331378	0.229527	0.2015
	2-3sigma(test)	Gini	-0.62149	0.114694	0.500602	0.3785
	2-3sigma(Ref)	LS	-0.68701	-0.01521	0.569765	0.5305
	2-3sigma(Ref)	IQR	-0.21054	0.329309	0.233034	0.202
	2-3sigma(Ref)	Gini	-0.62339	0.111546	0.488732	0.3585

Table B.12: Small sample power with LSCF, Gini and IQR

N=20	Outliers	Procedure	η	$\eta_{upperlimit}$	MSE_{ETA}	γ
	None	LS	-0.33511	-0.09585	0.023853	0.7655
	None	IQR	-0.35466	-0.09831	0.058604	0.691
	None	Gini	-0.34969	-0.09624	0.027208	0.785
	3sigma(test)	LS	-0.41458	-0.13581	0.036932	0.8065
	3sigma(test)	IQR	-0.40088	-0.11692	0.072375	0.711
	3sigma(test)	Gini	-0.42579	-0.12749	0.0413	0.81
	3sigma(Ref)	LS	-0.4146	-0.1357	0.033862	0.8125
	3sigma(Ref)	IQR	-0.40354	-0.11954	0.071861	0.712
	3sigma(Ref)	Gini	-0.42565	-0.12764	0.039228	0.8275
	6sigma(test)	LS	-0.72763	-0.28359	0.210995	0.8845
	6sigma(test)	IQR	-0.41471	-0.11775	0.081725	0.706
	6sigma(test)	Gini	-0.64399	-0.21253	0.141764	0.866
	6sigma(Ref)	LS	-0.72767	-0.28328	0.199033	0.906
	6sigma(Ref)	IQR	-0.41743	-0.12046	0.081942	0.7005
	6sigma(Ref)	Gini	-0.6438	-0.21299	0.137127	0.897
	2-3sigma(test)	LS	-1.03225	-0.42186	0.565053	0.9275
	2-3sigma(test)	IQR	-0.55668	-0.15988	0.195778	0.7125
	2-3sigma(test)	Gini	-0.96939	-0.32948	0.474843	0.8915
	2-3sigma(Ref)	LS	-1.03536	-0.42489	0.552421	0.955
	2-3sigma(Ref)	IQR	-0.55844	-0.16204	0.200687	0.6895
	2-3sigma(Ref)	Gini	-0.9713	-0.33335	0.465102	0.9305

2.4 Small sample PBE power and level with incremental outliers

In a small sample PBE analysis, the outliers range from zero to four sigma. One subject's data had outliers to see the effect of small outliers on the results.

- No Out : implies that the data with no outliers were considered
- 0.5 sd: implies that one subject's Test reading was having an outlier of 0.5 standard deviations
- 1 sd: implies that one subject's Test reading was having an outlier of 1 standard deviations
- 1.5 sd: implies that one subject's Test reading was having an outlier of 1.5 standard deviations
- 2 sd: implies that one subject's Test reading was having an outlier of 2 standard deviations
- 2.5 sd: implies that one subject's Test reading was having an outlier of 2.5 standard deviations
- 3 sd: implies that one subject's Test reading was having an outlier of 3 standard deviations
- 3.5 sd: implies that one subject's Test reading was having an outlier of 3.5 standard deviations
- 4 sd: implies that one subject's Test reading was having an outlier of 4 standard deviations

Table B.13: α with LSCF, Gini and IQR with incremental outliers

$Alpha(N=20)$	Outliers	Procedure	η	$\eta_{upperlimit}$	MSE_{ETA}	α
	None	LS	0.01324	0.377943	0.04378	0.0395
		IQR	-0.00516	0.382656	0.088684	0.1015
		Gini	-0.00019	0.384766	0.055147	0.0405
	.5sigma	LS	0.011064	0.376451	0.0442	0.04
		IQR	-0.00796	0.381309	0.088235	0.1005
		Gini	-0.00426	0.381959	0.055669	0.041
	1sigma	LS	0.004526	0.372062	0.045064	0.044
		IQR	-0.0169	0.376838	0.089676	0.1025
		Gini	-0.0126	0.378102	0.056613	0.044
	1.5sigma	LS	-0.00637	0.364821	0.046543	0.051
		IQR	-0.0274	0.371864	0.091819	0.105
		Gini	-0.02491	0.371981	0.058309	0.05
	2sigma	LS	-0.02164	0.354801	0.048923	0.057
		IQR	-0.03806	0.367389	0.095391	0.1075
		Gini	-0.04049	0.36439	0.060732	0.0545
	2.5sigma	LS	-0.04126	0.342095	0.052602	0.076
		IQR	-0.04676	0.365028	0.098596	0.1105
		Gini	-0.05846	0.356391	0.064335	0.0655
	3sigma	LS	-0.06623	0.326196	0.058345	0.088
		IQR	-0.05446	0.363629	0.10223	0.116
		Gini	-0.07937	0.347397	0.069362	0.076
	3.5sigma	LS	-0.0936	0.309087	0.066033	0.1075
		IQR	-0.05907	0.364411	0.104787	0.118
		Gini	-0.10148	0.33774	0.075287	0.086
	4sigma	LS	-0.12631	0.289037	0.077153	0.1355
		IQR	-0.06248	0.365499	0.106803	0.1225
		Gini	-0.12684	0.325979	0.083315	0.097

Table B.14: Power with LSCF, Gini and IQR with incremental outliers

$Power(N = 20)$	Outliers	Procedure	η	$\eta_{upperlimit}$	MSE_{ETA}	γ
	None	LS	-0.33511	-0.09585	0.023853	0.7655
		IQR	-0.35466	-0.09831	0.058604	0.691
		Gini	-0.34969	-0.09624	0.027208	0.785
	.5sigma	LS	-0.33729	-0.097	0.024063	0.767
		IQR	-0.35573	-0.09792	0.058461	0.69
		Gini	-0.35203	-0.09728	0.027623	0.7845
	1sigma	LS	-0.34383	-0.10037	0.024703	0.7715
		IQR	-0.36338	-0.10163	0.060228	0.7055
		Gini	-0.35908	-0.10027	0.028565	0.784
	1.5sigma	LS	-0.35473	-0.10593	0.025943	0.781
		IQR	-0.37296	-0.1057	0.06247	0.707
		Gini	-0.37046	-0.10506	0.030155	0.79
	2sigma	LS	-0.36999	-0.11364	0.02807	0.787
		IQR	-0.38308	-0.10977	0.065714	0.708
		Gini	-0.38552	-0.11128	0.032594	0.8015
	2.5sigma	LS	-0.38961	-0.12346	0.031482	0.7955
		IQR	-0.39216	-0.11326	0.068845	0.708
		Gini	-0.40386	-0.11872	0.036177	0.8035
	3sigma	LS	-0.41458	-0.13581	0.036932	0.8065
		IQR	-0.40088	-0.11692	0.072375	0.711
		Gini	-0.42579	-0.12749	0.0413	0.81
	3.5sigma	LS	-0.44195	-0.1492	0.044334	0.813
		IQR	-0.40621	-0.11848	0.075311	0.713
		Gini	-0.44862	-0.13658	0.047578	0.8185
	4sigma	LS	-0.47466	-0.16504	0.055144	0.826
		IQR	-0.41005	-0.11926	0.077348	0.7075
		Gini	-0.47441	-0.14674	0.055832	0.825

2.5 ABE power and level with LS and HL estimators

ABE procedure uses the LS and the Componentwise rank methods with a two one-sided hypothesis. The outliers vary from none to 3σ outliers. They are as shown:

- None : implies that the data with no outliers were considered
- 1.5 sigma: implies that one subject's reading was having an outlier of 1.5 standard deviations
- 1.5 (2) sigma: implies that two subject's readings had outliers of 1.5 standard deviations
- 3 sigma: implies that one subject's reading had an outlier of 3 standard deviations

Subjects with sample size 14, 16, 18, 20, 22 were considered for our simulation. This meant that each sequence had 7, 8, 9, 10, 11 subjects respectively.

Table B.15: ABE power and level with LS and HL estimators

Outlier	N	R				LS			
		Alpha	MSE	Power	MSE	Alpha	MSE	Power	MSE
None	14	0.071	0.77651	0.934	0.77651	0.079	0.774715	0.976	0.774715
	16	0.065	0.774958	0.957	0.774958	0.059	0.771496	0.982	0.771496
	18	0.07	0.765612	0.963	0.765612	0.062	0.763884	0.991	0.763884
	20	0.053	0.765926	0.978	0.765926	0.044	0.763879	0.993	0.763879
	22	0.054	0.756333	0.978	0.756333	0.05	0.754458	0.994	0.754458
1.5 sigma	14	0.075	0.787293	0.84	0.787293	0.092	0.793901	0.862	0.793901
	16	0.066	0.781173	0.881	0.781173	0.072	0.785535	0.892	0.785535
	18	0.065	0.774061	0.938	0.774061	0.059	0.77949	0.916	0.77949
	20	0.059	0.76901	0.931	0.76901	0.052	0.773608	0.931	0.773608
	22	0.061	0.762105	0.965	0.762105	0.048	0.76627	0.956	0.76627
1.5 (2) sigma	14	0.034	0.817761	0.847	0.817761	0.027	0.842522	0.875	0.842522
	16	0.035	0.807402	0.911	0.807402	0.022	0.830114	0.908	0.830114
	18	0.038	0.796192	0.899	0.796192	0.019	0.816352	0.911	0.816352
	20	0.029	0.790707	0.93	0.790707	0.02	0.809809	0.938	0.809809
	22	0.032	0.789121	0.956	0.789121	0.014	0.804202	0.955	0.804202
3 sigma	14	0.09	0.78694	0.589	0.78694	0.092	0.830632	0.23	0.830632
	16	0.06	0.782105	0.71	0.782105	0.065	0.819001	0.279	0.819001
	18	0.067	0.775185	0.836	0.775185	0.065	0.808115	0.367	0.808115
	20	0.051	0.770987	0.907	0.770987	0.044	0.800214	0.391	0.800214
	22	0.05	0.76767	0.928	0.76767	0.04	0.789038	0.461	0.789038

BIBLIOGRAPHY

Berger, R. & Hsu, J. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, **11**, 283–302.

Chinchilli, V. & Esinhart, J. (1996). Design and analysis of intra-subject variability in crossover experiments. *Statistics in Medicine*, **15**, 1619–1634.

Chow, S.C. & Wang, H. (2001). On sample size calculation in bioequivalence trials. *Journal of Pharmacokinetics and Pharmacodynamics*, **28**, 155–169.

Chow SC, W.H., Shao J (2002). Individual bioequivalence testing under 2x3 designs. *Statistics in Medicine*, **21**, 629–648.

Cornish, E. & Fisher, R. (1938). Moments and cumulants in the specification of distributions. *Revue de lInstitut International de Statistique*, **5**, 307320.

CPMP (2001). Note for guidance on the investigation of bioavailability and bioequivalence. Committee for Proprietary Medicinal Products-The European Agency for the Evaluation of Medicinal Products Evaluation of Medicines for Human Use.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer Texts in Statistics.

DasGupta, A. & Haff, L. (2006). Asymptotic expansions for the correlation between different measures of spread. *JSPI*, 2197–2212.

David, H.A. (1968). Gini's mean difference rediscovered. *Biometrika*, **55**, 573–575.

Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap.* Chapman and Hall New York.

FDA (1992). The fda guidance on statistical procedures for bioequivalence studies using a standard two treatment crossover design. U.S. Department of Health and Human Services Food and Drug Administration-Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research.

FDA (2001). Statistical approaches to establishing bioequivalence. U.S. Department of Health and Human Services Food and Drug Administration-Center for Drug Evaluation and Research.

FDA (2003a). Bioavailability and bioequivalence studies for orally administered drug products general considerations. U.S. Department of Health and Human Services Food and Drug Administration-Center for Drug Evaluation and Research.

FDA (2003b). Bioequivalence datasets.

Ghosh, P. & Gonen, M. (2007). Bayesian modeling of multivariate average bioequivalence. *Statistics in Medicine*.

Hauschke, D., Steinijans, V. & Pigeot, I. (2007). *Bioequivalence Studies in Drug Development*. John Wiley and Sons Ltd.

Hettmansperger, T.P. & McKean, J.W. (1998). *Kendalls Library of Statistics 5, Robust Nonparametric Statistical Methods*. An Arnold Publication.

Hollander, M. & Wolfe, D.A. (2001). *Nonparametric Statistical Methods*. John Wiley & Sons Ltd.

Howe, W. (1974). Approximate confidence limits on the mean of x+y where x and y are two tabled independent random variables. *Journal of the American Statistical Association*, **69**, 789–794.

Hui Quan, W.Y., Jim Bolognese (2001). Assessment of equivalence on multiple endpoints. *Statistics in Medicine*, **20**, 3159–3173.

Hwang, J.T.G. (1996). Comment on "bioequivalence trials, intersection-union tests and equivalence confidence sets". *Statistical Science*, **11**, 313315.

Jaschke, S.R. (2001). The cornish-fisher-expansion in the context of delta-gamma-normal approximations. *Humboldt-Universitt Berlin*.

Johnson, R.A. & Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. Prentice Hall.

Koul, H.L., Sievers, G.L. & McKean, J. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandinavian Journal of Statistics*, **14**, 131–141.

Lee Yonghee, J.S. & Chung, C.S. (2004). Modified large-sample confidence intervals for linear combinations of variance components: Extension, theory, and application. *Journal of the American Statistical Association*, **99**, 467–478.

Leena Choi, C.R., Brian Caffo (2008). A survey of the likelihood approach to bioequivalence trials. *Statistics in Medicine*, **27**, 4874–4894.

Lehmann, E. & Romano, J.P. (2005). *Testing Statistical Hypotheses*. Springer.

Lomnicki, Z.A. (1952). The standard error of gini's mean difference. *The Annals of Mathematical Statistics*, **23**, 635–637.

McKean, J. (2009). www.stat.wmich.edu/mckean.

Midha, O.E.H.J.M.G.H.E., K. & Gavalas, L. (1993). Logarithmic transformation in bioequivalence: application with two formulations of perphenazine. *Journal of Pharmaceutical Sciences*, **82**, 138144.

Nair, U.S. (1936). The standard error of gini's mean difference. *Biometrika*, **28**, 428–436.

Niazi, S.K. (2007). *Handbook of Bioequivalence Testing*. Pharmaceutical Scientist Inc. Deerfield, Illinois, USA.

Ola Hssjer, P.J.R. & Croux, C. (1996). Asymptotics of an estimator of a robust spread functional. *Statistica Sinica*, **6**, 375–388.

Rousseeuw, P.J. & Croux, C. (1992a). A class of high-breakdown scale estimators based on subranges. *Communications in Statistics: Theory and Methods*, **21**, 1935–1951.

Rousseeuw, P.J. & Croux, C. (1992b). Time-efficient algorithms for two highly robust estimators of scale, in computational statistics. *Computational Statistics*, **1**, 411–428.

Rousseeuw, P.J. & Croux, C. (1992c). Time-efficient algorithms for two highly robust estimators of scale, in computational statistics. *Computational Statistics*, **1**, 411–428.

Rousseeuw, P.J. & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, **88**, 1273–1283.

Schall, R. & Luus, H. (1993). On population and individual bioequivalence. *Statistics in Medicine*, **12**, 110924.

Schuirmann, D. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokin. Biopharm.*, **15**, 657–80.

Scott D Patterson, B.J. (2002). Replicate designs and average, individual, and population bioequivalence. Tech. rep., GlaxoSmithKline.

Serfling, R.J. (2001). *Approximation Theorems of Mathematical Statistics*. Wiley Interscience.

Skelly, J., Midha, K. & Blume, H. (1995). Bio-international 94, conference on bioavailability, bioequivalence and pharmacokinetic studies. *European Journal of Pharmaceutical Sciences*, **3**, 114–115.

Stefanescu, C. & Mehrotra, D.V. (2008). A more powerful average bioequivalence analysis for the 2x2 crossover. *Communications in Statistics - Simulation and Computation*, **37**, 212 – 221.

Sunkara, G., Sabo, R., Wang, Y., He, Y.L., Campestrini, J., Rosenberg, M., Howard, D. & Dole, W. (2007). Dose proportionality and the effect of food on vildagliptin, a novel dipeptidyl peptidase iv inhibitor, in healthy volunteers. *Journal of Clinical Pharmacology*, **47**, 1152–1158.

Terry Hyslop, F.H. & Holder, D.J. (2000). A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*, **19**, 2885–2897.

Von Luxburg, U. & Franz, V. (2004). Confidence sets for ratios: A purely geometric approach to fieller's theorem. *Max Planck Institute for Biological Cybernetics*, **133**.

Westlake, W. (1988). Bioavailability and bioequivalence of pharmeceutical formulations. *Biopharmeceutical Staistics for Drug development.*

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

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

Yates, R., Nairn, K., Dixon, R., Kemp, J. & Dane, A. (2002). Pharmacokinetics, dose proportionality, and tolerability of single and repeat doses of a nasal spray formulation of zolmitriptan in healthy volunteers. *Journal of Clinical Pharmacology*, **42**, 1244–1250.