A novel approach to testing for bioequivalence based on modeling the within-period dependence structure

Rameela Chandrasekhar[1], Alan D. Hutson[1] and Gregory E. Wilding[1*]

[1]*Department of Biostatistics, University at Buffalo, 3435 Main Street, Buffalo NY 14214.*

Abstract

Bioequivalence trials are commonly conducted to assess therapeutic equivalence between a generic and an innovator brand drug. In such trials, drug concentrations are obtained repeatedly over time and a metric such as the area under the concentration versus time curve is computed for each subject. Standard methods are then applied to conduct tests of average bioequivalence. A major disadvantage of this approach is the loss of information encountered when ignoring the correlation structure between repeated measurements. We propose a general linear model approach, incorporating the within-subject covariance structure, for making inferences. We investigate and compare the inferential properties of our proposed method with the traditional two one-sided tests approach using Monte Carlo simulation studies. We also examine the properties of the method in the event of missing data. Simulations show that the proposed approach is a more cost-effective viable alternative to the traditional method with superior inferential properties. Inferential advantages are particularly apparent in the presence of missing data. To illustrate our approach, a real working example from an asthma study is utilized.

Keywords: Bioequivalence, Area under the curve, Trapezoidal rule, Longitudinal data.

1 Introduction

Bioequivalence is defined by the United States Food and Drug Administration (FDA) as, "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" (FDA, 2008). Benet & Goyan (1995) described three situations that require bioequivalence trials: (i) when the proposed marketed dosage form is different from that used in pivotal clinical trials, (ii) when significant changes are made in

*Corresponding author: Department of Biostatistics, University at Buffalo, 3435 Main Street, Buffalo NY 14214. E-mail: gwilding@buffalo.edu

the manufacturing of the marketed formulation and (iii) when a new generic formulation is tested against the innovator's marketed product.

Bioequivalence studies conducted to assess therapeutic equivalence between two drug formulations are generally carried out between an innovator brand formulation and a generic formulation. Before a new drug is introduced into the market, clinical trials are conducted to test for safety, tolerability and effectiveness. Once the patent protection expires, other manufacturers are permitted to formulate a generic equivalent of the brand name counterpart. For a generic brand to replicate the clinical trial process would require substantial time and resources and may also be considered to be unethical such as in scenarios where the original drug was shown to be superior as compared to control. Instead, the FDA requires that the generic brand demonstrate equivalence to the innovator brand before it can be approved and marketed. An abbreviated new drug application (ANDA) is filed with the FDA along with evidence of bioequivalence between the test and reference formulations. Generic drug formulations that demonstrate bioequivalence are concluded to be therapeutically equivalent and interchangeable with the brand name counterpart.

Bioequivalence trials involve the administration of the test and reference drug in subjects and documentation of the relevant clinical parameters from blood or plasma measured at prespecified time intervals, generating the drug concentration versus time curve or bioavailability profile. Pharmacokinetic parameters commonly used to assess bioequivalence are the area under the concentration versus time curve (AUC), maximum concentration (C_{max}) and time to reach maximum concentration (t_{max}). The AUC is considered a measure of overall drug exposure in the body and for purposes of our work, we focus on it as the primary outcome. Various methods for approximating the AUC have been referenced in the literature such as the linear trapezoidal rule, log-linear trapezoidal rule, Lagrange method and the spline method. In this paper, we consider use of the most commonly used linear trapezoidal method. Let Y_{ij} be the observed concentration for subject i at the j^{th} time point t_j, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Then the AUC between time points t_j and t_{j-1} computed using the trapezoidal rule can be defined as

$$\begin{aligned} [AUC]_{t_{j-1}}^{t_j} &= \int_{t_{j-1}}^{t_j} y dt, \\ &= 0.5 * (t_j - t_{j-1})(Y_{ij} + Y_{i(j-1)}). \end{aligned} \tag{1}$$

Though multiple measurements spanning over a period of time are obtained, the AUC may sometimes be calculated only for a clinically relevant subset of time points. This is often the case when early or late exposure is of interest and is termed as the partial AUC.

The primary objective of bioequivalence trials is to determine if the effect of two treatments differ by more than a prespecified amount in either direction. Let μ_T and μ_R denote the population mean AUC for

the treatment and reference, respectively. The general bioequivalence hypotheses can be written as

$$H_0 : \frac{\mu_T}{\mu_R} \leq \theta_L \text{ or } \frac{\mu_T}{\mu_R} \geq \theta_U,$$
$$H_a : \theta_L < \frac{\mu_T}{\mu_R} < \theta_U, \tag{2}$$

where θ_L and θ_U are bioequivalence limits (BEL) recommended by the FDA. Bioequivalence is concluded if the null hypothesis is rejected. The bioequivalence hypotheses can also be stated in terms of the difference in areas represented by $\theta = \mu_T - \mu_R$. Then the bioequivalence hypotheses in terms of the absolute difference can be written as

$$H_0 : \mu_T - \mu_R \leq \theta_L \text{ or } \mu_T - \mu_R \geq \theta_U,$$
$$H_a : \theta_L < \mu_T - \mu_R < \theta_U. \tag{3}$$

Though parallel designs may be used, the classic experimental design adopted in bioequivalence studies are typically crossover in nature. The replicated 2×2 crossover experimental design commonly followed is illustrated in Figure 1. Subjects are randomized and a sequence of treatments is administered with a reasonable wash-out period (usually 3 half-lives of the drug) between treatments. Higher-order crossover designs for two formulations such as 2×4 or 2×3 are more often considered in recent times and current FDA guidelines recommend the use of a 2×4 crossover design to assess bioequivalence (FDA, 2001). For more on the design and analysis of crossover trials, see Jones & Kenward (2003) and Senn (2002).

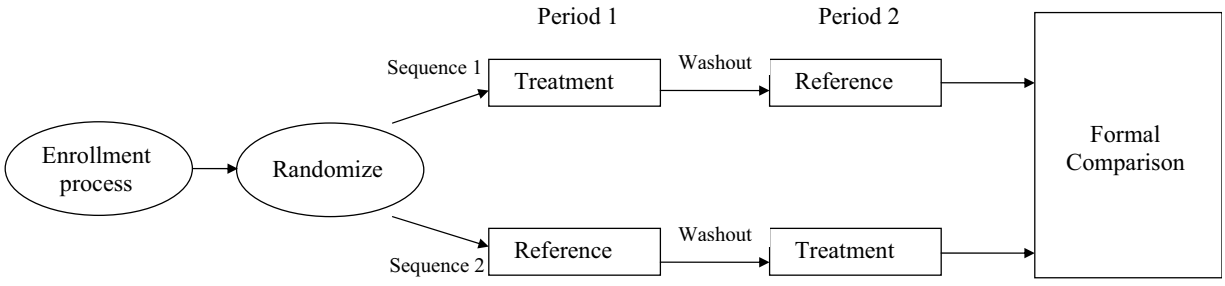


Figure 1: 2×2 crossover design

Though summary metrics such as the AUC are commonly used, significant questions exist regarding the capture of important features of the bioavailability profile. Bioequivalence testing using summary statistics may give misleading inferences if characteristic features of the bioavailability profile are not captured (Mauger & Chinchilli, 2000a). For example, features such as the within-subject variability or the shape of the bioavailability profile are ignored in the analysis of bioequivalence based on the AUC. Several authors

have proposed summarizing profile differences using indices that measure profile similarity to overcome this problem (see Marston & Polli, 1997; Mauger & Chinchilli, 2000b; Rescigno, 1992). Multivariate tests of simultaneous assessment of bioequivalence on multiple summary measures have also been developed (see Berger & Hsu, 1996; Chinchilli & Elswick, 1997; Wang et al., 1999).

Though the AUC is an attractive summary measure, assessment of bioequivalence using the AUC ignores the within-subject correlation component. This motivates us to consider a linear model-based approach that encapsulates this important feature of the subject response profile when making inferences on the mean AUC. Estimation of parameters will be approached via the restricted maximum likelihood (REML) method along with the Newton-Raphson iterative algorithm to optimize the likelihood function. Contrasts and corresponding confidence intervals will be constructed to conduct tests of bioequivalence. This will be explained further in Section 2.2.

The rest of the paper is structured as follows. In Section 2, we review the current methods to assess bioequivalence and develop the proposed approach. Our proposed method is compared with the traditional two one-sided tests approach using simulations in Section 3. In Section 4, a real working example from an asthma study is utilized for illustration purposes and we conclude and summarize this paper in Section 5.

2 Assessment of bioequivalence

The three types of bioequivalence evaluated are average bioequivalence (ABE), population bioequivalence (PBE) and individual bioequivalence (IBE). Though various articles have addressed the limitations of ABE as being that between-patient variability of these formulations is not taken into account, ABE still remains a popular criterion in evaluating bioequivalence due to its ease of interpretation (Ghosh & Rosner, 2007). For the purposes of our work, we concentrate on the methods used to evaluate for ABE. The following decision rules are most commonly recommended. The ± 20 rule states that bioequivalence is concluded if the average bioavailability of the test formulation is within $\pm 20\%$ of that of the reference formulation with a certain assurance. Then the hypotheses are formulated as in (2) using the ratio of means, with $\theta_L = 0.80$ and $\theta_U = 1.20$. When using the absolute difference as the parameter of interest, the hypotheses are formulated as in (3) with $(\theta_L, \theta_U) = (-0.20\mu_R, +0.20\mu_R)$. A major drawback with using $\pm 0.20\mu_R$ as bioequivalence limits for the difference estimator is that is that $\widehat{\mu}_R$ is assumed to be the true μ_R and the variability of $\widehat{\mu}_R$ is thus not taken into account. The 80/125 rule states that bioequivalence is concluded if the average bioavailability of the test formulation is within (80%, 125%) that of the reference formulation, with a certain assurance. In practice, it is generally assumed that pharmacokinetic data are log-normally distributed (Julious & Debarnot, 2000; Lacey et al., 1997). The FDA recommends log-transforming the data and testing the null hypotheses

listed in (2), where $\theta_L = 0.80$ and $\theta_U = 1.25$, which represents a symmetric range on the log-scale (FDA, 2001). Also, in using the 80/125 rule for the analysis of log-transformed data, the probability of concluding bioequivalence is at a maximum if the ratio of averages is in fact 1.

Arguments for log-transformation seem to be based on the perception that defining equivalence in terms of the ratio of parameters is more difficult than dealing with it in terms of the difference of parameters (Berger & Hsu, 1996). Berger & Hsu (1996) noted that if appropriate statistical procedures can be used to make inferences about the mean ratios directly, then there seems to be no need for a log transformation. Liu & Weng (1994) used a simulation study to demonstrate how analyses of bioequivalence based on the original scale are always more powerful than those based on the transformed scale when the distribution is normal. Based on the above arguments, in this paper we deal with the case where pharmacokinetic parameters are considered to be normally distributed. The null hypothesis listed in (2) is tested and the 80% and 120% limits are used to assess bioequivalence although difference limits may be accommodated.

2.1 Traditional approaches

For bioequivalence to be concluded with 100(1-2α)% assurance, the Schuirmann's method is commonly implemented. Schuirmann's method (Schuirmann, 1987), also termed as the two one-sided tests (TOST), is an Intersection-Union test that evaluates for bioequivalence by decomposing the hypotheses in (3) into two sets of one-sided hypotheses specified as

$$\begin{aligned} H_{01} &: \mu_T - \mu_R \leq \theta_L \\ H_{a1} &: \mu_T - \mu_R > \theta_L, \end{aligned} \tag{4}$$

and

$$\begin{aligned} H_{02} &: \mu_T - \mu_R \geq \theta_U, \\ H_{a2} &: \mu_T - \mu_R < \theta_U. \end{aligned} \tag{5}$$

Bioequivalence is concluded only if the null hypotheses in (4) and (5) are both rejected. Let $\widehat{\theta} = \widehat{\mu}_T - \widehat{\mu}_R$ be an estimate for $\mu_T - \mu_R$ and $\widehat{\sigma}_\theta$ be the standard deviation of the period differences for each subject. Assuming normality, the null is rejected if

$$T_L = \frac{\widehat{\theta} - \theta_L}{\widehat{\sigma}_\theta \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{(\alpha, n_1+n_2-2)}, \tag{6}$$

and

$$T_U = \frac{\widehat{\theta} - \theta_U}{\widehat{\sigma}_\theta \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{(\alpha, n_1+n_2-2)}, \tag{7}$$

where n_1 and n_2 ($n_1 + n_2 = n$) are the sample sizes randomized to respective sequences and $t_{(\alpha,n_1+n_2-2)}$ is the upper 100α percentile of a Student's t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Westlake (1972) and Metzler (1974) suggested the use of confidence intervals to evaluate therapeutic equivalence between drug formulations based on the argument that assessment of bioequivalence involves determining if the means of the formulations were sufficiently close. Various approaches based on confidence intervals to assess bioequivalence have been developed. The classic confidence interval, also known as the shortest confidence interval, is constructed based on the t statistic for the absolute difference in mean areas and is operationally equivalent to the TOST approach (Schuirmann, 1987; Westlake, 1976). This confidence interval is then transformed into that for the ratio by dividing the upper and lower limit by the observed reference mean and adding 1, assuming $\widehat{\mu}_R = \mu_R$ (Chow & Liu, 2008). Let the ratio of AUC means be denoted by $\Delta = \mu_T/\mu_R$. Then using the ± 20 rule, ABE is concluded if a 100(1-2α) confidence interval for Δ is within 80% and 120% or within $\pm 20\%\mu_R$ for θ. To overcome the drawbacks associated with the classic confidence interval approach, other confidence interval methods such as Westlake's symmetric confidence interval (Westlake, 1976), Chow and Shao's joint confidence region (Chow & Shao, 1990) and an interval based on Fieller's method (Fieller, 1954) have been developed.

Under the current guidance from the FDA, the two one-sided tests approach is suggested to evaluate for average bioequivalence. In the original scale, this method has been found to be biased since it ignores the variability of the least squares mean of the reference as it substitutes $\widehat{\mu}_R$ for the unknown reference average when using bioequivalence limits of $\pm 0.20\mu_R$ (Liu & Weng, 1995; Vuorinen & Tuominen, 1994). Though simplistic in nature, this approach suffers from lack of power and is conservative (Berger & Hsu, 1996). Several versions that improve upon Schuirmann's test have been proposed (Cao & Mathew, 2008; Liu & Weng, 1995; Locke, 1984; Yee, 1986). Anderson & Hauck (1983) proposed a test for average bioequivalence that is more powerful than the TOST approach though it does not control for Type I error for small samples. Another test proposed by Brown et al. (1997) was shown to be unbiased and uniformly more powerful than the TOST method. Berger & Hsu (1996) constructed a test based on the intersection-union principle that was approximately unbiased but was more powerful than the TOST method.

2.2 Proposed Approach

In this section, we develop our proposed approach to evaluate for bioequivalence. The relationship between the AUC and the linear combination of time point specific sample means is discussed and inference methods

based on a general linear model are examined. For simplicity we focus on a 2×2 crossover design to develop our proposed approach, but our method is intended to be flexible in that it can be extended to parallel as well as higher order crossover designs.

2.2.1 Re-expression of the mean AUC as a linear combination of means

Assuming all subjects to be observed at the same set of m occasions, the total AUC for the i^{th} individual can be computed using the trapezoidal rule as follows. Letting $\mathbf{Y}_i = (Y_{i1}\ Y_{i2}\ Y_{i3}\ \ldots\ Y_{im})'$, the AUC for the i^{th} subject can be written as $AUC_i = \mathbf{c}'\mathbf{Y}_i$, where $\mathbf{c} = (c_1\ c_2\ c_3\ \ldots\ c_m)'$, and

$$c_j = \begin{cases} \frac{t_{j+1}-t_j}{2}, & j = 1, \\ \frac{t_j-t_{j-1}}{2}, & j = m, \\ \frac{t_{j+1}-t_{j-1}}{2}, & otherwise. \end{cases} \tag{8}$$

Let $\overline{\mathbf{Y}} = (\overline{Y}_{\cdot 1},\ \overline{Y}_{\cdot 2},\ \ldots,\ \overline{Y}_{\cdot m})'$, where $\overline{Y}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}$. Furthermore let $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_m)'$, where $\mu_j = E(\overline{Y}_{\cdot j})$. The following theorem, taken from Wilding et al. (2011), establishes that the estimate of the mean AUC based on the trapezoidal rule can be written as a linear combination of the mean responses at each time point.

Theorem 1. *The estimate of the mean AUC obtained using the trapezoidal rule can be re-expressed as a linear combination of the mean outcome values at each sampling time, that is* $\widehat{\mu}_{AUC} = \mathbf{c}'\overline{\mathbf{Y}}$ *with* $E(\widehat{\mu}_{AUC}) = \mathbf{c}'\mu$ *and* $Var(\widehat{\mu}_{AUC}) = \mathbf{c}'Var(\overline{\mathbf{Y}})\mathbf{c}$. *(Refer to Wilding et al. (2011), for proof).*

2.2.2 The Model

Expressing the parameter of interest to be of the form $\mathbf{c}'\boldsymbol{\mu}$ leads us to propose an alternative model-based procedure for the assessment of bioequivalence. Though the actual estimate of AUC has the same interpretation as before, we are able to obtain more efficient estimates by incorporating the within-subject covariance structure. Using our proposed approach also lends an advantage in that profiles with missing observations can also be accomodated. We assume a linear model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9}$$

where $\mathbf{Y}$ is a $2nm$ column vector consisting of concentrations from n individuals at m time points from both periods, $\mathbf{X}$ is the design matrix and $\boldsymbol{\beta}$ is a function of parameters such as treatment, time, sequence, period and selected interactions. Furthermore $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$. The matrix $\boldsymbol{\Sigma}$ is a block diagonal

matrix, with the within-subject covariance structure $\mathbf{\Sigma}^*$ representing blocks on the main diagonal and zeros elsewhere.

Unlike for a parallel design, the errors for a crossover design in this context are correlated not just within a single treatment, but also between treatments for a single individual. One may attempt to accommodate the dependencies through inclusion of a single random subject effect as is commonly done in the analysis of data obtained from crossover designs. The inherent problem associated with such a strategy is that the linear correlations between any two time points are assumed equal which is most likely not reasonable. For simplicity, the within-subject covariance structure can be assumed to be the Kronecker product between two matrices. That is, we may let $\mathbf{\Sigma}^* = \mathbf{\Psi} \otimes \mathbf{\Omega}$, where $\mathbf{\Psi}$ and $\mathbf{\Omega}$ represents the between and within treatment or period covariance structure. This enables us to specify the variance-covariance matrix of the error within each study unit and thus to examine the intra and inter treatment correlations of the errors (Gao et al., 2006). For example, let us consider a two period, three within-period time point study where an unstructured between period covariance structure can be assumed and the structure for an individual within-period can be described by a first-order autoregressive structure. Using the Kronecker product, $\mathbf{\Sigma}^*$ can be specified as

$$\mathbf{\Psi} \otimes \mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}.$$

Other within-subject covariance structures for $\mathbf{\Psi}$ and $\mathbf{\Omega}$ include compound symmetric, autoregressive or unstructured. Specification of the covariance structure may be guided using historic information or indices of goodness-of-fit such as the Akaike Information Criterion (AIC) (Ferron et al., 2002; Keselman et al., 1998).

Let us define the population mean AUC for the treatment in the k^{th} sequence and l^{th} period to be $\mu_{kl} = g_{kl}(\mathbf{c})\boldsymbol{\beta} = \mathbf{d}'_{kl}\boldsymbol{\beta}$, where $g_{kl}(\mathbf{c})$ denotes a function of $\mathbf{c}$ and $\mathbf{c}$ is as defined in (8). Using the fact that the estimate can be defined as a linear combination of the mean outcome values in that block, it can be expressed as $\widehat{\mu}_{kl} = g_{kl}(\mathbf{c})\widehat{\boldsymbol{\beta}} = \mathbf{d}'_{kl}\widehat{\boldsymbol{\beta}}$. For the less than full rank parameterization of the model, $\mu_{kl} = \mathbf{d}'_{kl}A\boldsymbol{\beta} \neq \mathbf{d}'_{kl}\boldsymbol{\beta}$, where $A = G^-G$ is a matrix and G^- is the generalized inverse such that it satisfies the condition $GG^-G = G$. Then the estimate of AUC obtained is not unique and depends on the particular generalized inverse used. However, inferences are based on the reparameterized full rank model ensuring a unique solution.

Theorem 2. *Assume model* (9) *with* $E(\boldsymbol{\epsilon}) = 0$ *and* $Var(\boldsymbol{\epsilon}) = \mathbf{\Sigma}$. *Then the estimate of the mean area for the treatment in sequence* k *and period* l *is* $\widehat{\mu}_{kl} = \boldsymbol{d}'_{kl}\widehat{\boldsymbol{\beta}}$ *with* $E(\boldsymbol{d}'_{kl}\widehat{\boldsymbol{\beta}}) = \boldsymbol{d}'_{kl}\boldsymbol{\beta}$ *and* $Var(\boldsymbol{d}'_{kl}\widehat{\boldsymbol{\beta}}) = \boldsymbol{d}'_{kl}Var(\widehat{\boldsymbol{\beta}})\boldsymbol{d}_{kl}$. *Furthermore, assuming* $\boldsymbol{\epsilon} \sim N(0, \mathbf{\Sigma})$, *the distribution is given by* $\boldsymbol{d}'_{kl}\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{d}'_{kl}\boldsymbol{\beta}, \boldsymbol{d}'_{kl}Var(\widehat{\boldsymbol{\beta}})\boldsymbol{d}_{kl})$ *(Refer to Wilding et al. (2011), for proof).*

For a 2×2 crossover design, a pair of areas are obtained for each subject at periods 1 and 2. Using Theorems 1 and 2, the mean AUC for the treatment and reference can be obtained from the fitted model using the predicted means at each sampling time point.

Corollary 1. *The estimate for the mean AUC for treatment T is $\widehat{\mu}_T = (\widehat{\mu}_{k1} + \widehat{\mu}_{k'2})/2$, where $\widehat{\mu}_{kl}$ is the estimate of mean AUC for sequence k and period l with $E(\widehat{\mu}_T) = \mu_T = (\boldsymbol{d}'_{k1}\boldsymbol{\beta} + \boldsymbol{d}'_{k'2}\boldsymbol{\beta})/2$ and $Var(\widehat{\mu}_T) = Var(\widehat{\mu}_{k1} + \widehat{\mu}_{k'2})/4 = Var(\boldsymbol{d}'_{k1}\widehat{\boldsymbol{\beta}} + \boldsymbol{d}'_{k'2}\widehat{\boldsymbol{\beta}})/4$. The estimate for $\widehat{\mu}_R$ may be similarly defined.*

The ratio of AUC means can then be estimated as $\widehat{\Delta} = \widehat{\mu}_T/\widehat{\mu}_R$, where $\widehat{\mu}_T$ and $\widehat{\mu}_R$ are the estimates of the treatment and reference drug effect respectively.

Theorem 3. *Under model (9), if the estimate for the mean AUC for the treatment and reference is represented using $\widehat{\mu}_T$ and $\widehat{\mu}_R$ respectively, then the ratio of mean areas $\Delta = \mu_T/\mu_R$ can be estimated as $\widehat{\Delta} = \widehat{\mu}_T/\widehat{\mu}_R$ with distribution $\widehat{\Delta} \sim AN(\mu_T/\mu_R, (\widehat{\sigma}^2_{TT} - 2\widehat{\Delta}\widehat{\sigma}_{TR} + \widehat{\Delta}^2\widehat{\sigma}^2_{RR})/(\widehat{\mu}_R)^2)$.*

Proof. For the estimate of mean areas, the Taylor series expansion about (μ_T, μ_R) is given by,

$$\frac{\widehat{\mu}_T}{\widehat{\mu}_R} \approx \frac{\mu_T}{\mu_R} - \frac{\mu_T}{\mu^2_R}(\widehat{\mu}_R - \mu_R) + \frac{1}{\mu_R}(\widehat{\mu}_T - \mu_T). \tag{10}$$

Taking expectations on both sides,

$$E\left(\frac{\widehat{\mu}_T}{\widehat{\mu}_R}\right) \approx \frac{\mu_T}{\mu_R}. \tag{11}$$

Taking the variance of both sides of (10) yields

$$\widehat{Var}\left(\frac{\widehat{\mu}_T}{\widehat{\mu}_R}\right) \approx \frac{\left(\widehat{\sigma}^2_{TT} - 2\widehat{\Delta}\widehat{\sigma}_{TR} + \widehat{\Delta}^2\widehat{\sigma}^2_{RR}\right)}{(\widehat{\mu}_R)^2}, \tag{12}$$

where $\sigma^2_{TT} = Var(\widehat{\mu}_T)$, $\sigma^2_{RR} = Var(\widehat{\mu}_R)$ and $\sigma_{TR} = Cov(\widehat{\mu}_T, \widehat{\mu}_R)$. By the Delta theorem,

$$\frac{\widehat{\mu}_T}{\widehat{\mu}_R} - \frac{\mu_T}{\mu_R} \xrightarrow{d} N\left(0, Var\left(\frac{\widehat{\mu}_T}{\widehat{\mu}_R}\right)\right). \tag{13}$$

□

Thus, the appropriate contrasts can be constructed to estimate the mean area in each treatment. In addition, the null hypotheses can be expressed in terms of the vector of coefficients $\mathbf{c}$ and the parameter vector $\boldsymbol{\beta}$. Let $\mu_T = \mathbf{d}'_1\boldsymbol{\beta}$ and $\mu_R = \mathbf{d}'_2\boldsymbol{\beta}$, where $\mathbf{d}_k = \mathbf{d}'_{k1}\boldsymbol{\beta} + \mathbf{d}'_{k'2}\boldsymbol{\beta}$. Then the hypotheses specified in (2)

can be written as

$$H_0 : \frac{\mathbf{d}_1'\boldsymbol{\beta}}{\mathbf{d}_2'\boldsymbol{\beta}} \leq \theta_L \text{ or } \frac{\mathbf{d}_1'\boldsymbol{\beta}}{\mathbf{d}_2'\boldsymbol{\beta}} \geq \theta_U,$$
$$H_a : \theta_L < \frac{\mathbf{d}_1'\boldsymbol{\beta}}{\mathbf{d}_2'\boldsymbol{\beta}} < \theta_U. \tag{14}$$

Remark 1. *Parameter effects to be included in the model could be selected such that predicted values at each time point correspond to sample means. When the parameters of the model are subject to appropriate restrictions, the model based estimate will not be equal to the traditional estimate, but will still be unbiased. Researchers are encouraged to work with the over-specified model when uncertainity of the presence of certain effects is of concern.*

Estimation of $\boldsymbol{\beta}$ involves the estimation of the unknown variance and covariance components in $\boldsymbol{\Sigma}$, which is approached via the restricted maximum likelihood approach (REML) along with an iterative algorithm to optimize the likelihood function. Though estimation can be performed using other methods like maximum likelihood and minimum variance quadratic unbiased estimation, good properties of the REML approach dictate its use in our analysis. Though the estimate of the AUC is similar to that obtained using the trapezoidal rule, we are able to incorporate within-subject variability leading to more efficient estimates.

2.2.3 Inferences

In order to make inferences on the ratio of means and evaluate for bioequivalence, confidence intervals can be constructed for the purpose of testing the hypotheses outlined in (2). The confidence interval approach states that bioequivalence can be concluded if a 100(1-2α)% for the ratio of the averages fall within the bioequivalence limits of 0.80 and 1.20 (Westlake, 1972). Using our fitted model, the generalized least square estimate of mean areas are obtained using the vector of coefficients $\mathbf{c}$ and the REML estimates of variance components are used to construct a confidence interval. Since exact solutions cannot be obtained, several approximate methods exist to construct a confidence interval for the ratio of two correlated AUC means. A few methods for constructing confidence intervals are outlined below.

By Theorem 3, the finite sample distribution of $\widehat{\Delta}$ can be approximated to be normal with mean μ_T/μ_R and variance $(\widehat{\sigma}^2_{TT} - 2\widehat{\Delta}\widehat{\sigma}_{TR} + \widehat{\Delta}^2\widehat{\sigma}^2_{RR})/(\widehat{\mu}_R)^2$, and therefore a 100(1-2α)% CI can be constructed as

$$\widehat{\Delta} \pm z_\alpha \sqrt{\widehat{Var}(\widehat{\Delta})},$$

where z_α is the standard normal α quantile. This will be referred to as the Delta method through the remainder of this note.

Fieller's method may also be employed to derive the confidence interval for the ratio of two means. Fieller's theorem has been used for constructing the confidence intervals for cost-effectiveness ratios (Willan & O'Brien, 1996), in the study of linkage disequilibrium mapping (Cordell & Elston, 1999) and in the assessment of average bioequivalence (Vuorinen & Tuominen, 1994). A 100(1-2α)% confidence interval for Δ based on Fieller's theorem for a 2×2 crossover design can be constructed by defining a test statistic

$$T = \frac{(\widehat{\mu}_T - \Delta\widehat{\mu}_R)}{\sqrt{\omega(\widehat{\sigma}_{TT} - 2\Delta\widehat{\sigma}_{TR} + \Delta^2\widehat{\sigma}_{RR})}}\ , \tag{15}$$

which follows the t-distribution with degrees of freedom of ν. With $P(T^2 \leq t^2_{(\alpha,\nu)}) = 1-2\alpha$, a $100(1-2\alpha)\%$ CI is constructed by solving the quadratic equation

$$(\widehat{\mu_T} - \Delta\widehat{\mu_R})^2 - t^2_{(\alpha,\nu)}\omega(\widehat{\sigma}^2_{TT} - 2\Delta\widehat{\sigma}^2_{TR} + \Delta^2\widehat{\sigma}^2_{RR}), \tag{16}$$

and is given by

$$\frac{1}{1-G}\left[\left(\frac{\widehat{\mu}_T}{\widehat{\mu}_R} - G\frac{\widehat{\sigma}_{TR}}{\widehat{\sigma}^2_{RR}}\right) \pm \left(t_{(\alpha,\nu)}\frac{\sqrt{\omega\widehat{\sigma}^2_{RR}}}{\widehat{\mu}_R}G^*\right)\right], \tag{17}$$

where, $\omega = \frac{1}{4}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$, G $= t^2_{(\alpha,\nu)}[\omega\frac{\widehat{\sigma}^2_{RR}}{\widehat{\mu}^2_R}]$ and $G^* = (\frac{\widehat{\mu}_T}{\widehat{\mu}_R})^2 + \frac{\widehat{\sigma}^2_{TT}}{\widehat{\sigma}^2_{RR}}(1-G) + \frac{\widehat{\sigma}_{TR}}{\widehat{\sigma}^2_{RR}}(G\frac{\widehat{\sigma}_{TR}}{\widehat{\sigma}^2_{RR}} - 2(\frac{\widehat{\mu}_T}{\widehat{\mu}_R}))$ (Chow & Liu, 2008).

Remark 2. *It may be noted that when the variances and covariances are known, the confidence intervals obtained using Fieller's theorem are exact.*

In implementing our approach, the degrees of freedom is obtained from the fitted model using a procedure such as Kenward-Roger or Satterthwaite's approximation. Fieller's theorem offers an exact confidence interval that are asymmetric about the parameter estimate. Due to the nature of the formula, Fieller's confidence intervals may not always exist, though finite intervals exist in practice. Locke (1984) notes how the bivariate normal distributional assumption may be relaxed and it is sufficient that the sets of differences $AUC_{ik1} - \Delta AUC_{ik2}$ be normally distributed with a common variance. Then, the Fieller's confidence interval will only be approximate in nature. When small samples are encountered and normality assumptions cannot be validated, a bootstrap approach can be used. For further details refer to Chow & Liu (2008) and Locke (1984).

Using the Transformation method, a confidence interval for the ratio of areas can also be obtained by dividing a 100(1-2α)% interval obtained for $\theta = \mu_T - \mu_R$ by the observed reference mean (Chow & Liu, 2004). If (L_1, U_1) is a 100(1-2α)% interval obtained for $\mu_T - \mu_R$ using our proposed method, then the 100(1-2α)%

lower and upper confidence limits for the ratio is given by

$$L_2 = \left(\frac{L_1}{\widehat{\mu}_R} + 1\right) \text{ and } U_2 = \left(\frac{U_1}{\widehat{\mu}_R} + 1\right).$$

This method does not maintain coverage probability since it ignores the variability of the estimate of the reference mean. For comparison of operational characteristics of Fieller's method and the Delta method, see Herson (1975), Cox (1990) and Faraggi et al. (2003).

3 Simulation Study

To examine the properties of the proposed approaches in testing for bioequivalence we conducted a simulation study. Under a 2×2 crossover design, $n = 30$ random samples were randomized equally to two sequences with alternating treatments with multivariate normal mean vectors $\boldsymbol{\mu}_1 = (2, 2.5, 3, 3, 2.5, 2)'$ and $\boldsymbol{\mu}_2 = \gamma\boldsymbol{\mu}_1$, measured at six time points $\mathbf{t} = (1, 2, 3, 4, 5, 6)'$. The value of the coefficient γ ranged from 1.10 to 1.25, where $\gamma = 1.10$ and 1.25 maps to the ratio parameter (Δ) of 0.90 and 0.80 respectively. We assumed no carryover effect for ease of interpretation. To reduce the number of parameters assumed, the longitudinal data with mean $\boldsymbol{\mu}_1$ was generated with the within-period and between-period correlation modeled under the homogeneous first-order autoregressive structure assuming variance homogeneity. The variance σ^2 was set to be equal to 0.1, the within-period correlation parameter ρ was allowed to range from 0.3 to 0.7 and the between-period correlation parameter was set to 0.5. This translates to a common intra-period correlation of ρ and a common between-period correlation of 0.5. Type I error was evaluated at the boundary of the null hypothesis ($\Delta = 0.80$).

Simulations were performed for 5,000 iterations for each combination of ρ and γ. The estimate of interest considered was the ratio of the treatment and reference areas. For our proposed approach, 90% confidence intervals for the estimate were computed using the delta method, Fieller's method and the transformation method as outlined in Section 2.2.3. The traditional test of bioequivalence was performed using Schuirmann's TOST approach (PROC TTEST in SAS). Our approach was implemented using a no-intercept repeated measures model (PROC MIXED in SAS) with covariates sequence, period, treatment, time, as a categorical variable and the interaction between treatment and time. Appropriate contrasts were constructed to obtain the estimate of the ratio of areas and its corresponding standard error. The unstructured covariance structure was considered to model the between-period covariance while compound symmetric, autoregressive structure of order one and the unstructured covariance structure was assumed to model the within-period dependencies. The AIC criterion was also utilized in our simulation study to guide the selection of a covariance structure.

The denominator degrees of freedom was specified using the Kenward-Roger method and a nominal coverage probability of 90% was considered. To compare approaches, inferential properties such as type 1 error, power, coverage probability, confidence interval width, bias and standard error of the estimate under various parameter combinations were assessed.

Our results demonstrated comparable power curves for the transformation and the TOST approach while the power curves for the delta and Fieller's method were slightly higher in the neighbourhood of the null hypotheses for all the assumed covariance structures. For the Fieller and delta method, coverage was occasionally observed to be under the nominal coverage probability of 90% while the transformation and TOST approach exhibited conservativeness in that regard. The delta method was seen to maintain Type I error whereas the TOST approach was consistently conservative in nature. Type I error for Fieller's method and transformation method were inflated and conservative in comparison. For all approaches, estimates were observed to be approximately unbiased and standard errors and confidence interval width were comparable.

To evaluate the effect of missing outcomes, concentrations at each time point were generated under a constant missing probability, p assuming MCAR. Simulations were performed for p ranging from 0% to 20%. Simulation results obtained under various scenarios are summarized in Tables 1 - 4. Power curves for the different methods have been graphically illustrated in Figures 2 - 5.

It can be noted that when missing concentrations were encountered, an increase in p translated to an increase in distance between power curves of the traditional and repeated measures models. The estimates were still noted to be unbiased though the standard error and confidence interval width were comparatively larger for the TOST approach.

All but the repeated measures model assuming an unstructured within-period covariance converged 100% of the time. The convergence of this model ranged from 73 - 83 %. Only models that converged contributed to values used to compute estimates. In practice, a more restrictive covariance structure could be assumed to overcome this issue. In our simulations, the AIC model chose the true covariance structure 93 - 99% of the time. It demonstrated comparable inferential properties with those obtained from the true covariance structure indicating that the model selection can be often guided by the AIC without experiencing loss of power or precision.

4 Application: Asthma Study

We now apply the methods developed in this paper to the data associated with the study described in Graff-Lonnevig & Browaldh (1990). The aim of this double-blind crossover study was to evaluate the twelve hours' bronchodilating effect of inhaled Formoterol (F) versus Salbutamol (S) in children with asthma. Fourteen

children with perennial asthma were enrolled and peak expiratory flow rate (PEFR) was measured before the test dose was given and 20, 40, 60, and 90 min after the dose, and then every hour up to 8 hr in the hospital. Further measurements were recorded at home after 10, 11 and 12 hr. Since the AUC is considered a measure of drug exposure in pharmacokinetic studies, we use this example to compare the bioavailability of formoterol versus salbutamol. A higher PEFR indicates greater effectiveness of the drug and hence a higher area under the PEFR versus time curve indicates a probability for greater bioavailability.

The subject profiles presented in Figures 6(a) and 6(b) for each sequence in the study enable us to understand the data further. A lower AUC for the Salbutamol group versus Formoterol in both sequences can be observed. Observations from the Formoterol-Salbutamol sequence are also seen to be more variable when compared to the Salbutamol-Formoterol sequence. A sequence by period graph (Figure 6(c)) generated indicates a higher mean AUC for the Formoterol treatment in both periods. From the PEFR versus time curve in Figure 6(d), PEFR for Formoterol is seen to be higher than Salbutamol in the terminal phase of the curve. The PEFR of Salbutamol is seen to peak in the early phase of the outcome-time curve and then decay at a rapid rate compared to Formoterol which exhibits a slower and steady decay rate. Salbutamol is also seen to have a higher peak and higher time to reach the peak than Formoterol.

Assuming no carry over effects, the model based on the proposed approach is a function of treatment, time as a categorical variable, period, sequence and the interaction between treatment and time. From the fitted model, estimated mean flow rate at the fifteen time points are obtained and the mean AUC for each intervention is estimated using contrasts specified in Table 5. Bioequivalence is evaluated by constructing 90% confidence interval for the areas using the transformation, delta and Fieller's method as detailed in Section 2.2. Results from our analysis using the traditional and the proposed approach are provided in Table 6. The 90% confidence interval of (1.04260,1.11063) obtained from the traditional method provided evidence to conclude Formoterol to be bioequivalent to Salbutamol. Confidence intervals obtained using the proposed approach also yielded similar results.

5 Discussion and Conclusion

In this paper we introduced a linear model-based approach to make inferences on AUC means and evaluate bioequivalence. Properties of our approach were evaluated using simulations studies. The proposed method demonstrated better inferential properties when compared to the traditional approach, markedly noticeable in the presence of missing data.

The flexibility of our approach lies in that various experimental designs can be accomodated. Using our proposed approach, the impact of covariates or baseline values on bioequivalence can be investigated. When

treatment differences are a function of covariates such as gender, age, etc., the covariate may be accounted for by its inclusion in the linear model. The adjusted mean AUC can then be obtained as a function of the vector of estimated conditional means at each time point. It also allows the covariates to vary over clusters and response to vary over clusters after controlling for covariates.

Depending on the effects included in the model, the estimate of mean AUC computed using our approach may be a function of the predicted values not equal to the sample means. Though the parameter is redefined and might have a slightly different yet clinically significant interpretation, the estimate remains unbiased.

Although we have modeled the within and between-subject covariances using the Kronecker product structure, it can also be accomplished with a mixed-model formulation. Random coefficients may be fit with a different variance matrix for observations in each period providing an added advantage that it facilitates higher order crossover designs with more than two periods.

Bioequivalence parameters such as the AUC are commonly observed to be log-normally distributed while in this paper, our estimates are based on the normal distribution. Future research aims to relax the above mentioned assumption. This framework may be applied to a variety of sampling designs observed in vivo animal studies such as batch, serial and sparse sampling. Non-parametric alternatives such as a permutation test may be considered when the distribution of the test statistic under the null is suspect, while a bootstrap version can be considered for extremely small sample sizes.

Table 1: Comparison of power curves and coverage for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$ and $\sigma^2 = 0.1$. Simulations were performed for 5,000 iterations.

p	γ	Power(Coverage): Delta Method					% convergent (UN)
		UN	CS	AR(1)	AIC	TOST	
$\rho = 0.3$							
0	1.1	1(0.9027)	1(0.8888)	1(0.8984)	1(0.9)	1(0.9184)	74.6
	1.15	1(0.9061)	1(0.894)	1(0.905)	1(0.9045)	1(0.9276)	72.86
	1.17	1(0.901)	1(0.8886)	1(0.9)	1(0.8999)	1(0.9246)	73.32
	1.2	0.9698(0.9032)	0.9772(0.8898)	0.9744(0.8984)	0.9728(0.8954)	0.9576(0.9254)	73.58
	1.23	0.4054(0.8951)	0.4378(0.8844)	0.4158(0.8928)	0.4144(0.8902)	0.3448(0.9242)	73.56
	1.25	0.0503(0.903)	0.0578(0.8872)	0.051(0.9004)	0.0511(0.9008)	0.0356(0.9298)	74.4
0.2	1.1	1(0.9003)	1(0.8826)	1(0.8908)	1(0.8946)	0.9992(0.9142)	73.22
	1.15	1(0.9017)	1(0.8902)	1(0.8962)	1(0.8952)	0.9506(0.9166)	74.08
	1.17	0.9997(0.9052)	0.9998(0.892)	0.9998(0.8998)	0.9997(0.9001)	0.8522(0.929)	73.64
	1.2	0.9412(0.894)	0.9512(0.8852)	0.9486(0.8956)	0.9474(0.8956)	0.49(0.926)	74.12
	1.23	0.3492(0.8981)	0.37(0.8822)	0.3584(0.8896)	0.3592(0.8903)	0.1198(0.929)	74.22
	1.25	0.0529(0.8918)	0.0552(0.8852)	0.0504(0.894)	0.0523(0.8921)	0.0276(0.9348)	73
$\rho = 0.7$							
0	1.1	1(0.8965)	1(0.8836)	1(0.8906)	1(0.8935)	1(0.911)	81.54
	1.15	0.9995(0.8907)	0.9996(0.88)	0.9996(0.893)	0.9995(0.8909)	0.9992(0.914)	81.76
	1.17	0.9932(0.8987)	0.9938(0.8842)	0.9938(0.8954)	0.9941(0.8962)	0.985(0.9232)	81.9
	1.2	0.8052(0.9029)	0.8196(0.8828)	0.813(0.8914)	0.812(0.8944)	0.7506(0.9248)	83.18
	1.23	0.2541(0.8976)	0.2766(0.8832)	0.2614(0.8932)	0.2625(0.8937)	0.204(0.9238)	81.08
	1.25	0.048(0.8983)	0.0562(0.8866)	0.0496(0.8968)	0.05(0.8956)	0.034(0.931)	82
0.2	1.1	1(0.8965)	1(0.8844)	1(0.8908)	1(0.8916)	0.9978(0.915)	81
	1.15	0.9983(0.9006)	0.9988(0.8882)	0.9988(0.8968)	0.9985(0.8996)	0.9242(0.9292)	81.06
	1.17	0.9867(0.8985)	0.9872(0.8842)	0.9866(0.8938)	0.9874(0.8948)	0.7824(0.926)	81.02
	1.2	0.7763(0.8966)	0.7874(0.8842)	0.7822(0.8894)	0.7832(0.8904)	0.424(0.9252)	80.46
	1.23	0.2456(0.903)	0.2632(0.8878)	0.2524(0.8956)	0.2552(0.8956)	0.1236(0.9256)	81.44
	1.25	0.0474(0.9034)	0.0564(0.8876)	0.052(0.8968)	0.0501(0.8979)	0.0342(0.9316)	82.64

Table 2: Comparison of power curves and coverage for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$ and $\sigma^2 = 0.1$. Simulations were performed for 5,000 iterations.

p	γ	Power(Coverage): Fieller Method					% convergent (UN)
		UN	CS	AR(1)	AIC	TOST	
$\rho = 0.3$							
0	1.1	1(0.8906)	1(0.8764)	1(0.8864)	1(0.8874)	1(0.9184)	74.6
	1.15	1(0.8855)	1(0.8784)	1(0.8866)	1(0.8839)	1(0.9276)	72.86
	1.17	1(0.883)	1(0.866)	1(0.8776)	1(0.8778)	1(0.9246)	73.32
	1.2	0.975(0.8804)	0.9804(0.8668)	0.9794(0.8746)	0.9777(0.8722)	0.9576(0.9254)	73.58
	1.23	0.4492(0.8681)	0.4728(0.8574)	0.467(0.8636)	0.466(0.8613)	0.3448(0.9242)	73.56
	1.25	0.0659(0.8766)	0.0738(0.8584)	0.0654(0.8716)	0.0672(0.8696)	0.0356(0.9298)	74.4
0.2	1.1	1(0.9036)	1(0.8892)	1(0.8952)	1(0.8984)	0.9992(0.9142)	73.22
	1.15	1(0.8988)	1(0.8872)	1(0.893)	1(0.892)	0.9506(0.9166)	74.08
	1.17	0.9997(0.9033)	0.9998(0.8912)	0.9998(0.8948)	0.9997(0.8946)	0.8522(0.929)	73.64
	1.2	0.9414(0.8891)	0.9524(0.8794)	0.9506(0.8864)	0.949(0.8859)	0.49(0.926)	74.12
	1.23	0.36(0.8946)	0.3812(0.8792)	0.376(0.8788)	0.3802(0.8793)	0.1198(0.929)	74.22
	1.25	0.057(0.8822)	0.0594(0.8742)	0.0588(0.879)	0.0605(0.877)	0.0276(0.9348)	73
$\rho = 0.7$							
0	1.1	1(0.8864)	1(0.8718)	1(0.8782)	1(0.882)	1(0.911)	81.54
	1.15	0.9995(0.8784)	0.9996(0.8616)	0.9996(0.8738)	0.9995(0.8721)	0.9992(0.914)	81.76
	1.17	0.9946(0.8823)	0.9948(0.8666)	0.995(0.878)	0.9951(0.8781)	0.985(0.9232)	81.9
	1.2	0.8324(0.8815)	0.845(0.8626)	0.8408(0.872)	0.8401(0.8738)	0.7506(0.9248)	83.18
	1.23	0.2972(0.8705)	0.3154(0.8578)	0.3036(0.867)	0.3041(0.8661)	0.204(0.9238)	81.08
	1.25	0.0637(0.8717)	0.0724(0.8592)	0.0652(0.87)	0.0654(0.868)	0.034(0.931)	82
0.2	1.1	1(0.8928)	1(0.8788)	1(0.8848)	1(0.8862)	0.9978(0.915)	81
	1.15	0.9995(0.8882)	0.999(0.8782)	0.9992(0.8844)	0.999(0.886)	0.9242(0.9292)	81.06
	1.17	0.9891(0.883)	0.9884(0.8738)	0.9892(0.8796)	0.9899(0.8795)	0.7824(0.926)	81.02
	1.2	0.7957(0.8817)	0.807(0.8692)	0.8078(0.8704)	0.8079(0.873)	0.424(0.9252)	80.46
	1.23	0.2773(0.8819)	0.2914(0.865)	0.2872(0.8758)	0.2903(0.8752)	0.1236(0.9256)	81.44
	1.25	0.0595(0.8778)	0.0694(0.8636)	0.0664(0.8712)	0.0639(0.8715)	0.0342(0.9316)	82.64

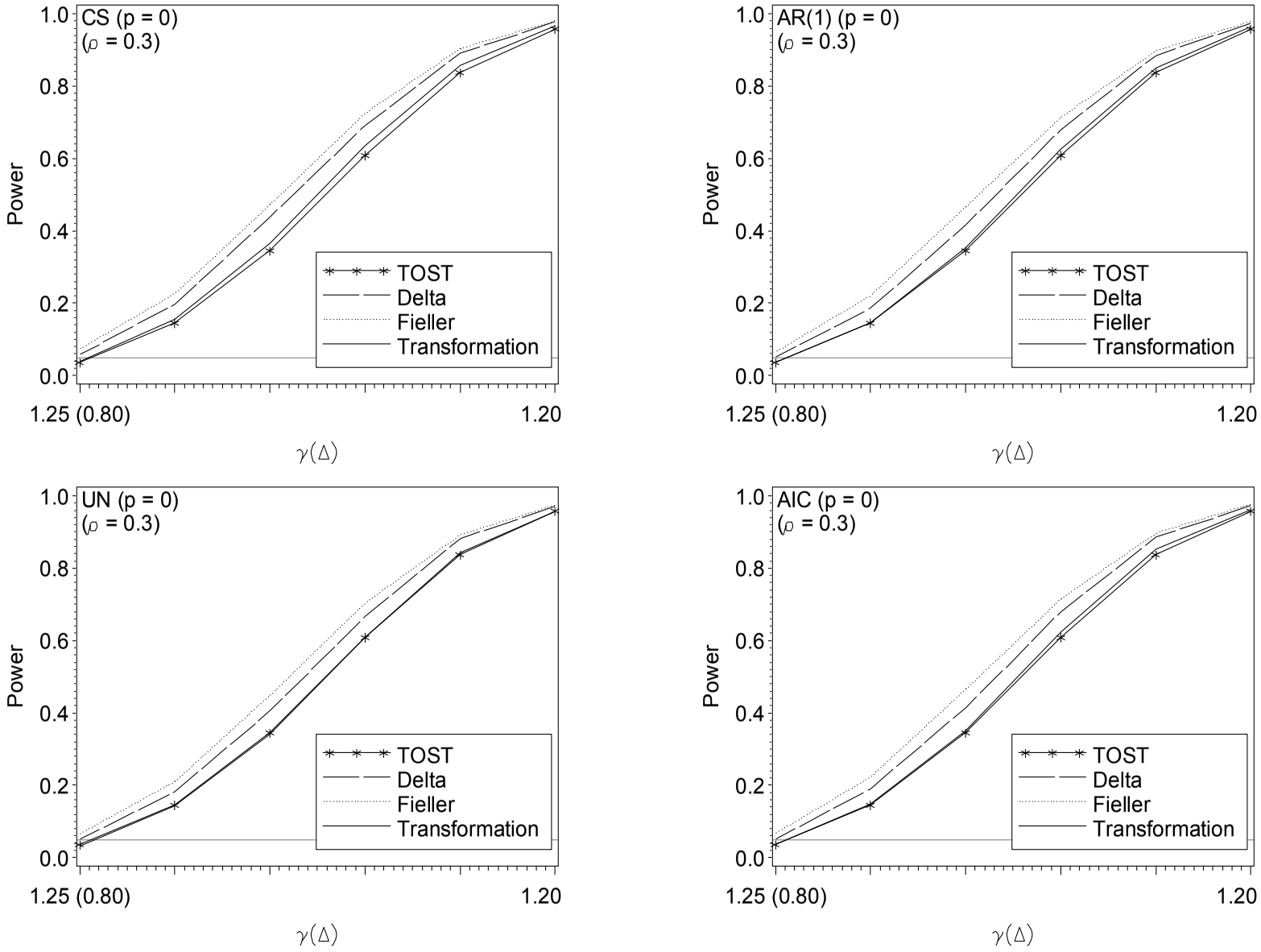


Figure 2: Comparison of power for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.3$ and p = 0%. Simulations were performed for 5,000 iterations.

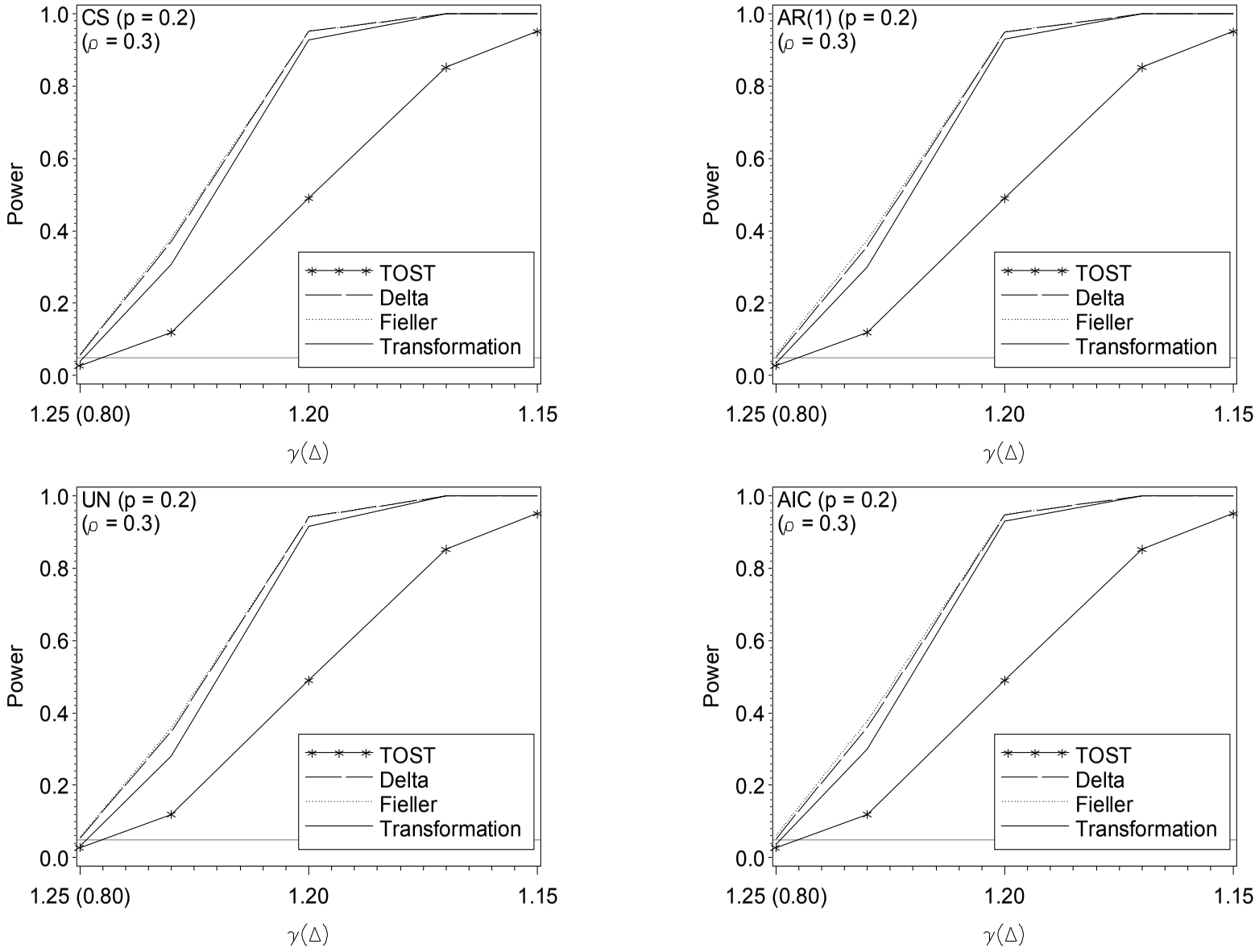


Figure 3: Comparison of power for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.3$ and p = 20%. Simulations were performed for 5,000 iterations.

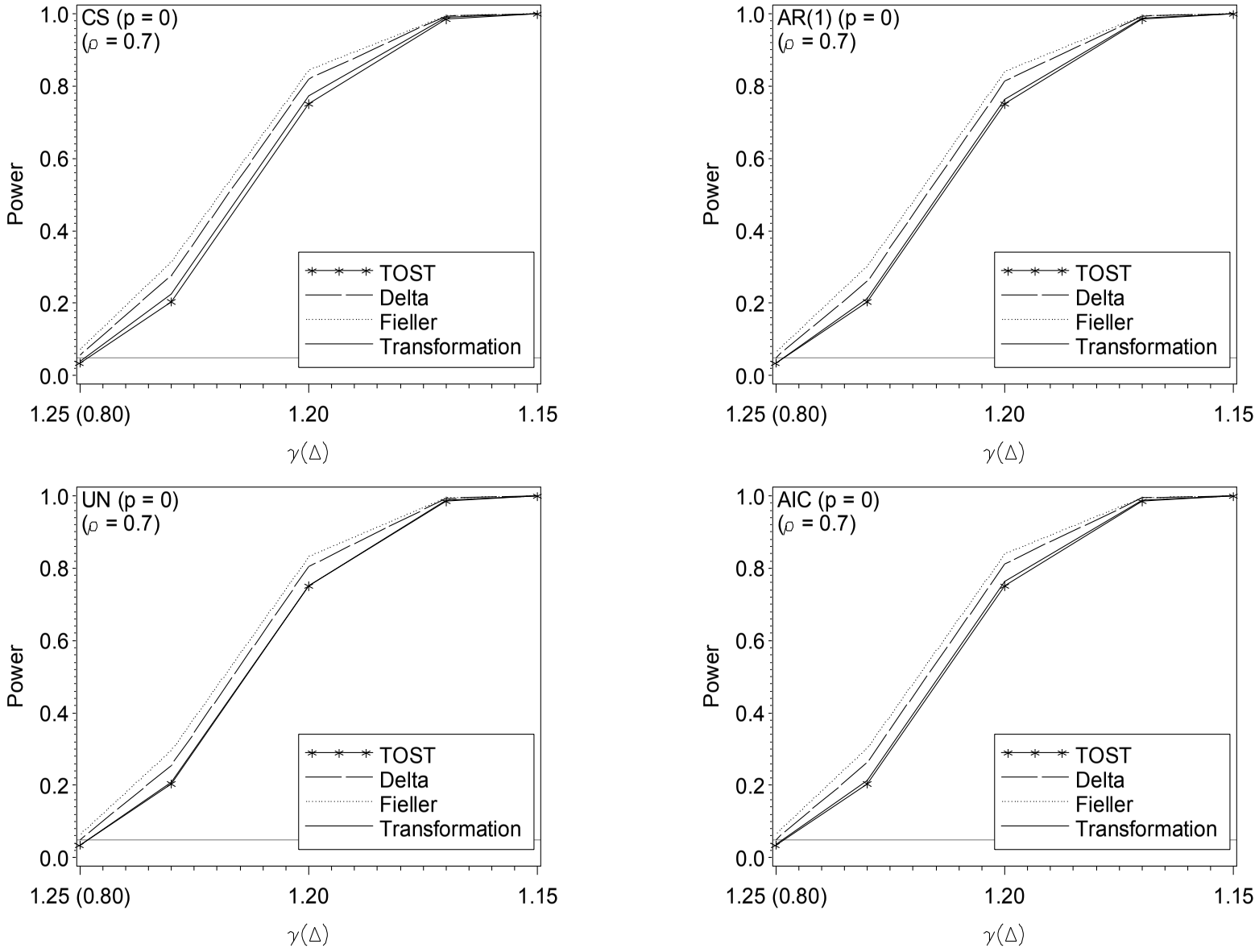


Figure 4: Comparison of power for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.7$ and p = 0%. Simulations were performed for 5,000 iterations.

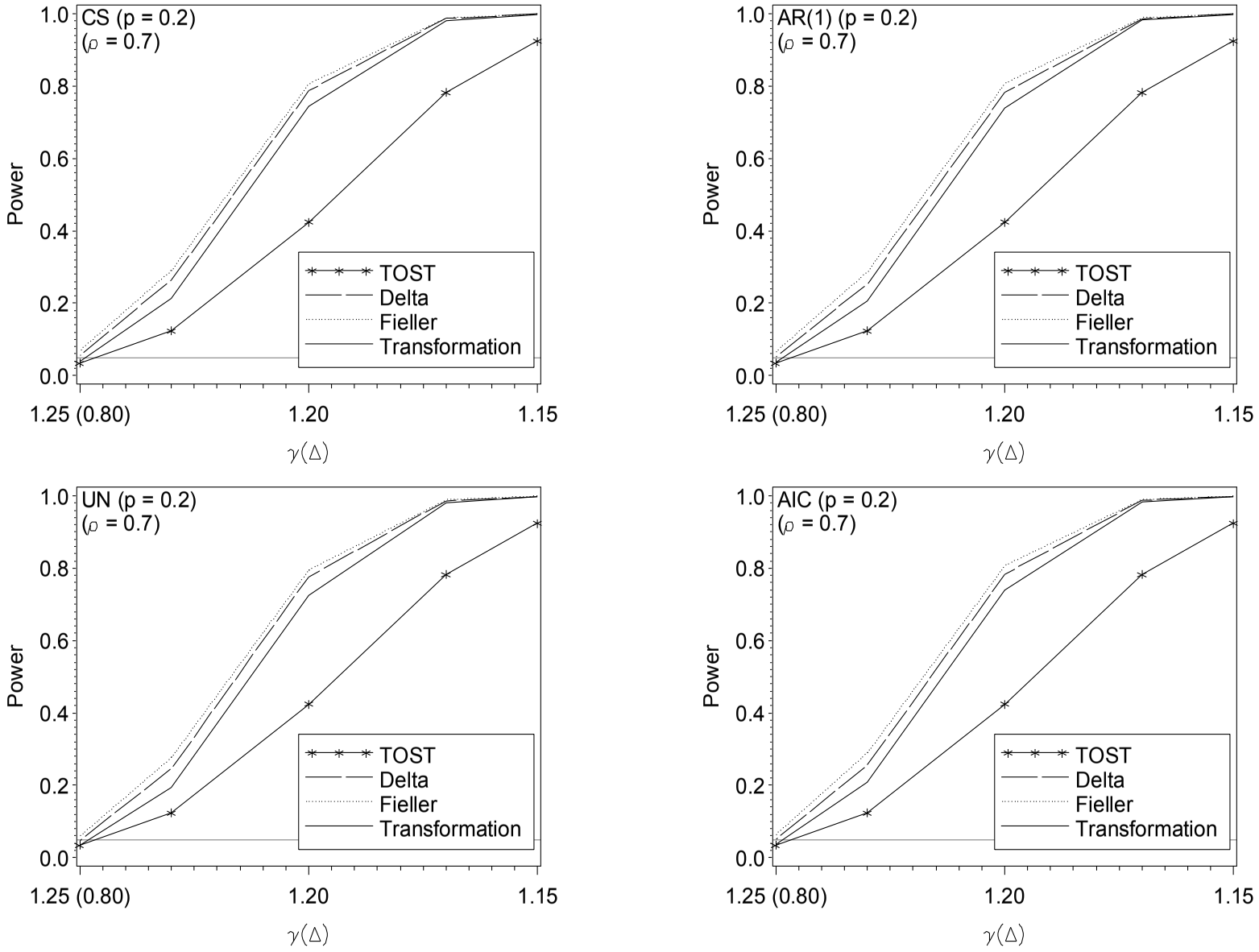
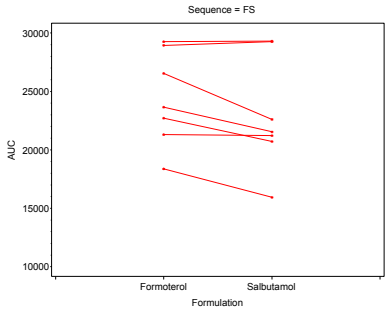


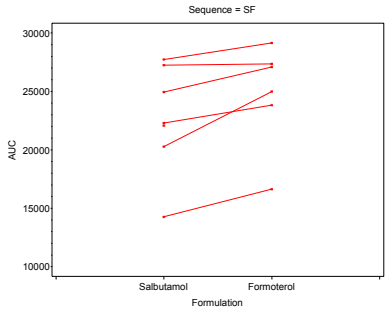
Figure 5: Comparison of power for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.7$ and p = 20%. Simulations were performed for 5,000 iterations.

Table 3: Comparison of power curves and coverage for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$ and $\sigma^2 = 0.1$. Simulations were performed for 5,000 iterations.

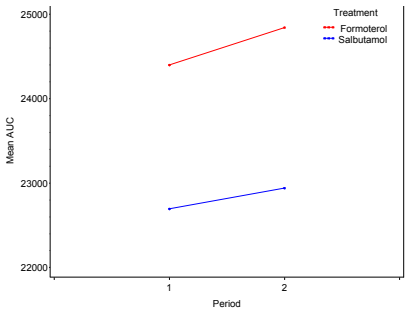
p	γ	Power(Coverage): Transformation Method					% convergent (UN)
		UN	CS	AR(1)	AIC	TOST	
$\rho = 0.3$							
0	1.1	1(0.9233)	1(0.9058)	1(0.9144)	1(0.9161)	1(0.9184)	74.6
	1.15	1(0.9286)	1(0.9166)	1(0.9256)	1(0.9256)	1(0.9276)	72.86
	1.17	1(0.9304)	1(0.9186)	1(0.926)	1(0.9272)	1(0.9246)	73.32
	1.2	0.9571(0.9307)	0.9658(0.9204)	0.9636(0.9296)	0.9625(0.9274)	0.9576(0.9254)	73.58
	1.23	0.3399(0.9266)	0.3662(0.9172)	0.3526(0.9266)	0.3502(0.9228)	0.3448(0.9242)	73.56
	1.25	0.0325(0.9344)	0.0388(0.9218)	0.0356(0.9298)	0.0358(0.9317)	0.0356(0.9298)	74.4
0.2	1.1	1(0.9186)	1(0.902)	1(0.9096)	1(0.9115)	0.9992(0.9142)	73.22
	1.15	1(0.9239)	1(0.913)	1(0.9202)	1(0.9166)	0.9506(0.9166)	74.08
	1.17	0.9997(0.9261)	0.9998(0.92)	0.9998(0.9254)	0.9997(0.9267)	0.8522(0.929)	73.64
	1.2	0.9145(0.9231)	0.927(0.914)	0.9296(0.919)	0.929(0.9185)	0.49(0.926)	74.12
	1.23	0.2813(0.9313)	0.308(0.9174)	0.2996(0.9204)	0.301(0.9221)	0.1198(0.929)	74.22
	1.25	0.0321(0.9318)	0.0392(0.9194)	0.035(0.928)	0.0359(0.9271)	0.0276(0.9348)	73
$\rho = 0.7$							
0	1.1	1(0.9173)	1(0.9032)	1(0.908)	1(0.9105)	1(0.911)	81.54
	1.15	0.999(0.919)	0.9996(0.906)	0.9996(0.9148)	0.9995(0.9132)	0.9992(0.914)	81.76
	1.17	0.9868(0.927)	0.9896(0.9148)	0.9884(0.9228)	0.9888(0.9241)	0.985(0.9232)	81.9
	1.2	0.7507(0.9293)	0.7738(0.9162)	0.7638(0.9236)	0.7629(0.9262)	0.7506(0.9248)	83.18
	1.23	0.2072(0.9299)	0.224(0.9148)	0.2128(0.9212)	0.2134(0.9238)	0.204(0.9238)	81.08
	1.25	0.0354(0.9283)	0.0388(0.9208)	0.0354(0.927)	0.0366(0.9246)	0.034(0.931)	82
0.2	1.1	1(0.9205)	1(0.9056)	1(0.9118)	1(0.9114)	0.9978(0.915)	81
	1.15	0.9978(0.9257)	0.9982(0.9146)	0.9984(0.921)	0.998(0.923)	0.9242(0.9292)	81.06
	1.17	0.9798(0.9257)	0.9806(0.9116)	0.9826(0.9172)	0.9835(0.9198)	0.7824(0.926)	81.02
	1.2	0.7248(0.9279)	0.745(0.9176)	0.7398(0.9218)	0.7402(0.9215)	0.424(0.9252)	80.46
	1.23	0.193(0.9354)	0.2128(0.9206)	0.2046(0.9276)	0.208(0.9288)	0.1236(0.9256)	81.44
	1.25	0.0353(0.931)	0.0394(0.9224)	0.0368(0.9294)	0.0365(0.9288)	0.0342(0.9316)	82.64



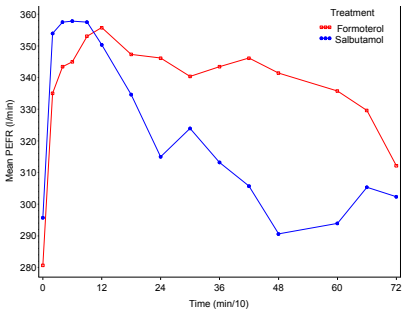
(a) Subject profiles for seq 1



(b) Subject profiles for seq 2



(c) Sequence by period AUC means



(d) Mean PEFR vs time curve

Table 4: Comparison of Bias, Std and CI width for the test of BE. Data was generated under the AR(1) assumption within period and between period correlation of 0.5 with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$ and $\sigma^2 = 0.1$. Simulations were performed for 5,000 iterations.

						CI width									
		Bias (std)				Delta			Fieller			Transformation			TOST
p	γ	UN	CS	AR(1)	TOST	UN	CS	AR(1)	UN	CS	AR(1)	UN	CS	AR(1)	
$\rho = 0.3$															
0	1.1	0(0.0107)	0(0.0107)	0(0.0107)	0(0.0107)	0.0355	0.034	0.0349	0.0349	0.0333	0.034	0.0378	0.0361	0.0369	0.0376
	1.15	0(0.0099)	0(0.0098)	0(0.0098)	0(0.0098)	0.0334	0.0319	0.0328	0.0322	0.0308	0.0314	0.0361	0.0345	0.0353	0.036
	1.17	0(0.0097)	0(0.0097)	0(0.0097)	0(0.0097)	0.0325	0.0311	0.032	0.0312	0.0298	0.0304	0.0355	0.0339	0.0346	0.0353
	1.2	0(0.0095)	0(0.0094)	0(0.0094)	0(0.0094)	0.0315	0.0302	0.031	0.0299	0.0286	0.0292	0.0346	0.0331	0.0338	0.0344
	1.23	0(0.0093)	0(0.0092)	0(0.0092)	0(0.0092)	0.0305	0.0292	0.03	0.0287	0.0275	0.0281	0.0338	0.0323	0.033	0.0336
	1.25	0(0.0088)	0(0.0089)	0(0.0089)	0(0.0089)	0.0298	0.0285	0.0294	0.028	0.0267	0.0273	0.0332	0.0317	0.0324	0.033
0.2	1.1	0(0.012)	0(0.0119)	0(0.0118)	0(0.0209)	0.0396	0.0378	0.0384	0.0409	0.039	0.0392	0.0422	0.0402	0.0406	0.0722
	1.15	0(0.0112)	0(0.0111)	0(0.011)	0(0.0198)	0.0372	0.0356	0.0361	0.0376	0.036	0.0361	0.0404	0.0385	0.0389	0.0703
	1.17	0(0.0107)	0(0.0107)	0(0.0106)	0(0.0191)	0.0364	0.0348	0.0353	0.0366	0.035	0.0352	0.0397	0.0379	0.0383	0.0695
	1.2	0(0.0107)	0(0.0106)	0(0.0105)	0(0.0186)	0.0351	0.0336	0.0341	0.035	0.0335	0.0336	0.0386	0.0369	0.0373	0.0683
	1.23	0(0.0101)	0(0.0101)	0(0.0101)	0(0.0181)	0.0339	0.0325	0.033	0.0335	0.0321	0.0323	0.0376	0.036	0.0364	0.0674
	1.25	0(0.0102)	0(0.01)	0(0.01)	0(0.018)	0.0333	0.0318	0.0323	0.0328	0.0313	0.0314	0.0372	0.0354	0.0358	0.0668
$\rho = 0.7$															
0	1.1	0(0.015)	0(0.015)	0(0.015)	0(0.015)	0.0496	0.0478	0.049	0.0486	0.0469	0.0478	0.0528	0.0509	0.0519	0.0528
	1.15	0(0.0141)	0(0.014)	0(0.014)	0(0.014)	0.0465	0.0448	0.0459	0.0447	0.0431	0.044	0.0503	0.0485	0.0495	0.0503
	1.17	0(0.0137)	0(0.0137)	0(0.0137)	0(0.0137)	0.0456	0.044	0.045	0.0436	0.0421	0.043	0.0496	0.0479	0.0489	0.0497
	1.2	0(0.0134)	0(0.0134)	0(0.0134)	0(0.0134)	0.0442	0.0426	0.0436	0.0421	0.0405	0.0413	0.0485	0.0467	0.0477	0.0484
	1.23	0(0.0129)	0(0.0128)	0(0.0128)	0(0.0128)	0.0426	0.0411	0.0421	0.0401	0.0387	0.0394	0.0472	0.0454	0.0464	0.0473
	1.25	0(0.0127)	0(0.0126)	0(0.0126)	0(0.0126)	0.0417	0.0402	0.0413	0.0391	0.0377	0.0385	0.0464	0.0447	0.0457	0.0465
0.2	1.1	0(0.0154)	0(0.0155)	0(0.0154)	0.001(0.0229)	0.0514	0.0498	0.0504	0.0513	0.0499	0.05	0.0548	0.0531	0.0535	0.0804
	1.15	0(0.0143)	0(0.0144)	0(0.0143)	0(0.0212)	0.0481	0.0467	0.0472	0.0472	0.0459	0.0461	0.0522	0.0506	0.051	0.0776
	1.17	0(0.0143)	0(0.0144)	0(0.0143)	0(0.0212)	0.0471	0.0456	0.0461	0.0458	0.0446	0.0446	0.0514	0.0498	0.0501	0.0769
	1.2	0(0.0137)	0(0.0139)	0(0.0138)	0(0.0207)	0.0455	0.0441	0.0446	0.044	0.0428	0.0429	0.0501	0.0485	0.0489	0.0753
	1.23	0(0.0132)	0(0.0132)	0(0.0131)	0(0.0204)	0.0441	0.0427	0.0432	0.0423	0.041	0.0411	0.049	0.0473	0.0477	0.0742
	1.25	0(0.013)	0(0.0131)	0(0.013)	0.001(0.0199)	0.0432	0.0419	0.0423	0.0412	0.0401	0.0401	0.0482	0.0466	0.047	0.0735

Table 5: Vector of coefficients ($\mathbf{c}$) to test for bioequivalence

c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}
2	3	4	5	6	9	12	12	12	12	12	18	18	12	6

Table 6: Summary of results from the Asthma Study

Method	*Estimate* (*ratio*)	90% *CI*	*Conclusion*
TOST	1.07661	(1.04260, 1.11063)	Bioequivalent
Transformation	1.01397	(0.88163, 1.14632)	Bioequivalent
Delta	1.01397	(0.88591, 1.14203)	Bioequivalent
Fieller	1.01397	(0.96615, 1.06306)	Bioequivalent

References

Anderson, S., & Hauck, W. W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics - Theory and Methods*, *12*(23), 2663–2692.

Benet, L. Z., & Goyan, J. E. (1995). Bioequivalence and narrow therapeutic index drugs. *Pharmacotherapy*, *15*(4), 433–440.

Berger, R. L., & Hsu, J. C. (1996). Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science*, *11*(4).

Brown, L. D., Hwang, J. T. G., & Munk, A. (1997). An Unbiased Test for the Bioequivalence Problem. *The Annals of Statistics*, *25*(6).

Cao, L., & Mathew, T. (2008). A simple numerical approach towards improving the two one-sided test for average bioequivalence. *Biometrical journal. Biometrische Zeitschrift*, *50*(2), 205–211.

Chinchilli, V. M., & Elswick, R. K. (1997). The multivariate assessment of bioequivalence. *Journal of Biopharmaceutical Statistics*, *7*(1), 113–123.

Chow, S. C., & Liu, J. P. (2004). *Design and analysis of clinical trials: concepts and methodologies*. Wiley-Interscience; 2 edition (December 8, 2003).

Chow, S. C., & Liu, J. P. (2008). *Design and analysis of bioavailability and bioequivalence studies.*. CRC Press.

Chow, S. C., & Shao, J. (1990). An Alternative Approach for the Assessment of Bioequivalence Between Two Formulations of a Drug. *Biometrical Journal*, *32*(8), 969–976.

Cordell, H. J., & Elston, R. C. (1999). Fieller's theorem and linkage disequilibrium mapping. *Genetic epidemiology*, *17*(4), 237–252.

Cox, C. (1990). Fieller's Theorem, the Likelihood and the Delta Method. *Biometrics*, *46*(3).

Faraggi, D., Izikson, P., & Reiser, B. (2003). Confidence intervals for the 50 per cent response dose. *Statist. Med.*, *22*(12), 1977–1988.

FDA (2001). Guidance for industry: Statistical approaches to establishing bioequivalence. U.S. department of health and human services, food and drug administration, center for drug evaluation and research (CDER).

FDA (2008). *Code of Federal Regulations, Title 21, Part 320. Bioavailability and bioequivalence requirements. US Food and Drug Administration Web site (April 2008). http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm? CFRPart=320 Accessed January 19, 2010.*.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of Misspecifying the First-Level Error Structure in Two-Level Models of Change. *Multivariate Behavioral Research*, *37*(3), 379–403.

Fieller, E. C. (1954). Some Problems in Interval Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, *16*(2), 175–185.

Gao, F., Thompson, P., Xiong, C., & Miller, J. P. (2006). Analyzing multivariate longitudinal data using SAS. In *Proceedings of the Thirty-first Annual SAS Users Group International Conference*, (pp. 187–731).

Ghosh, P., & Rosner, G. L. (2007). A semi-parametric Bayesian approach to average bioequivalence. *Statistics in Medicine*, *26*(6), 1224–1236.

Graff-Lonnevig, V., & Browaldh, L. (1990). Twelve hours' bronchodilating effect of inhaled formoterol in children with asthma: a double-blind cross-over study versus salbutamol. *Clinical & Experimental Allergy*, *20*(4), 429–432.

Herson, J. (1975). Fieller's theorem vs. The delta method for significance intervals for ratios. *Journal of Statistical Computation and Simulation*, *3*(3), 265–274.

Jones, B., & Kenward, M. G. (2003). *Design and Analysis of Cross-over Trials*. Chapman & Hall/CRC: London, Boca Raton.

Julious, S. A., & Debarnot, C. A. (2000). Why are pharmacokinetic data summarized by arithmetic means? *Journal of biopharmaceutical statistics*, *10*(1), 55–71.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Simulation and Computation*, *27*(3), 591–604.

Lacey, L. F., Keene, O. N., Pritchard, J. F., & Bye, A. (1997). Common noncompartmental pharmacokinetic variables: are they normally or log-normally distributed? *Journal of Biopharmaceutical Statistics*, *7*(1), 171–178.

Liu, J. P., & Weng, C. S. (1994). Evaluation of log-transformation in assessing bioequivalence. *Communications in Statistics - Theory and Methods*, *23*(2), 421–434.

Liu, J. P., & Weng, C. S. (1995). Bias of two one-sided tests procedures in assessment of bioequivalence. *Statist. Med.*, *14*(8), 853–861.

Locke, C. S. (1984). An exact confidence interval from untransformed data for the ratio of two formulation means. *Journal of Pharmacokinetics and Pharmacodynamics*, *12*(6), 649–655.

Marston, S. A., & Polli, J. E. (1997). Evaluation of Direct Curve Comparison Metrics Applied to Pharmacokinetic Profiles and Relative Bioavailability and Bioequivalence. *Pharmaceutical Research*, *14*(10), 1363–1369.

Mauger, D. T., & Chinchilli, V. M. (2000a). An alternative index for assessing profile similarity in bioequivalence trials. *Statist. Med.*, *19*(20), 2855–2866.

Mauger, D. T., & Chinchilli, V. M. (2000b). Profile Similarity in Bioequivalence Trials. *Sankhy: The Indian Journal of Statistics, Series B*, *62*(1).

Metzler, C. M. (1974). Bioavailability – A Problem in Equivalence. *Biometrics*, *30*(2).

Rescigno, A. (1992). Bioequivalence. *Pharmaceutical Research*, *9*(7), 925–928.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics*, *15*(6), 657–680.

Senn, S. (2002). *Cross-over Trials in Clinical Research*. John Wiley & Sons.

Vuorinen, J., & Tuominen, J. (1994). Fieller's confidence intervals for the ratio of two means in the assessment of average bioequivalence from crossover data. *Statistics in medicine*, *13*(23-24), 2531–2545.

Wang, W., Gene Hwang, J. T., & Dasgupta, A. (1999). Statistical tests for multivariate bioequivalence. *Biometrika*, *86*(2), 395–402.

Westlake, W. J. (1972). Use of confidence intervals in analysis of comparative bioavailability trials. *Journal of Pharmaceutical Sciences*, *61*(8), 1340–1341.

Westlake, W. J. (1976). Symmetrical confidence intervals for bioequivalence trials. *Biometrics*, *32*(4), 741–744.

Wilding, G. E., Chandrasekhar, R., & Hutson, A. D. (2011). A new linear model-based approach for inferences about the mean area under the curve. Tech. rep., University at Buffalo (Available at http://sphhp.buffalo.edu/biostat/research/techreports/UB_Biostatistics_TR1104.pdf).

Willan, A. R., & O'Brien, B. J. (1996). Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health economics*, *5*(4), 297–305.

Yee, K. F. (1986). The Calculation of Probabilities in Rejecting Bioequivalence. *Biometrics*, *42*(4).