An Exact Method for Power Calculation for a Three-arm Clinical Endpoint Bioequivalence Study

Aotian Yang[1], Wanjie Sun[2]

[1] The George Washington University, 2121 I Street, NW, Washington, DC, 20052, United States

[2] U.S. Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD, 20993, United States

Correspondence: Wanjie Sun, U.S. Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD, 20993, United States. E-mail: wanjie.sun@fda.hhs.gov

Abstract

A clinical endpoint bioequivalence (BE) study aims to establish BE between a generic drug (TEST) and an innovator drug (REF). A placebo (PLB) is usually included to demonstrate the sensitivity of the study. BE is established if TEST is shown to be superior to PLB, REF superior to PLB, and TEST equivalent to REF. Therefore, an overall BE test for a clinical endpoint BE study is composed of two superiority tests (TEST vs. PLB and REF vs. PLB) and one equivalence test (TEST vs. REF).

Previously, Chang et al (2014) calculated the sample size and power for an overall BE test based on one superiority test (TEST vs. PLB) and an equivalence test (TEST vs. REF) using the joint distribution of sample means and sample variances because 'it is not easy to derive the sample size based on the multivariate t-distribution' (we call this a Z-ChiSquare method). In this paper, we propose an exact method to calculate the power and sample size for an overall BE test based on two superiority tests (TEST vs. PLB, REF vs. PLB) and one equivalence test (TEST vs. REF) using a multivariate non-central t distribution directly, which we call an Exact-t method. We also extended the Z-ChiSquare method to an overall BE test with two superiority tests and one equivalence test, rather than one superiority and one equivalence test as in Chang et al's paper.

Simulation shows that our proposed Exact-t method is computationally more efficient than the Z-ChiSquare method without self-writing codes to numerically calculate the conditional expectation of a multivariate normal distribution conditional upon a truncated Chi-Square distribution. When sample size is small, the Exact-t method generates more accurate results than the Z-ChiSquare method.

The Exact-t method is recommended when calculating power and determining sample size for a three-arm clinical endpoint BE study.

Keywords: bioequivalence test, three-arm four-step clinical trial, Z-ChiSquare method, exact-t

1. Introduction

Most generic drugs are systematic drugs where drugs are intended to be absorbed into the bloodstream. For systematically absorbed drugs, pharmacokinetic (PK) studies are usually used to establish BE between a generic drug (TEST) and an innovator drug (REF) by assessing whether TEST and REF containing the same active moiety (chemically equivalent) in the same dosage form (pharmaceutically equivalent) reach the system circulation at an equivalent relative rate and extent in healthy subjects (Davit et al, 2009). For locally acting generic drugs, pharmacokinetic (PK) studies are generally not adequate for establishing bioequivalence and clinical endpoint BE studies are usually requested for establishing BE for these drugs. Clinical endpoint BE studies aim to establish bioequivalence by performing a comparison of clinical effects based on pre-defined clinical endpoints between TEST and RLD (Yu & Li, 2014). A placebo or vehicle arm (PLB) is usually included in these studies in order to demonstrate that the study is sufficiently sensitive to identify the clinical effect in the study population enrolled in the study. Therefore, the typical study design of a clinical endpoint BE study for topical products is a double-blind, randomized 3-arm parallel clinical trial. In order establish bioequivalence between TEST and REF, TEST has to be shown to be superior to PLB, REF shown to be superior to PLB, and TEST shown to be equivalent to REF. Therefore, an overall BE test for a clinical endpoint BE study is composed of two superiority tests (TEST vs. PLB and REF vs. PLB) and one equivalence test (TEST vs. REF).

For generic drugs, the two-sided tests (TOST) procedure (Schuirmann, 1987) is the standard test for average equivalence. Previously, Phillips (1990) calculated power for the TOST in a two-phase two-period PK cross-over study using a bivariate non-central t distribution. The equivalence test in a TOST is based on the different of means (DOM) for PK parameters

after log-transformation, which are assumed to have a log-normal distribution. Chow et al (2003) determined power and sample size for the TOST in a PK cross-over study using the normal Z approximation. Hauschke et al (1999) derived exact methods for calculation of power and sample sizes for the ratio of means (ROM) equivalence test using a bivariate non-central t distribution, assuming the original (untransformed) outcome is normally distributed. Hauschke et al (1999) calculated power and sample sizes for both the parallel study design and the two-period cross-over designs. Chang et al (2014) calculated the sample size and power for an overall BE test based on one superiority test (TEST vs. PLB) and an equivalence test (TEST vs. REF) when the metric of treatment effect is DOM, ROM, or ratio of mean difference. Chang et al used the joint distribution of sample means and sample variances to calculate the expectation of normal distribution conditional upon a Chi-square distribution because "it is not easy to derive the sample size based on the multivariate t-distribution". We call this a Z-ChiSquare method. Shen et al (2015) also used a similar method to calculate power for a two-period two-treatment cross over study.

In this paper, we propose an exact method to calculate the power and sample size for an overall BE test based on two superiority tests (TEST vs. PLB, REF vs. PLB) and one equivalence test (TEST vs. REF) using a multivariate non-central t distribution directly, which we call an Exact-t method. We also applied Chang and Shens Z-ChiSquare method to an overall BE test with two superiority tests and one equality test. In this paper, we focus on two commonly used metrics of treatment effect: difference of means and ratio of means. Simulation was used to compare the proposed Exact-t method and the Z-ChiSquare method in power and sample size calculation against the empirical power, as well as in computation efficiency.

2. Method

The Method section describes in detail how the study was conducted, including conceptual and operational definitions of the variables used in the study, Different types of studies will rely on different methodologies; however, a complete description of the methods used enables the reader to evaluate the appropriateness of your methods and the reliability and the validity of your results, It also permits experienced investigators to replicate the study, If your manuscript is an update of an ongoing or earlier study and the method has been published in detail elsewhere, you may refer the reader to that source and simply give a brief synopsis of the method in this section.

2.1 Exact-t Method

We assume that the continuous outcomes of the three independent treatment groups TEST, REF and PLB X_T, X_R, X_P are normally distributed with means μ_T, μ_R and μ_P and variances σ_T^2, σ_R^2, σ_P^2. The sample size for the three treatment arms are n_T, n_R and n_P, respectively.

2.1.1 Difference of Means as Metric of Treatment Effect

In a three-arm clinical BE study, an overall BE test is composed of two superiority tests (TEST vs PLB, REF vs PLB) each at $\alpha = 0.025$ level and one equivalent test (TEST vs REF) at $\alpha = 0.05$. The hypotheses of superiority tests are given by

$$H_{01} : \mu_R - \mu_P \leqslant 0 \quad \text{vs.} \quad H_{11} : \mu_R - \mu_P > 0$$
$$H_{02} : \mu_T - \mu_P \leqslant 0 \quad \text{vs.} \quad H_{12} : \mu_T - \mu_P > 0$$

Here we assume that a larger value indicates a more desirable effect. The hypothesis of equivalence test is given by

$$H_0 : \theta_1 < \mu_T - \mu_R < \theta_2 \quad \text{vs.} \quad H_1 : \mu_T - \mu_R \leqslant \theta_1 \text{ or } \mu_T - \mu_R \geqslant \theta_2$$

where θ_1 and θ_2 are pre-defined equivalence margins. Since the equivalence test is operationally the same as two one-sided tests (TOST) (Schuirmann, 1987), it can be expressed as follows:

$$H_{03} : \mu_T - \mu_R \leqslant \theta_1 \quad \text{vs.} \quad H_{13} : \mu_T - \mu_R > \theta_1$$
$$H_{04} : \mu_T - \mu_R \geqslant \theta_2 \quad \text{vs.} \quad H_{14} : \mu_T - \mu_R < \theta_2$$

The test statistic for H_{01} is $T_1 = \frac{\overline{X}_T - \overline{X}_P}{Se(\overline{X}_T - \overline{X}_P)} = \frac{\overline{X}_T - \overline{X}_P}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_P^2}{n_P}}}$ Therefore H_{01} is rejected when $T_1 \geqslant t_{0.025, v_1}$. For unequal variances,

the degree of freedom v_1 can be approximated by the Satterthwaite approach (Satterthwaite, 1946) $v_1 = \frac{(\frac{\sigma_T^2}{n_T} + \frac{\sigma_P^2}{n_P})^2}{\frac{\sigma_T^4/n_T^2}{n_T-1} + \frac{\sigma_P^4/n_P^2}{n_P-1}}$.

Likewise, the test statistic for H_{02} is $T_2 = \frac{\overline{X}_R - \overline{X}_P}{Se(\overline{X}_R - \overline{X}_P)} = \frac{\overline{X}_R - \overline{X}_P}{\sqrt{\frac{s_R^2}{n_R} + \frac{s_P^2}{n_P}}}$ and H_{02} is rejected when $T_2 \geqslant t_{0.025, v_2}$ and v_2 can be

estimated by $v_2 = \frac{(\frac{\sigma_R^2}{n_R} + \frac{\sigma_P^2}{n_P})^2}{\frac{\sigma_R^4/n_R^2}{n_R-1} + \frac{\sigma_P^4/n_P^2}{n_P-1}}$. For H_{03}, the test statistic is $T_3 = \frac{\overline{X}_T - \overline{X}_R - \theta_1}{Se(\overline{X}_T - \overline{X}_R)} = \frac{\overline{X}_T - \overline{X}_R - \theta_1}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_R^2}{n_R}}}$ and H_{03} is rejected when $T_3 \geqslant t_{0.05, v_3}$

and v_3 can be estimated by $v_3 = \frac{(\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R})^2}{\frac{\sigma_T^4/n_T^2}{n_T-1} + \frac{\sigma_R^4/n_R^2}{n_R-1}}$. Likewise, for H_{04}, $T_4 = \frac{\overline{X}_T - \overline{X}_R - \theta_2}{Se(\overline{X}_T - \overline{X}_R)} = \frac{\overline{X}_T - \overline{X}_R - \theta_2}{\sqrt{\frac{S_T^2}{n_T} + \frac{S_R^2}{n_R}}}$ and H_{04} is rejected when

$T_4 \leqslant -t_{0.05,v_4}$ and v_4 can be estimated by $v_4 = \frac{(\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R})^2}{\frac{\sigma_T^4/n_T^2}{n_T-1} + \frac{\sigma_R^4/n_R^2}{n_R-1}}$.

Due to the Intercept-Union Test (IUT), the family-wise type 1 error is controlled under 0.05 (Hauck & Anderson, 1984). The power function for the overall BE test can be calculated as

$$1 - \beta$$
$$= P(T_1 \geqslant t_{0.025,v_1}, T_2 \geqslant t_{0.025,v_2}, T_3 \geqslant t_{0.05,v_3}, T_4 \leqslant -t_{0.05,v_4} \mid H_{11}, H_{12}, H_{13}, H_{14})$$

Without loss of generality, we assume equal variance among the three treatment groups $\sigma_T^2 = \sigma_R^2 = \sigma_P^2 = \sigma^2$ and the pooled variance is defined as

$$S_{Pool}^2 = \frac{\sum_{i=1}^{n_T}(x_i - \overline{X}_T)^2 + \sum_{j=1}^{n_R}(x_j - \overline{X}_R)^2 + \sum_{k=1}^{n_P}(x_k - \overline{X}_P)^2}{n_T + n_R + n_P - 3}$$

We have equal degrees of freedom $v = v_1 = v_2 = v_3 = v_4 = n_T + n_R + n_P - 3$ and the test statistics

$$T_1 = \frac{\overline{X}_T - \overline{X}_P}{S_{Pool}\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}} = \frac{N_1}{D_1}, \quad T_2 = \frac{\overline{X}_R - \overline{X}_P}{S_{Pool}\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} = \frac{N_2}{D_2}$$

$$T_3 = \frac{\overline{X}_T - \overline{X}_R - \theta_1}{S_{Pool}\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}} = \frac{N_3}{D_3}, \quad T_4 = \frac{\overline{X}_T - \overline{X}_R - \theta_2}{S_{Pool}\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}} = \frac{N_4}{D_4}$$

Under the alternative hypotheses, T_1, T_2, T_3 and T_4 follow a multivariate non-central t distribution with non-centrality parameters

$$\Theta_1 = \frac{\mu_T - \mu_P}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}}, \quad \Theta_2 = \frac{\mu_R - \mu_P}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}}, \quad \Theta_3 = \frac{\mu_T - \mu_R - \theta_1}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}}, \quad \Theta_4 = \frac{\mu_T - \mu_R - \theta_2}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}}.$$

and the correlation matrix

$$\rho = \begin{pmatrix} 1 & \frac{\frac{1}{n_P}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}} & \frac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & \frac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} \\ \frac{\frac{1}{n_P}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}} & 1 & \frac{-\frac{1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & \frac{-\frac{1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} \\ \frac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & \frac{-\frac{1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & 1 & 1 \\ \frac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & \frac{-\frac{1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} & 1 & 1 \end{pmatrix}$$

In a balanced test, the three arms have equal sample sizes $n_T = n_R = n_P = n$, the test statistics are

$$T_1 = \frac{\overline{X}_T - \overline{X}_P}{S_{Pool}\sqrt{\frac{2}{n}}}, \quad T_2 = \frac{\overline{X}_R - \overline{X}_P}{S_{Pool}\sqrt{\frac{2}{n}}}, \quad T_3 = \frac{\overline{X}_T - \overline{X}_R - \theta_1}{S_{Pool}\sqrt{\frac{2}{n}}}, \quad T_4 = \frac{\overline{X}_T - \overline{X}_R - \theta_2}{S_{Pool}\sqrt{\frac{2}{n}}}$$

with non-centrality parameters

$$\Theta_1 = \frac{\mu_T - \mu_P}{\sigma\sqrt{\frac{2}{n}}}, \quad \Theta_2 = \frac{\mu_R - \mu_P}{\sigma\sqrt{\frac{2}{n}}}, \quad \Theta_3 = \frac{\mu_T - \mu_R - \theta_1}{\sigma\sqrt{\frac{2}{n}}}, \quad \Theta_4 = \frac{\mu_T - \mu_R - \theta_2}{\sigma\sqrt{\frac{2}{n}}}.$$

and the correlation matrix can thus be simplified as

$$\rho = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & 1 & 1 \\ \frac{1}{4} & -\frac{1}{4} & 1 & 1 \end{pmatrix}$$

With the density function of a multivariate non-central t distribution, we can calculate the power for the overall BE test as

$$
\begin{aligned}
1 - \beta \\
= \int_{t_{0.025,\nu}}^{+\infty} \int_{t_{0.025,\nu}}^{+\infty} \int_{t_{0.05,\nu}}^{+\infty} \int_{-\infty}^{-t_{0.05,\nu}} \frac{\Gamma(\frac{\nu+4}{2})}{(\pi\nu)^2 |\sigma^2 \rho|^{\frac{1}{2}} \Gamma(\frac{\nu}{2})} \times \\
\left(1 + \frac{(x-\mu)^T \rho^{-1}(x-\mu)}{\sigma^2 \nu} \right)^{-\frac{\nu+4}{2}} dT_1 dT_2 dT_3 dT_4
\end{aligned}
\tag{1}
$$

where x is the vector of (T_1, T_2, T_3, T_4), μ is the mean of x. We use R and package "mvtnorm" and function "pmvt" to calculate the power for the Exact-t method.

2.1.2 Ratio of Means as Metric of Treatment Effect

Ratio of means is more often used to test equivalence in a BE study. Likewise, the overall BE test contains two superiority tests each at a significance level of 0.025 and an equivalence test at $\alpha = 0.05$. The superiority tests are the same as before for the difference of means. The equivalence test is instead given by

$$
H_{03} : \quad \frac{\mu_T}{\mu_R} \leqslant \theta_1 \quad \text{vs.} \quad H_{13} : \quad \frac{\mu_T}{\mu_R} > \theta_1
$$

$$
H_{04} : \quad \frac{\mu_T}{\mu_R} \geqslant \theta_2 \quad \text{vs.} \quad H_{14} : \quad \frac{\mu_T}{\mu_R} < \theta_2
$$

For H_{03}, the test statistic is

$$
T_3 = \frac{\overline{X}_T - \theta_1 \overline{X}_R}{S_e(\overline{X}_T - \theta_1 \overline{X}_R)} = \frac{\overline{X}_T - \theta_1 \overline{X}_R}{\sqrt{\frac{S_T^2}{n_T} + \frac{\theta_1^2 S_R^2}{n_R}}}
$$

and H_{03} is rejected when $T_3 \geqslant t_{0.05,\nu_3}$ where ν_3 is estimated by the Satterthwaite approach (Satterthwaite, 1946), where $\nu_3 = \frac{(\frac{\sigma_T^2}{n_T} + \frac{\theta_1^2 \sigma_R^2}{n_R})^2}{\frac{\sigma_T^4/n_T^2}{n_T-1} + \frac{\theta_1^4 \sigma_R^4/n_R^2}{n_R-1}}$. Likewise, for H_{04},

$$
T_4 = \frac{\overline{X}_T - \theta_2 \overline{X}_R}{S_e(\overline{X}_T - \theta_2 \overline{X}_R)} = \frac{\overline{X}_T - \theta_2 \overline{X}_R}{\sqrt{\frac{S_T^2}{n_T} + \frac{\theta_2^2 S_R^2}{n_R}}}
$$

and H_{04} is rejected when $T_4 \leqslant -t_{0.05,\nu_4}$ where ν_4 is estimated by $\nu_4 = \frac{(\frac{\sigma_T^2}{n_T} + \frac{\theta_2^2 \sigma_R^2}{n_R})^2}{\frac{\sigma_T^4/n_T^2}{n_T-1} + \frac{\theta_2^4 \sigma_R^4/n_R^2}{n_R-1}}$. The power function for the overall BE test is thus

$$
\begin{aligned}
1 - \beta \\
= P(T_1 \geqslant t_{0.025,\nu_1}, T_2 \geqslant t_{0.025,\nu_2}, T_3 \geqslant t_{0.05,\nu_3}, T_4 \leqslant -t_{0.05,\nu_4} \mid H_{11}, H_{12}, H_{13}, H_{14})
\end{aligned}
$$

Assume equal variances $\sigma_T^2 = \sigma_R^2 = \sigma_P^2 = \sigma^2$, we have equal degrees of freedom $\nu = n_T + n_R + n_P - 3$ and the test statistics

$$
T_1 = \frac{\overline{X}_T - \overline{X}_P}{S_{Pool}\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}} = \frac{N_1}{D_1}, \quad T_2 = \frac{\overline{X}_R - \overline{X}_P}{S_{Pool}\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} = \frac{N_2}{D_2}
$$

$$
T_3 = \frac{\overline{X}_T - \theta_1 \overline{X}_R}{S_{Pool}\sqrt{\frac{1}{n_T} + \theta_1^2 \frac{1}{n_R}}} = \frac{N_3}{D_3}, \quad T_4 = \frac{\overline{X}_T - \theta_2 \overline{X}_R}{S_{Pool}\sqrt{\frac{1}{n_T} + \theta_2^2 \frac{1}{n_R}}} = \frac{N_4}{D_4}
$$

Under the alternative hypotheses, T_1, T_2, T_3 and T_4 follows multivariate non-central t distribution with non-centrality parameters

$$
\Theta_1 = \frac{\mu_T - \mu_P}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}}, \quad \Theta_2 = \frac{\mu_R - \mu_P}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}}, \quad \Theta_3 = \frac{\mu_T - \theta_1 \mu_R}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_1^2}{n_R}}}, \quad \Theta_4 = \frac{\mu_T - \theta_2 \mu_R}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_2^2}{n_R}}}.
$$

and the correlation matrix

$$
\rho = \begin{pmatrix}
1 & \dfrac{\frac{1}{n_P}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}} & \dfrac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}} & \dfrac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} \\[2.5em]
\dfrac{\frac{1}{n_P}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}} & 1 & \dfrac{-\frac{\theta_1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}} & \dfrac{-\frac{\theta_2}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} \\[2.5em]
\dfrac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}} & \dfrac{-\frac{\theta_1}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}} & 1 & \dfrac{\frac{1}{n_T}+\frac{\theta_1\theta_2}{n_R}}{\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} \\[2.5em]
\dfrac{\frac{1}{n_T}}{\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} & \dfrac{-\frac{\theta_2}{n_R}}{\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} & \dfrac{\frac{1}{n_T}+\frac{\theta_1\theta_2}{n_R}}{\sqrt{\frac{1}{n_T}+\frac{\theta_1^2}{n_R}}\sqrt{\frac{1}{n_T}+\frac{\theta_2^2}{n_R}}} & 1
\end{pmatrix}
$$

In a balanced test, the three arms have equal sample sizes $n_T = n_R = n_P = n$, the test statistics are

$$
T_1 = \frac{\overline{X}_T - \overline{X}_P}{S_{Pool}\sqrt{\frac{2}{n}}}, \quad
T_2 = \frac{\overline{X}_R - \overline{X}_P}{S_{Pool}\sqrt{\frac{2}{n}}}, \quad
T_3 = \frac{\overline{X}_T - \theta_1\overline{X}_R}{S_{Pool}\sqrt{\frac{1+\theta_1^2}{n}}}, \quad
T_4 = \frac{\overline{X}_T - \theta_2\overline{X}_R}{S_{Pool}\sqrt{\frac{1+\theta_2^2}{n}}}
$$

with non-centrality parameters

$$
\Theta_1 = \frac{\mu_T - \mu_P}{\sigma\sqrt{\frac{2}{n}}}, \quad
\Theta_2 = \frac{\mu_R - \mu_P}{\sigma\sqrt{\frac{2}{n}}}, \quad
\Theta_3 = \frac{\mu_T - \theta_1\mu_R}{\sigma\sqrt{\frac{1+\theta_1^2}{n}}}, \quad
\Theta_4 = \frac{\mu_T - \theta_2\mu_R}{\sigma\sqrt{\frac{1+\theta_2^2}{n}}}.
$$

and the correlation matrix can be simplified as

$$
\rho = \begin{pmatrix}
1 & \frac{1}{2} & \frac{1}{\sqrt{2(1+\theta_1^2)}} & \frac{1}{\sqrt{2(1+\theta_2^2)}} \\[1em]
\frac{1}{2} & 1 & \frac{-\theta_1}{\sqrt{2(1+\theta_1^2)}} & \frac{-\theta_2}{\sqrt{2(1+\theta_2^2)}} \\[1em]
\frac{1}{\sqrt{2(1+\theta_1^2)}} & \frac{-\theta_1}{\sqrt{2(1+\theta_1^2)}} & 1 & \frac{1+\theta_1\theta_2}{\sqrt{(1+\theta_1^2)(1+\theta_2^2)}} \\[1em]
\frac{1}{\sqrt{2(1+\theta_2^2)}} & \frac{-\theta_2}{\sqrt{2(1+\theta_2^2)}} & \frac{1+\theta_1\theta_2}{\sqrt{(1+\theta_1^2)(1+\theta_2^2)}} & 1
\end{pmatrix}
$$

The power function for the overall BE test is the same as before in (1), which can be estimated by the same R package as for the difference of means.

2.2 Extension of Z-ChiSquare Method

Chang et al (2014) and Shen et al (2015) both used Z-ChisSquare method to calculate the power of BE test.

2.2.1 Difference of Means as Metric of Treatment Effect

Assuming homogeneity across treatment groups, power for an overall BE test including two superiority tests and one equivalence test can be written as

$$
\begin{aligned}
1 - \beta \\
= P(T_1 &\geqslant t_{0.025,\nu_1}, T_2 \geqslant t_{0.025,\nu_2}, T_3 \geqslant t_{0.05,\nu_3}, T_4 \leqslant -t_{0.05,\nu_4}) \\
= P\Bigg(&\frac{\overline{X}_T - \overline{X}_P}{S_{pool}\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}} \geqslant t_{0.025,\nu_1}, \ \frac{\overline{X}_R - \overline{X}_P}{S_{pool}\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}} \geqslant t_{0.025,\nu_2}, \\
&\frac{\overline{X}_T - \overline{X}_R - \theta_1}{S_{pool}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} \geqslant t_{0.05,\nu_3}, \ \frac{\overline{X}_T - \overline{X}_R - \theta_2}{S_{pool}\sqrt{\frac{1}{n_T}+\frac{1}{n_R}}} \leqslant -t_{0.05,\nu_4} \Bigg)
\end{aligned}
$$

Using Chang et al (2014) and Shen et al (2015)'s approach, let $U^2 = \frac{\nu S_{pool}^2}{\sigma^2} \sim \chi_\nu^2$, $Z_1 = \frac{\overline{X}_T - \overline{X}_P - \mu_{TP}}{\sigma\sqrt{\frac{1}{n_T}+\frac{1}{n_P}}}$, $Z_2 = \frac{\overline{X}_R - \overline{X}_P - \mu_{RP}}{\sigma\sqrt{\frac{1}{n_R}+\frac{1}{n_P}}}$,

$Z_3 = \frac{\overline{X}_T - \overline{X}_R - \mu_{TR}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}}$ where $\mu_{TP} = \mu_T - \mu_P$, $\mu_{RP} = \mu_R - \mu_P$, $\mu_{TR} = \mu_T - \mu_R$, the power can be written as

$$1 - \beta$$
$$= P\left(Z_1 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{TP}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}}, Z_2 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{RP}}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}},\right.$$
$$\left. t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_1 - \mu_{TR}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}} \leqslant Z_3 \leqslant -t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_2 - \mu_{TR}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}}\right)$$

Note that U^2 need to be smaller than or equal to $\frac{(\theta_2 - \theta_1)^2 \nu}{4\sigma^2 t_{0.05,\nu}^2(\frac{1}{n_T} + \frac{1}{n_R})}$ in order to hold the validity of the probability such that the upper bound of Z_3 is larger than its lower bound. Therefore, the power function can be written in the form of expectation of multivariate normal distribution conditional upon a chi-square distribution:

$$1 - \beta$$
$$= E_{U^2}\left[P\left(Z_1 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{TP}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}}, Z_2 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{RP}}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}},\right.\right.$$
$$t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_1 - \mu_{TR}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}} \leqslant Z_3 \leqslant -t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_2 - \mu_{TR}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_R}}}$$
$$\left.\left.\left| U^2 \leqslant \frac{(\theta_2 - \theta_1)^2 \nu}{4\sigma^2 t_{0.05,\nu}^2(\frac{1}{n_T} + \frac{1}{n_R})}\right)\right]\right.$$

Under the assumption of equal sample sizes $n_T = n_R = n_P = n$, we have

$$1 - \beta$$
$$= E_{U^2}\left[P\left(Z_1 \geqslant t_{0.025,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{TP}}{\sigma\sqrt{\frac{2}{n}}}, Z_2 \geqslant t_{0.025,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{RP}}{\sigma\sqrt{\frac{2}{n}}},\right.\right.$$
$$t_{0.05,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_1 - \mu_{TR}}{\sigma\sqrt{\frac{2}{n}}} \leqslant Z_3 \leqslant -t_{0.05,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} + \frac{\theta_2 - \mu_{TR}}{\sigma\sqrt{\frac{2}{n}}}$$
$$\left.\left.\left| U^2 \leqslant \frac{(\theta_2 - \theta_1)^2 n\nu}{8\sigma^2 t_{0.05,3n-3}^2}\right)\right]\right.$$

where

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}\right)$$

and $U^2 = \frac{\nu S_p^2}{\sigma^2} \sim \chi_{3n-3}^2$.

We use R and package "mvtnorm" to calculate the power. The numerical integral is calculated in the following steps:

1. Use function "rchisq" to generate a sample of U^2 based on the truncated chi-square distribution.

2. For each U^2, calculate the inside probability with multivariate normal cumulative density function "pmvnorm".

3. Take the mean of the probabilities to calculate the expectation with respect to U^2 as the Z-ChiSquare power.

2.2.2 Ratio of Means as Metric of Treatment Effect

Similar to the difference of means, the power function for the overall BE test based on the ratio of means can be written as

$$
1 - \beta
$$
$$
= \quad P(T_1 \geqslant t_{0.025,\nu_1}, T_2 \geqslant t_{0.025,\nu_2},
$$
$$
T_3 \geqslant t_{0.05,\nu_3}, T_4 \leqslant -t_{0.05,\nu_4})
$$
$$
= P\left(\frac{\overline{X}_T - \overline{X}_P}{S_{pool}\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}} \geqslant t_{0.025,\nu_1} , \frac{\overline{X}_R - \overline{X}_P}{S_{pool}\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} \geqslant t_{0.025,\nu_2} , \right.
$$
$$
\left. \frac{\overline{X}_T - \theta_1\overline{X}_R}{S_{pool}\sqrt{\frac{1}{n_T} + \frac{\theta_1^2}{n_R}}} \geqslant t_{0.05,\nu_3} , \frac{\overline{X}_T - \theta_2\overline{X}_R}{S_{pool}\sqrt{\frac{1}{n_T} + \frac{\theta_2^2}{n_R}}} \leqslant -t_{0.05,\nu_4} \right)
$$

Likewise, this can be written in the form of Z_1, Z_2, Z_3, Z_4:

$$
1 - \beta
$$
$$
= \quad P\left(Z_1 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{TP}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}} , Z_2 \geqslant t_{0.025,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{RP}}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} , \right.
$$
$$
\left. Z_3 \geqslant t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{T\theta_1 R}}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_1^2}{n_R}}} , Z_4 \leqslant -t_{0.05,\nu} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{T\theta_2 R}}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_2^2}{n_R}}} \right)
$$

where $U^2 = \frac{\nu S_{pool}^2}{\sigma^2} \sim \chi_\nu^2$, $\mu_{T\theta_1 R} = \mu_T - \theta_1\mu_R$, $\mu_{T\theta_2 R} = \mu_T - \theta_2\mu_R$,

$$
Z_1 = \frac{\overline{X}_T - \overline{X}_P - \mu_{TP}}{\sigma\sqrt{\frac{1}{n_T} + \frac{1}{n_P}}} , Z_2 = \frac{\overline{X}_R - \overline{X}_P - \mu_{RP}}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} ,
$$
$$
Z_3 = \frac{\overline{X}_T - \theta_1\overline{X}_R - \mu_{T\theta_1 R}}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_1^2}{n_R}}} , Z_4 = \frac{\overline{X}_T - \theta_2\overline{X}_R - \mu_{T\theta_2 R}}{\sigma\sqrt{\frac{1}{n_T} + \frac{\theta_2^2}{n_R}}}
$$

Under balanced group assumption, $n_T = n_R = n_P = n$, the power function can be simplified as

$$
1 - \beta
$$
$$
= \quad P\left(Z_1 \geqslant t_{0.025,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{TP}}{\sigma\sqrt{\frac{2}{n}}} , Z_2 \geqslant t_{0.025,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{RP}}{\sigma\sqrt{\frac{2}{n}}} , \right.
$$
$$
\left. Z_3 \geqslant t_{0.05,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{T\theta_1 R}}{\sigma\sqrt{\frac{1+\theta_1^2}{n}}} , Z_4 \leqslant -t_{0.05,3n-3} \cdot \sqrt{\frac{U^2}{\nu}} - \frac{\mu_{T\theta_2 R}}{\sigma\sqrt{\frac{1+\theta_2^2}{n}}} \right)
$$

where

$$
\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{\sqrt{2(1+\theta_1^2)}} & \frac{1}{\sqrt{2(1+\theta_2^2)}} \\ \frac{1}{2} & 1 & -\frac{\theta_1}{\sqrt{2(1+\theta_1^2)}} & -\frac{\theta_2}{\sqrt{2(1+\theta_2^2)}} \\ \frac{1}{\sqrt{2(1+\theta_1^2)}} & -\frac{\theta_1}{\sqrt{2(1+\theta_1^2)}} & 1 & \frac{1+\theta_1\theta_2}{\sqrt{(1+\theta_1^2)(1+\theta_2^2)}} \\ \frac{1}{\sqrt{2(1+\theta_2^2)}} & -\frac{\theta_2}{\sqrt{2(1+\theta_2^2)}} & \frac{1+\theta_1\theta_2}{\sqrt{(1+\theta_1^2)(1+\theta_2^2)}} & 1 \end{pmatrix} \right)
$$

and $U^2 = \frac{\nu S_p^2}{\sigma^2} \sim \chi_{3n-3}^2$

3. Simulation

In order to examine the performance of Exact-t and Z-ChiSquare method, we compare the power of the two methods against the empirical power by simulation under different scenarios.

3.1 Data Generation

We assume normal distribution and equal variance across the groups ($\sigma_T^2 = \sigma_R^2 = \sigma_P^2 = \sigma^2$).

1. Generate response for each treatment group as follows:

$$X_{Ti} \sim N(\mu_T, \sigma^2), \quad i = 1, \ldots, n_T ,$$

$$X_{Rj} \sim N(\mu_R, \sigma^2), \quad j = 1, \ldots, n_R ,$$

$$X_{Pk} \sim N(\mu_P, \sigma^2), \quad k = 1, \ldots, n_P ,$$

where n_T, n_R and n_P are the sample sizes for each group. In this simulation, we assume $0.35 < \mu_T < 0.65$, $\mu_R = 0.5$, $\mu_P = 0.3$, $\sigma \in (0.1, 0.2)$. We assume an equal sample size across treatment groups: $n = 12, 24, 36, 48, 60, 100, 200, 300$ per arm.

2. Calculate bioequivalence test statistics T_1, T_2, T_3 and T_4.

3. Perform tests of superiority, equivalence, and overall BE test.

4. Iterate previous steps N times ($N = 10000$). Empirical power is calculated based on the proportion of the times when the null hypothesis is rejected.

5. Calculate power for the Exact-t method and Z-ChiSquare method as outlined in the Methods section.

4. Results

4.1 Difference of Means Test

We first compare the powers for difference of mean tests under the condition that test group, reference group, and placebo group all follow a normal distribution where $\mu_r = 0.5$, $\mu_p = 0.3$ with equal standard deviation $\sigma = 0.2$ between the groups.

4.1.1 Power of Exact-t Method vs. Effect Size by Different Variances and Sample Sizes

Figure 1 shows the empirical power by effect size for (A) two superiority tests (TEST vs. PLB, RLD vs PLB); (B) one equivalence test (TEST vs. REF); and (C) the overall BE test including two superiority tests and one equivalence test, when $\mu_R = 0.5$, $\mu_P = 0.3$, and $\sigma = 0.2$.
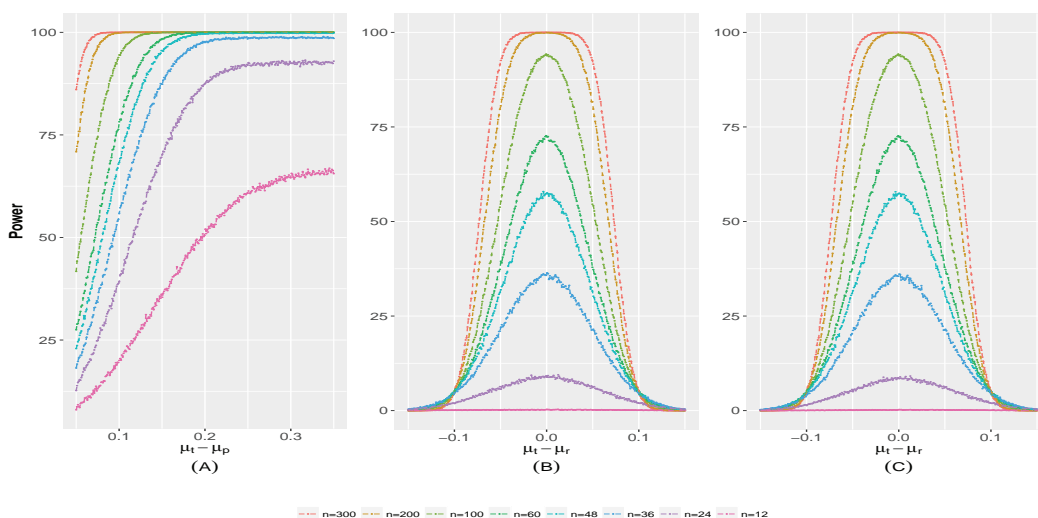

Figure 1. Empirical power when $\mu_R = 0.5$, $\mu_P = 0.3$, $\sigma = 0.2$ (A) Two arm superiority tests: $\alpha = 0.025$ (B) Equivalence test (TEST vs. REF): $\alpha = 0.05$ (C) Bioequivalence tests: 2 superiority tests (TEST vs. PLB, REF vs. PLB) + 1 equivalence test (TEST vs. REF), $\alpha = 0.05$
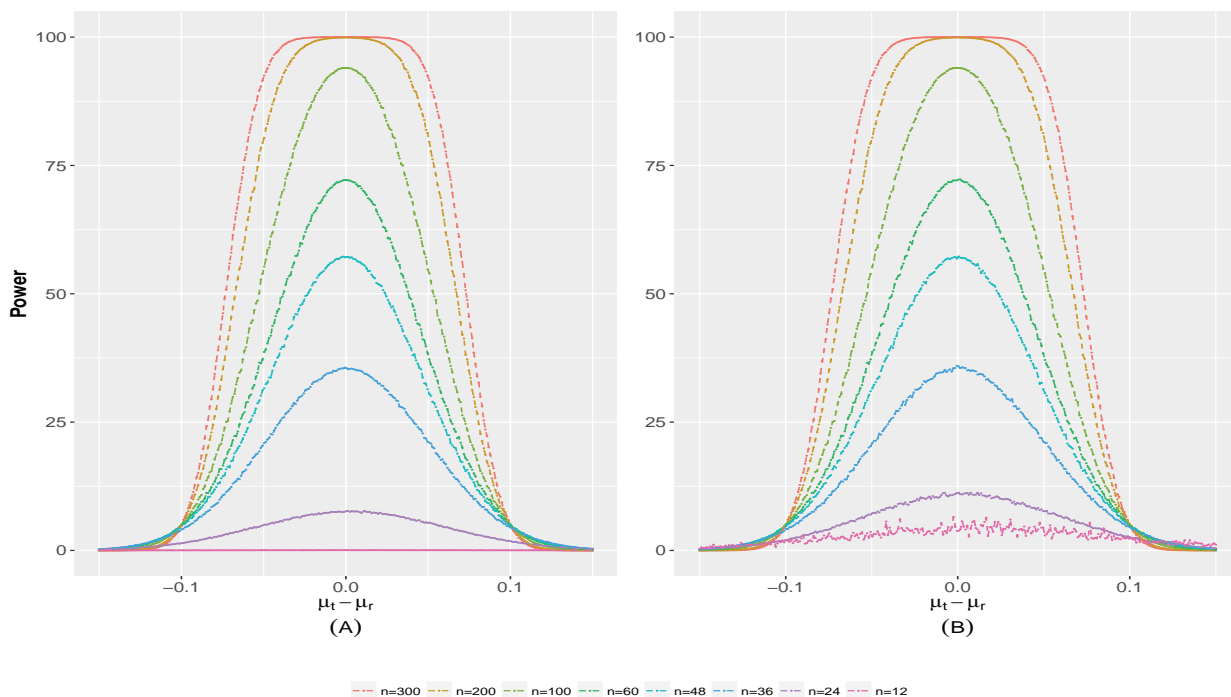
Figure 2. Compare Exact-t and Z-ChiSquare Method: Power for Overall BE Test: 2 superiority tests (TEST vs. PLB, REF vs. PLB) + 1 equivalence test (TEST vs. REF), $\mu_R = 0.5$, $\mu_P = 0.3$, $\sigma = 0.2$, $\alpha = 0.05$ (A) Exact non-central t distribution. (B) Z-ChiSq

We can see that the power of the superiority test increases when the difference between test group and placebo groups increases. The empirical power for equivalence and bioequivalence is bell shaped and reaches maximum when means of test group and reference group are equal. Figure 2 compares the power calculated using (A) the Exact-t method and (B) the Z-Chi-Square method for the overall BE test, by different effect sizes and different sample sizes. Both Exact-t and Z-ChiSquare are bell shaped and are generally good estimates of the empirical powers. However, when sample size is small, the Z-ChiSqure method has unstable estimates of power, whereas the Exact-t method has a smooth and accurate estimation of the empirical power.

4.1.2 Power vs. Sample Size by Different Methods

To better understand the difference between Exact-t and Z-ChiSquare power function, and to understand which tests (superiority test or equivalence test) drives the power of the overall BE test, we study power vs sample size for two superiority tests, one equivalence test, and the overall BE test based on the empirical power, as well as the power of the overall BE test calculated from the Exact-t method and the Z-ChiSquare method. Four different scenarios are considered in Figure 3 and Figure 4:

1. $\mu_R = 0.5, \mu_T = 0.475, \mu_P = 0.3, \sigma = 0.2$

2. $\mu_R = 0.5, \mu_T = 0.475, \mu_P = 0.4, \sigma = 0.2$

3. $\mu_R = 0.5, \mu_T = 0.475, \mu_P = 0.4, \sigma = 0.3$

4. $\mu_R = 0.5, \mu_T = 0.55, \mu_P = 0.3, \sigma = 0.2$

When $\mu_P = 0.3$ (3A and 3D), i.e., the effect size is large, the power for superiority is high and the overall BE test is mostly driven by the equivalence test. On the other hand, when $\mu_P = 0.4$ (3B and 3C), i.e., the placebo effect is high, power for the superiority tests becomes the driven force for the overall BE test instead. Between 3B and 3C, as variance increases, power decreases and the power curve shift to the right in 3(C).

A zoom-in plot (Figure 4) further compares the Exact-t method and Z-ChiSquare method against the empirical power in the small sample size range. Figure 4 shows that the power curve of the Z-ChiSquare method (green) deviates from the empirical BE power curve (purple) when sample size is small, whereas the power curve of the Exact-t method (blue) overlaps perfectly with that of the empirical power.
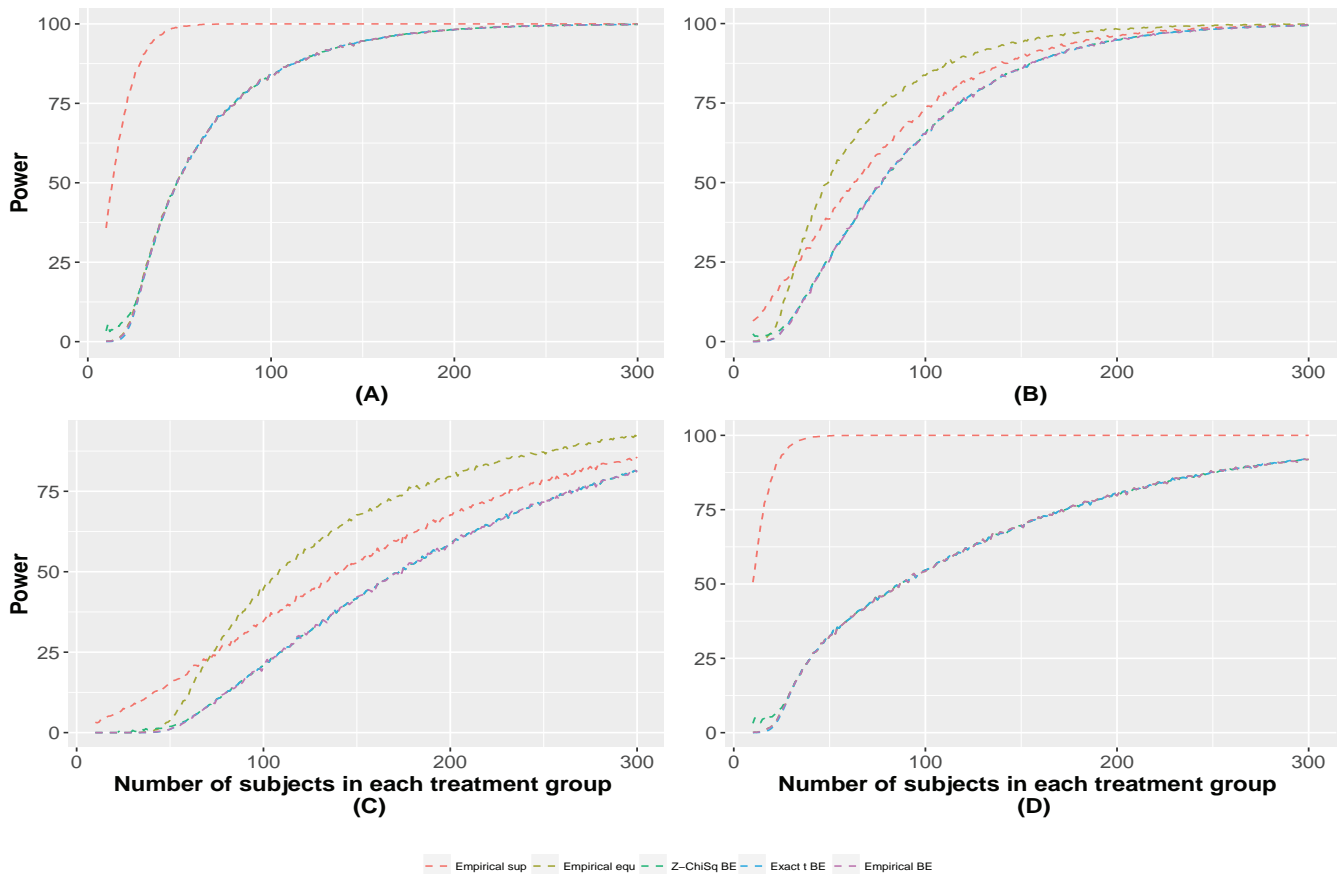

Figure 3. Comparison of empirical power from 2 superiority test: 1 equivalence test, BE tests (2 Sup + 1 Equ), and powers from exact t and Z-ChiSq by sample size (A) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.3$, $\sigma = 0.2$ (B) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.2$ (C) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.3$ (D) $\mu_R = 0.5$, $\mu_T = 0.55$, $\mu_P = 0.3$, $\sigma = 0.2$
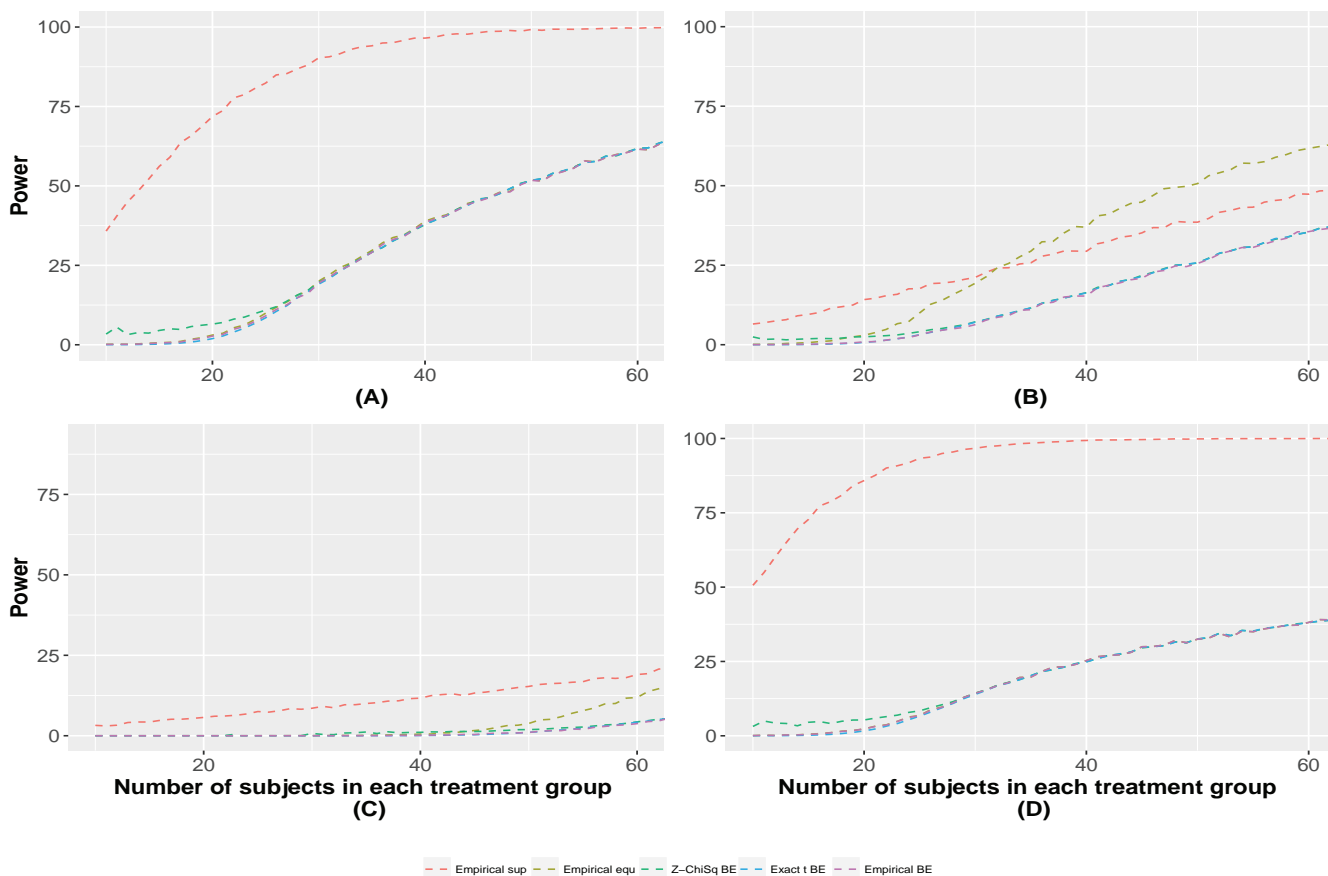
Figure 4. Zoomed-in Comparison of Exact=t Method and Z-ChiSquare Method against the Empirical Power for the Overall BE Test in the Small Sample Size Range (A) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.3$, $\sigma = 0.2$ (B) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.2$ (C) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.3$ (D) $\mu_R = 0.5$, $\mu_T = 0.55$, $\mu_P = 0.3$, $\sigma = 0.2$

4.2 Ratio of Means Test

Similar to the difference of means test, we plotted the empirical power for ratio of means by effect sizer in Figure 5 for (A) two superiority tests (TEST vs. PLB, RLD vs PLB); (B) one equivalence test (TEST vs. REF); and (C) the overall BE test including two superiority tests and one equivalence test, when $\mu_R = 0.5$, $\mu_P = 0.3$, and $\sigma = 0.2$. The results are similar to those of difference of means (Figure 1 and 2), except that the Z-ChiSquare Method (Figure 6B) matches the Exact-t Method (Figure 6A) and the empirical power (Figure 5C) very well even when sample size is small.

Likewise, the empirical power vs sample size for superiority equivalent test, and overall BE test, as well as the power calculated from the Exact-t test and Z-ChiSquare test based on ratio of means in Figure 7 have a very similar trend as those in Figure 4 for difference of means.
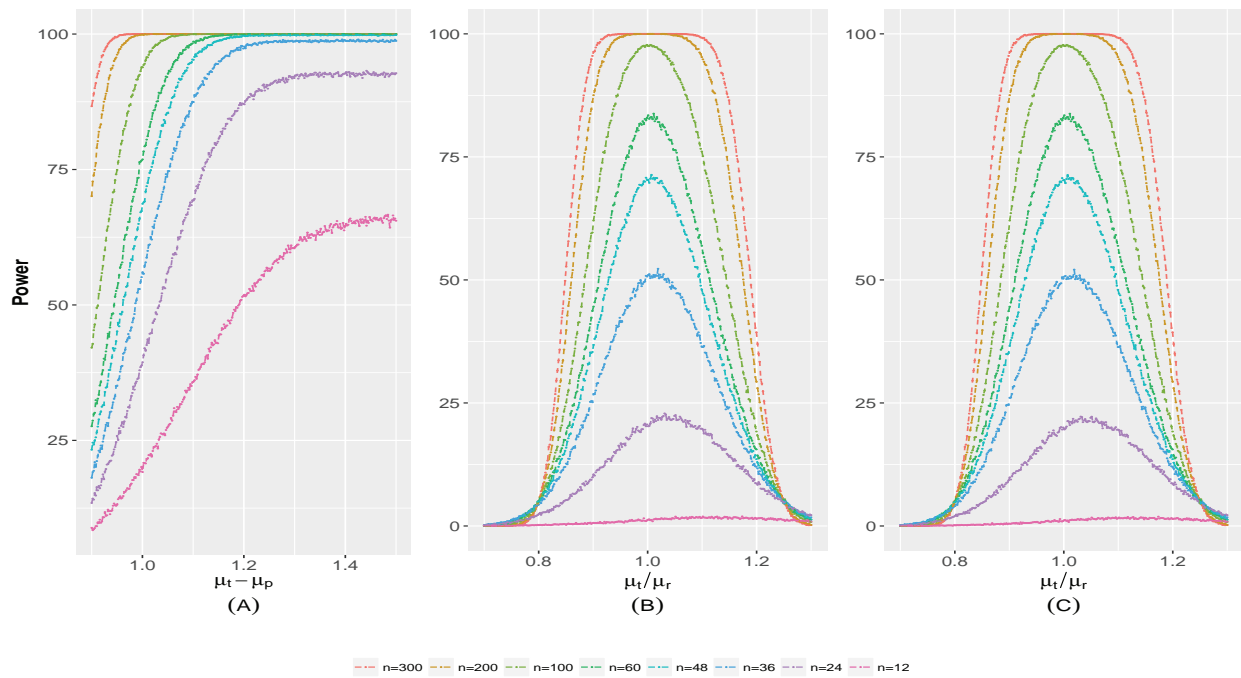
Figure 5. Empirical power when $\mu_R = 0.5$, $\mu_P = 0.3$, $\sigma = 0.2$ (A) Two arm superiority tests: $alpha = 0.025$ (B) Equivalence test (TEST vs. REF): $\alpha = 0.05$ (C) Bioequivalence tests: 2 superiority tests (TEST vs. PLB, REF vs. PLB) + 1 equivalence test (TEST vs. REF), $\alpha = 0.05$
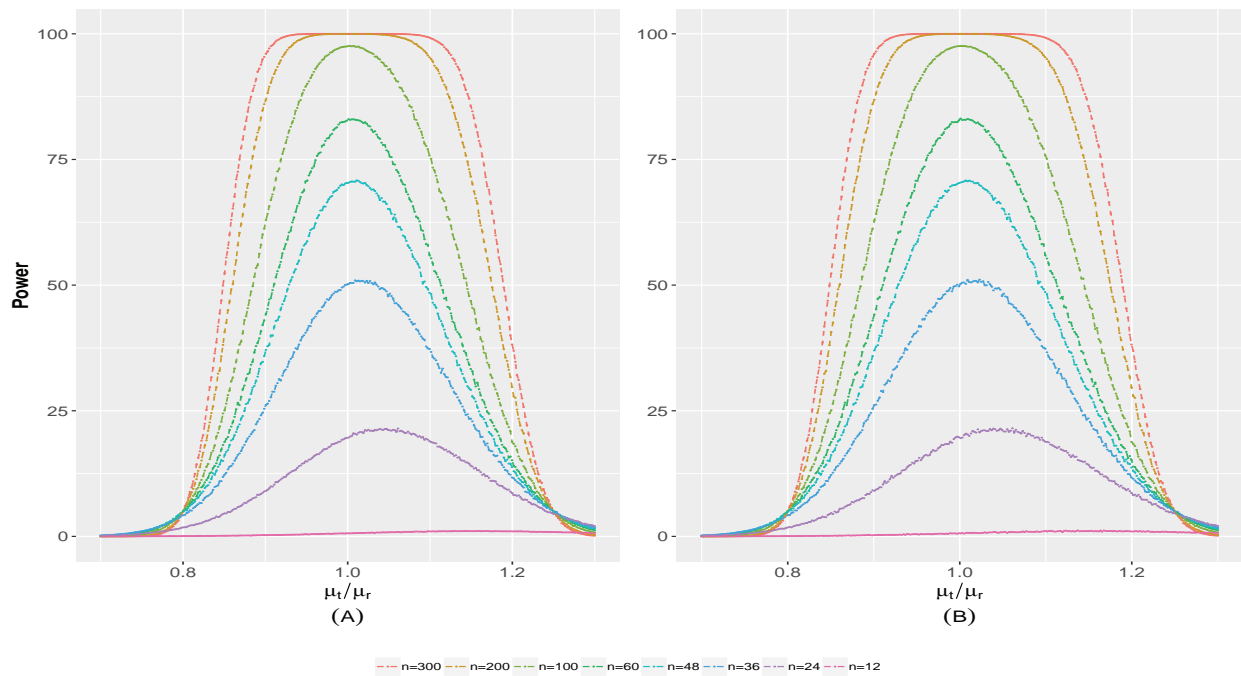

Figure 6. Compare Exact-t and Z-ChiSquare Method: Power for Overall BE Test: 2 superiority tests (TEST vs. PLB, REF vs. PLB) + 1 equivalence test (TEST vs. REF), $\mu_R = 0.5$, $\mu_P = 0.3$, $\sigma = 0.2$, $\alpha = 0.05$ (A) Exact non-central t distribution. (B) Z-ChiSq
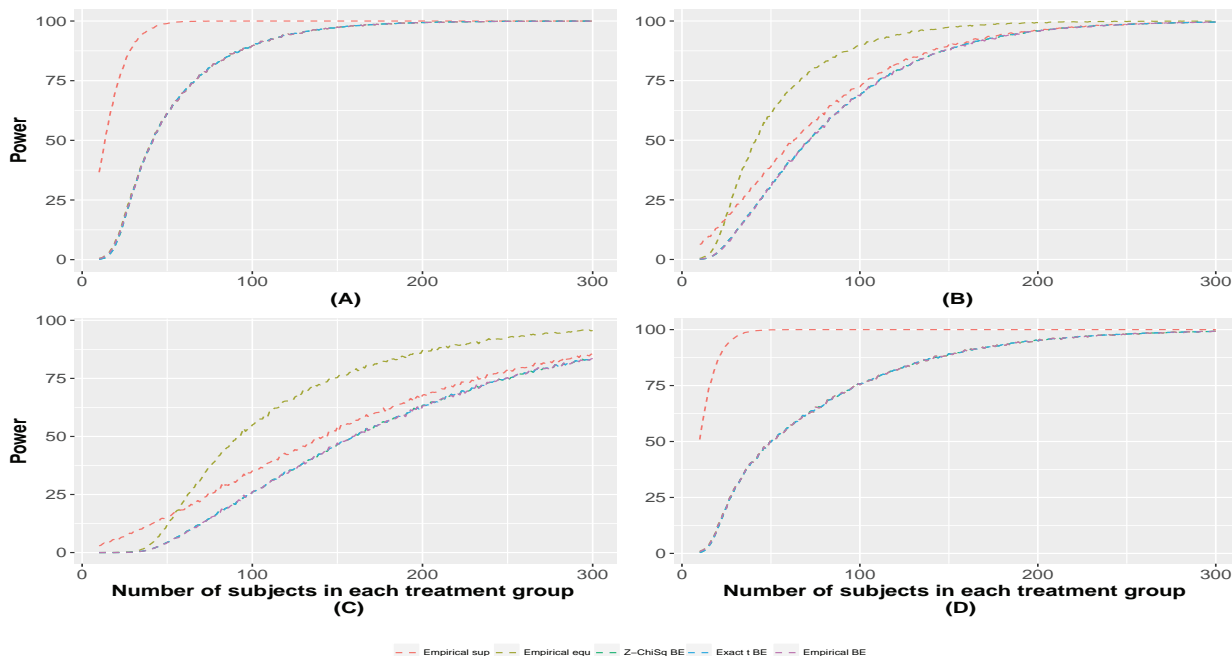
Figure 7. Comparison of empirical power from 2 superiority test: 1 equivalence test, BE tests (2 Sup + 1 Equ), and powers from exact t and Z-ChiSq by sample size (A) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.3$, $\sigma = 0.2$ (B) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.2$ (C) $\mu_R = 0.5$, $\mu_T = 0.475$, $\mu_P = 0.4$, $\sigma = 0.3$ (D) $\mu_R = 0.5$, $\mu_T = 0.55$, $\mu_P = 0.3$, $\sigma = 0.2$

5. Computation Time

Table 1. Computation time comparison for difference of means test (seconds)

	Difference of Means	
Group size	Exact-t	Z-ChiSquare
12	4.90	5.57
24	5.03	20.72
36	5.38	27.46
48	5.73	27.49
60	6.16	28.12
100	6.62	28.27
200	8.70	30.12
300	9.62	30.94

6. Discussion

Previously, Chang et al (2014) calculated the sample size and power for an overall BE test including one superiority test (TEST vs. PLB) and one equivalence test (TEST vs. REF) for the clinically end-point BE study. Chang et al used the joint distribution of sample means and sample variances to calculate the expectation of a normal distribution conditional upon a Chi-square distribution because "it is not easy to derive the sample size based on the multivariate-t distribution".

In this paper, we propose an exact method to calculate the power and sample size for an overall BE test based on two (rather than one) superiority tests (TEST vs. PLB, REF vs. PLB) and one equivalence test (TEST vs. REF) using a multivariate non-central t distribution directly. The proposed Exact-t method is computationally more efficient, especially when sample size gets larger. It can take advantage of an existing R package without self writing codes to numerically calculating the conditional expectation of a multivariate normal distribution based on a truncated Chi-Square distribution. This is more convenient to general users. Although the power of Exact-t and Z-ChiSquare method for the overall BE test overlap when sample size is large, the Z-ChiSquare method deviates from that of the empirical power when sample size is small. On the other hand, the power of the proposed Exact-t method overlaps with the empirical power of the overall

BE test from a small sample size to a large sample size. One interesting finding from the simulation results (Figure 3 and Figure 7) is that when the effect size of REF vs. PLB is large, the equivalence test is the driving force of the power of the overall BE test; however, when the placebo effect is high, i.e., the effect size is small, superiority tests instead are the driving force for the overall power. One future research area is to explore different study designs to improve the efficiency of the overall BE test.

For simplicity, we assume homogeneous variance across treatment groups in this paper. A future research area is to release the assumption of homogeneous variance across treatment groups to heterogeneous variances, because active treatment groups are likely to have different variances from PLB. One challenge for heterogeneous variance is that with Satterthwaite's approximation (Satterthwaite, 1946) for the four t test statistics, there will be four different degrees of freedom for the four test statistics. A multivariate non-central t distribution, however, only allows one degree of freedom. We would need to specify an approximate degree of freedom for the multivariate non-central t distribution. One possible extension is to use Dannenburg et al's updated degree of freedom (Dannenberg et al, 1994), which remains to be explored.

In this paper, we assume a normal distribution. However, in reality, data usually deviates from normality. Therefore, another future research area is to extend the normal distribution to other distributions such as log-normal distribution, etc.

In summary, we propose an exact multivariate non-central t distribution to calculate power and sample size for a three-arm BE study with two superiority tests (TEST vs PLB, REF vs. PLB) and one equivalence test. It is an advance from Chang et al (2014)'s method in that

1. we extended the overall BE test from one superiority test and one equivalent test to two superiority tests and one equivalence test.

2. Our proposed Exact-t method is shown to be more accurate than the Z-ChiSquare method when sample size is small by simulation, is computationally more efficient, and does not require self writing codes to numerically calculate the conditional expectation of the normal distribution given a truncated chi-square distribution.

Therefore, we recommend using the proposed Exact-t method when calculate power and sample size for a three-arm clinical endpoint BE study.

Disclaimer

The opinions and information in this presentation are those of this presenter, and do not represent the views and/or policies of the U.S. Food and Drug Administration.

References

Chang, Y. W., Tsong, Y., Dong, X., & Zhao, Z. (2014). Sample size determination for a three-arm equivalence trial of normally distributed responses. *Journal of biopharmaceutical statistics, 24*(6), 1190-1202. https://doi.org/10.1080/10543406.2014.948552

Davit, B. M., Nwakama, P. E., Buehler, G. J., Conner, D. P., Haidar, S. H., Patel, D. T., ... & Woodcock, J. (2009). Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration. *Annals of Pharmacotherapy, 43*(10), 1583-1597. https://doi.org/10.1345/aph.1M141

Lawrence, X. Y., & Li, B. V. (Eds.). (2014). *FDA bioequivalence standards (Vol. 13)*. Springer.

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. https://doi.org/10.1007/BF01068419

Phillips, K. F. (1990). Power of the two one-sided tests procedure in bioequivalence. *Journal of pharmacokinetics and biopharmaceutics, 18*(2), 137-144. https://doi.org/10.1007/BF01063556

Chow, S. C., & Liu, J. P. (2008). *Design and analysis of clinical trials: concepts and methodologies (Vol. 507)*. John Wiley & Sons.

Hauschke, D., Kieser, M., Diletti, E., & Burke, M. (1999). Sample size determination for proving equivalence based on the ratio of two means for normally distributed data. *Statistics in Medicine, 18*(1), 93-105.

Shen, M., Russek-Cohen, E., & Slud, E. V. (2015). Exact calculation of power and sample size in bioequivalence studies using two one-sided tests. *Pharmaceutical statistics, 14*(2), 95-101.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110-114.

Hauck, W. W., & Anderson, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Biopharmaceutics, 12*(1), 83-91.

Dannenberg, O., Dette, H., & Munk, A. (1994). An extension of Welch's approximate t-solution to comparative bioequivalence trials. *Biometrika, 81*(1), 91-101.

Copyrights