MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1666 Published online 5 December 2014 in Wiley Online Library

Exact calculation of power and sample size in bioequivalence studies using two one-sided tests†

Meiyu Shen,[a]* Estelle Russek-Cohen,[b] and Eric V. Slud[c,d]

The number of subjects in a pharmacokinetic two-period two-treatment crossover bioequivalence study is typically small, most often less than 60. The most common approach to testing for bioequivalence is the two one-sided tests procedure. No explicit mathematical formula for the power function in the context of the two one-sided tests procedure exists in the statistical literature, although the exact power based on Owen's special case of bivariate noncentral *t*-distribution has been tabulated and graphed.

Several approximations have previously been published for the probability of rejection in the two one-sided tests procedure for crossover bioequivalence studies. These approximations and associated sample size formulas are reviewed in this article and compared for various parameter combinations with exact power formulas derived here, which are computed analytically as univariate integrals and which have been validated by Monte Carlo simulations. The exact formulas for power and sample size are shown to improve markedly in realistic parameter settings over the previous approximations.

Keywords: bioequivalence; single endpoint; sample size; power; crossover; pharmacokinetic study

1. INTRODUCTION

In a typical pharmacokinetic two-period two-treatment crossover bioequivalence study, subjects are randomized to one group receiving the reference drug (R) first or the other group receiving the test drug (T) first. Each group receives the complementary drug in the second period. Within each period for each subject a drug plasma concentration profile over time is obtained. Then *AUC* (area under the concentration–time curve), $C_{\max}$ (maximum concentration), and $T_{\max}$ (time to reach maximum concentration) for each subject from each treatment are obtained from the observed concentration–time profiles. Because previous experience indicates that log(*AUC*) and log($C_{\max}$) are approximately normally distributed‡, we assume this throughout and denote by μ_T and μ_R the mean of the response variable of interest used, respectively for the test product and reference product. The number of subjects in a pharmacokinetic two-period two-treatment crossover bioequivalence study is typically small, most often less than 60, with range from 12 to 170 according to the survey by Davit *et al.* [1] of 12 years of bioequivalence studies submitted to the United States Food and Drug Administration.

In order to conclude the bioequivalence of the test product and the reference product, we should reject the null hypothesis in the following hypothesis test [2]:

$$\begin{aligned} H_0 &: \mu_T - \mu_R \leqslant \theta_1 \text{ or } \mu_T - \mu_R \geqslant \theta_2 \\ H_a &: \theta_1 < \mu_T - \mu_R < \theta_2 \end{aligned} \tag{1}$$

Here, θ_1 and θ_2 are pre-specified constants, also called equivalence margins, and $\theta_1 < \theta_2$.

The null hypothesis, H_0, states that μ_T and μ_R are not equivalent. The alternative hypothesis representing equivalence, H_a, is the intersection of the two one-sided parameter regions, $\{\theta_1 < \mu_T - \mu_R\}$ and $\{\mu_T - \mu_R < \theta_2\}$.

No exact analytical formula has previously been published for the probability of rejection in the two one-sided tests procedure for crossover bioequivalence studies under general parameter settings, but there are several published approximations [3–6]. All of these approximations, as well as the analytical formulas derived in this paper, apply to the two one-sample tests procedure when the response variable of interest, log(*AUC*) or log($C_{\max}$), is normally distributed. Phillips [3] and Diletti, Hauschke, and Steinijans [4] calculated the power and the sample size using Owen's [7] special case of the bivariate noncentral *t*-distribution and presented some sample size tables

[a] *Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, USA*

[b] *Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, USA*

[c] *Mathematics Department, University of Maryland, College Park, USA*

[d] *Center for Statistical Research and Methodology, US Census Bureau, Maryland, USA*

**Correspondence to: Meiyu Shen, DBVI, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993, USA.*
E-mail: meiyu.shen@fda.hhs.gov

† *This article reflects the views of the authors and should not be construed to be those of the US Food and Drug Administration.*

‡ *All logarithms used in this paper are natural, to base e.*

and power graphs. Phillips' [3] calculations were for equivalence margins $\theta_2 = -\theta_1 = 0.20$, different from the values $\theta_2 = -\theta_1 = \log 1.25 = 0.2231$ recommended in [8], the Food and Drug Administration (FDA) Guidance for Industry. Kieser and Hauschke [5] proposed approximate methods for sample size determination, and Chow and Wang [6] derived approximate formulas for sample size calculation under crossover or parallel designs.

In Section 2 of this article, we present the background, notation, and standard assumptions in two-stage crossover designs for tests of bioequivalence, and in Section 3, we briefly derive the exact formula for the power of the two one-sided tests procedure for testing bioequivalence based on a univariate normally distributed response variable and indicate how numerical integration easily provides accurate numerical values for power and sample size. In Section 4, we compare the numerical results of the exact method with the tabulated sample size values of Diletti *et al.* [4], the power values tabulated by Phillips [3] for $\theta_2 = -\theta_1 = 0.20$, and the sample size and power values generated by the approximate method of Chow and Wang [6].

2. LINEAR MIXED EFFECT MODEL FOR A TWO-PERIOD TWO-TREATMENT CROSSOVER BIOEQUIVALENCE STUDY

In a two-period two-treatment crossover design, there is a washout period between two periods of observation to limit or remove any carryover effect. Let Y_{ijk} be the response (e.g., log(*AUC*)) of the *k*th subject in the *j*th period of the *i*th sequence in the two-period two-treatment crossover study, where $i = 1, 2$, $j = 1, 2$, and $k = 1, \ldots, n_i$. Then, the linear mixed effect model for Y_{ijk} is

$$Y_{ijk} = \gamma + S_{ik} + P_j + F_{ij} + \varepsilon_{ijk} \tag{2}$$

where γ is the overall mean; P_j is the fixed effect of period *j*; F_{ij} is the fixed effect of the formulation administered in period *j* of sequence *i*; S_{ik} is the random effect of subject *k* in sequence *i*; and ε_{ijk} is the random error. From the treatment assignments, we know that $F_{11} = F_{22} = \mu_R$ and $F_{12} = F_{21} = \mu_T$. The parameters can be estimated only subject to restrictions: $P_1 + P_2 = \mu_T + \mu_R = 0$, because otherwise the parameters P_1, P_2, μ_T, and μ_R are not separately identifiable. The S_{ik} and the ε_{ijk} are all independent normal random variables with mean 0. The variance of S_{ik} is σ_S^2 and the variance of ε_{ijk} is σ_T^2 if $i \neq j, \forall i, j = 1, 2$ for the test formulation and σ_R^2 if $i = j, \forall i, j = 1, 2$ for the reference formulations.

Let $\theta^* = \mu_T - \mu_R$, where μ_T and μ_R are the true means of the test and reference formulations, respectively. An estimator of θ^* is given as

$$\hat{D} = \frac{\bar{Y}_{12\bullet} - \bar{Y}_{11\bullet} + \bar{Y}_{21\bullet} - \bar{Y}_{22\bullet}}{2}. \tag{3}$$

The pooled estimate of variance is

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{k=1}^{n_1}\left(Y_{12k} - Y_{11k} - \left(\vec{Y}_{12\bullet} - \vec{Y}_{11\bullet}\right)\right)^2 + \sum_{k=1}^{n_2}\left(Y_{21k} - Y_{22k} - \left(\vec{Y}_{21\bullet} - \vec{Y}_{22\bullet}\right)\right)^2\right]. \tag{4}$$

Clearly, the estimator $\hat{D}$ is the average of the averages of the intra-subject difference between the test and the reference for the two sequences, and S^2 is a pooled estimate of the variance of the intra-subject difference. For this two-period two-treatment crossover design, $\hat{D}$ is a normally distributed unbiased estimate of θ^* with variance $\sigma_{\hat{D}}^2 = \frac{\sigma_R^2 + \sigma_T^2}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$, and $\frac{(n_1+n_2-2)\bullet S^2}{\sigma_T^2 + \sigma_R^2}$ is distributed as χ^2 with $\nu = n_1 + n_2 - 2$ degrees of freedom.

3. EXACT POWER FUNCTION AND THE JOINT PROBABILITY DENSITY FUNCTION OF T_1 AND T_2

First, we define test statistics $T_1 = \frac{\hat{D} - \theta_1}{\frac{S}{2}\bullet\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ and $T_2 = \frac{\hat{D} - \theta_2}{\frac{S}{2}\bullet\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ as the basis for the hypothesis test (1). To correct the misstatement of Liu and Li [9] that test statistics (T_1, T_2) do not have a joint density (see first two lines after Equation (3) in [9]), we simply observe that the function mapping $\left(\hat{D}, S^2\right)$ to (T_1, T_2), with domain $(-\infty, \infty) \times (0, \infty)$ and range $\{(t_1, t_2) : t_1 > t_2\}$, is both differentiable and differentially invertible, while the independent variables $\hat{D}$ and S^2 have normal and gamma densities, respectively, as described in the previous text. For an exposition of the change-of-variable formula for differentiable one-to-one transformations of random vectors with a joint density, see Formula 4.3.2 on page 158 and Theorem 2.1.8 on page 53 of [10]. The joint density for (T_1, T_2) can also be obtained directly by differentiation with respect to t_1 and t_2 on the region $t_1 > t_2$ of the joint cumulative distribution function derived and published by Owen [7, Section 5].

Although the exact power function can easily be developed as a double integral from the joint pdf of T_1 and T_2, we will derive a simpler form of the exact power function by integrating the conditional power given S^2. This alternative formula, in the same spirit as Formulas (8) and (12) in Section 5 of [7], is simpler because it involves only a univariate integral over a bounded interval, after recognizing that the conditional power given S^2 is a readily evaluated normal tail probability.

The exact power function can be written, in terms of the α and $1-\alpha$ quantiles $t_\nu(\alpha)$ and $t_\nu(1-\alpha) = -t_\nu(\alpha)$ of the *t* distribution with ν degrees of freedom, as

$$\begin{aligned}
&P_1\left(\theta^*, n_1, n_2, \sigma_T, \sigma_R\right) \\
&= P\left(\text{Reject}H_0 | \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right) \\
&= \mathrm{P}\left(\frac{\hat{D} - \theta_1}{\frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geqslant t_\alpha(\nu) \cap \frac{\hat{D} - \theta_2}{\frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leqslant t_{1-\alpha}(\nu) \,|\mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right) \\
&= \mathrm{P}\left(\theta_1 + t_\alpha(\nu) \bullet \frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leqslant \hat{D} \leqslant \theta_2 - t_\alpha(\nu) \bullet \frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \,|\mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right).
\end{aligned}$$

This last expression is the expectation over S of a conditional probability given S, which is expressed simply in terms of the cumulative normal distribution function as

$$E_S\left[\left\{\Phi\left(\frac{\theta_2-\theta^*}{\sigma_{\hat{D}}}-t_\alpha(\nu)\bullet\frac{S}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)-\Phi\left(\frac{\theta_1-\theta^*}{\sigma_{\hat{D}}}+t_\alpha(\nu)\bullet\frac{S}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)\right\}I\left(S\leqslant\frac{\theta_2-\theta_1}{t_\alpha(\nu)}\left(\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)^{-1}\right)\right],$$

where $t_\alpha(\nu)$ is the $1-\alpha$ quantile of the t-distribution with $\nu=n_1+n_2-2$ degrees of freedom, $\Phi(\cdot)$ is the standard normal cumulative distribution function, $I(\cdot)$ is the indicator function, and recall that $\sigma_{\hat{D}}^2=\frac{\sigma_R^2+\sigma_T^2}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$. Using the fact that $\frac{\nu\hat{S}^2}{\sigma_T^2+\sigma_R^2}\sim\chi^2(\nu)$ and $\hat{D}\sim N\left(\theta^*,\sigma_{\hat{D}}^2\right)$ are independent, we write the previous expectation over S explicitly as

$$\begin{aligned}&P_1(\theta^*,n_1,n_2,\sigma_T,\sigma_R)\\&=\int_0^{c_1}\left(\Phi\left(\frac{\theta_2-\theta^*}{\sqrt{\frac{\sigma_T^2+\sigma_R^2}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}-t_\alpha(\nu)\bullet\sqrt{\frac{x}{\nu}}\right)\right.\\&\left.-\Phi\left(\frac{\theta_1-\theta^*}{\sqrt{\frac{\sigma_T^2+\sigma_R^2}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}+t_\alpha(\nu)\bullet\sqrt{\frac{x}{\nu}}\right)\right)\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}dx,\end{aligned}\tag{5}$$

where $c_1=\frac{(\theta_2-\theta_1)^2\nu}{t_\alpha^2(\nu)(\sigma_T^2+\sigma_R^2)}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}$.

The gamma density and integrand written using the normal cumulative distribution function are readily evaluated, so the integral (5) is easily evaluated in any good statistical computing package: **R** code for it is given in Appendix A.

In the two one-sided tests procedure for one single variable, the power (P_1) is the probability of rejecting H_0 when H_a in (1) is true. For sample size determination and power of two one-sided tests procedure in the bioequivalence literature, Schuirmann [2], Diletti *et al.* [4], and Chow and Wang [6] all assumed that $n_1=n_2=n/2$ and $\sigma_T^2=\sigma_R^2=\sigma^2$. Following this convention, we will compare the exact power with the approximate power of Chow and Wang under the assumption that $n_1=n_2=n/2$ and $\sigma_T^2=\sigma_R^2=\sigma^2$. The exact power function becomes

$$\begin{aligned}&P_1\left(\theta^*,n,\sigma\right)\\&=\int_0^{c_2}\left(\Phi\left(\frac{\theta_2-\theta^*}{\sigma\bullet\sqrt{2/n}}-t_\alpha(\nu)\bullet\sqrt{\frac{x}{\nu}}\right)\right.\\&\left.-\Phi\left(\frac{\theta_1-\theta^*}{\sigma\bullet\sqrt{2/n}}+t_\alpha(\nu)\bullet\sqrt{\frac{x}{\nu}}\right)\right)\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}dx,\end{aligned}\tag{6}$$

where $c_2=\frac{n(\theta_2-\theta_1)^2\nu}{8t_\alpha^2(\nu)\sigma^2}$.

The power values computed numerically from (5) using a standard numerical-integration routine, **integrate** in **R** [11], were carefully checked both in terms of their own estimated error bounds and by comparing them with Monte Carlo simulations of rejections in the two one-sided tests procedure. As examples of the results, we found using 10^6 Monte Carlo replications (with corresponding simulation standard errors less than 0.0005), for a combination of θ^* equal to 0, 0.1, 0.2, and log(1.25), and of $\sigma=0.2$ and 0.3, that the simulated and exact values were always within 0.001 of one another, and that the numerical-integration error bounds were less than 0.0001 (usually by one or more orders of magnitude).

Because $\nu\left\{t_\alpha^2(\nu)\right\}$ is an increasing function of ν (see the proof in Appendix B), both the upper limit of integration and the integrand in the integral formula (6) are directly seen to be monotone increasing as a function of n, so the integral (6) itself is also monotone increasing in n. Hence, we use a bisection search or other numerical root-finder to find the required sample size n by first solving for the continuous value $n=n*$ at which the exact power (6) with all parameters held fixed is equal to 1-β and then rounding it up to the smallest even number $n\geqslant n*$. Code lines for doing this in **R** are also supplied in Appendix A.

4. COMPARISON OF EXACT POWER WITH APPROXIMATE POWER

4.1. Comparison of exact power function with approximate power function of Chow and Wang

4.1.1. Approximate power formulas of Chow and Wang. Assuming $P\left\{\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}\leqslant t_\alpha(\nu)+\frac{\theta_1-\theta^*}{S\bullet\sqrt{1/n}}\right\}\approx 0$, Chow and Wang [6] approximated the power function $P_1(\theta^*,n,\sigma)$ for the two one-sided tests procedure (1) when $\theta^*>0$ and θ^* is relatively big to σ by

$$P^{CW}=P\left\{\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}\leqslant\frac{\theta_2-\theta^*}{S\bullet\sqrt{1/n}}-t_\alpha(\nu)\,|\mu_T-\mu_R=\theta^*,\theta^*\in(\theta_1,\theta_2),\sigma\right\}.$$

Clearly, from the right sides of P^{CW} and $P_1(\theta^*,n,\sigma)$, P^{CW} overestimates $P_1(\theta^*,n,\sigma)$ because $P\left\{\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}\leqslant t_\alpha(\nu)+\frac{\theta_1-\theta^*}{S\bullet\sqrt{1/n}}\right\}\geqslant 0$. Replacing S by $\sigma\sqrt{2}$ in the right-hand expression for P^{CW}, [6] further approximated $P_1(\theta^*,n,\sigma)$ for $\theta_2=-\theta_1=\theta$ and relatively large $\theta^*>0$ by $P^{CW}\approx P\left\{\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}\leqslant\frac{\theta-\theta^*}{\sigma\bullet\sqrt{2/n}}-t_\alpha(n-2)\,|\mu_T-\mu_R=\theta^*,\theta^*\in(\theta_1,\theta_2),\sigma\right\}$.

With this replacement, for $\theta_2=-\theta_1=\theta$ and $\theta^*=0$, Chow and Wang approximated $P_1(\theta^*,n,\sigma)$ by $P^{CW}=P\left\{\left|\frac{\hat{D}}{S\bullet\sqrt{1/n}}\right|\leqslant\frac{\theta}{\sigma\bullet\sqrt{2/n}}-t_\alpha(\nu)\right\}$. Hence, the sample size formulas of Chow and Wang are

$$n\geqslant\begin{cases}\frac{2\sigma^2\left(t_\alpha(n-2)+t_\beta(n-2)\right)^2}{(\theta-|\theta^*|)^2}\ \text{if}\ \theta^*\neq 0\\ \frac{2\sigma^2\left(t_\alpha(n-2)+t_{\beta/2}(n-2)\right)^2}{\theta^2}\ \text{if}\ \theta^*=0\end{cases}\tag{7}$$

4.1.2. Comparison of exact power and approximate power of Chow and Wang. The approximate power function of Chow and Wang [6] tends to overestimate power by removing the upper-tail

bound on $\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}$, but because of the further replacement of sample standard deviation by true σ in the inequality for $\frac{\hat{D}-\theta^*}{S\bullet\sqrt{1/n}}$, their approximation does not overestimate for all possible parameter combinations. The differences between exact power and approximate powers are illustrated in Figures 1–4 for wide ranges of standard deviations and true mean differences.

As defined in Section 2, σ^2 is the variance of log-transformed data from the reference product and the coefficient of variation (*CV*) in the untransformed data is $CV = \sqrt{e^{\sigma^2} - 1}$.

The product is considered a highly variable drug when *CV* for untransformed *AUC* or *Cmax* is greater than 0.3 [12,13]. Regardless of the magnitude of the standard deviation σ from examination of many numerical results, the approximate power of Chow and Wang [6] seems to overestimate the exact power when both *n* and θ^* are very small. The power curve of Chow and Wang has a peak because their calculation used different formulas for power when θ^* is zero and nonzero. The difference between the approximate power of [6] and the exact power decreases as the total sample size increases for the

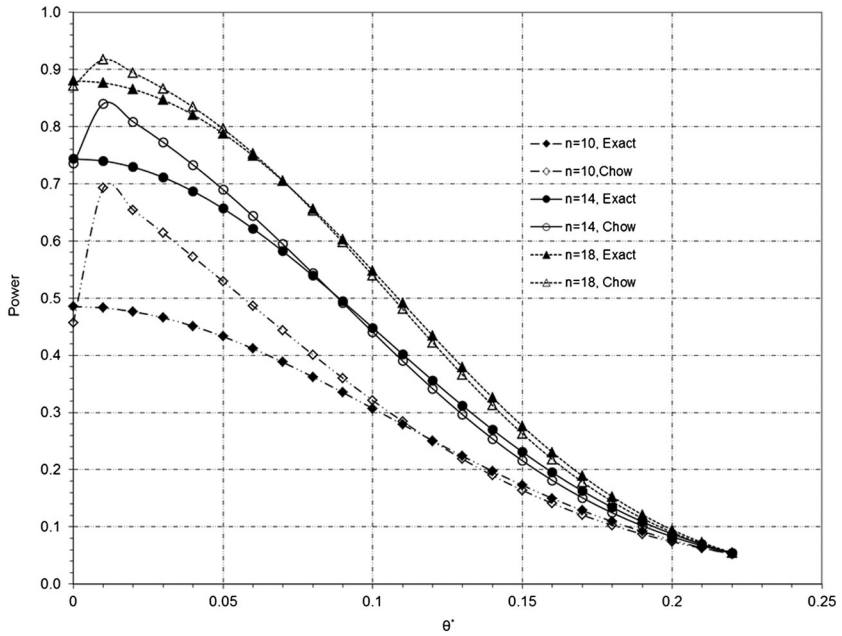

Figure 1. Approximate power of Chow and Wang versus exact power numerically calculated from Equation (6) against true mean difference θ^* when $\sigma = 0.2$, at different total sample sizes, *n*.

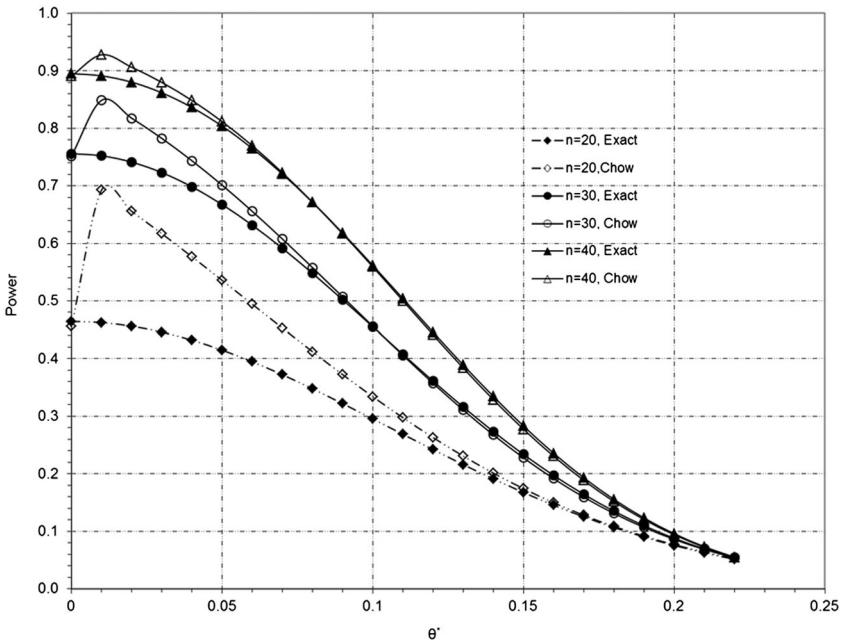

Figure 2. Approximate power of Chow and Wang versus exact power numerically calculated from Equation (6) against true mean difference θ^* when $\sigma = 0.3$, at different total sample sizes, *n*.

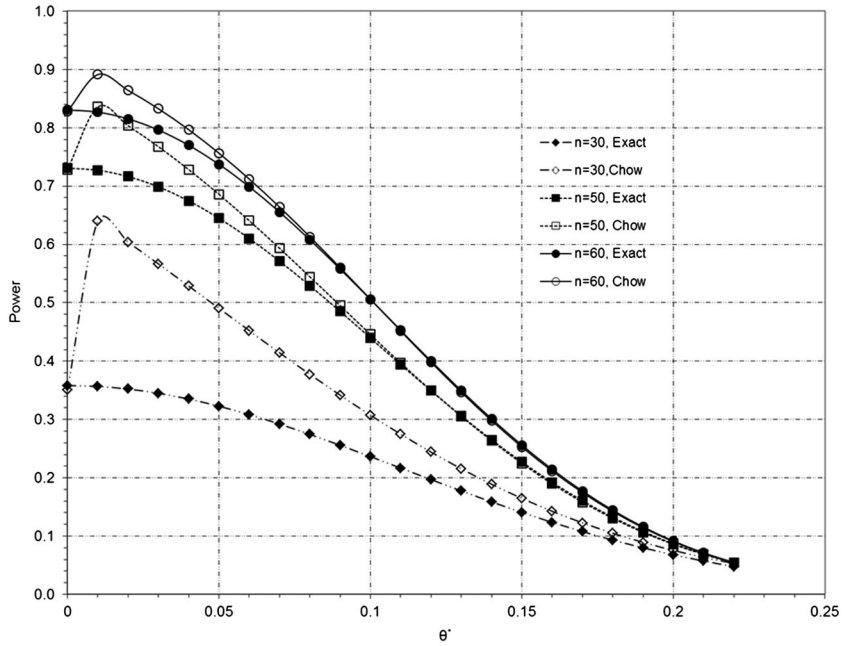

Figure 3. Approximate power of Chow and Wang versus exact power numerically calculated from Equation (6) against true mean difference θ^* when $\sigma = 0.4$, at different total sample sizes, *n*.

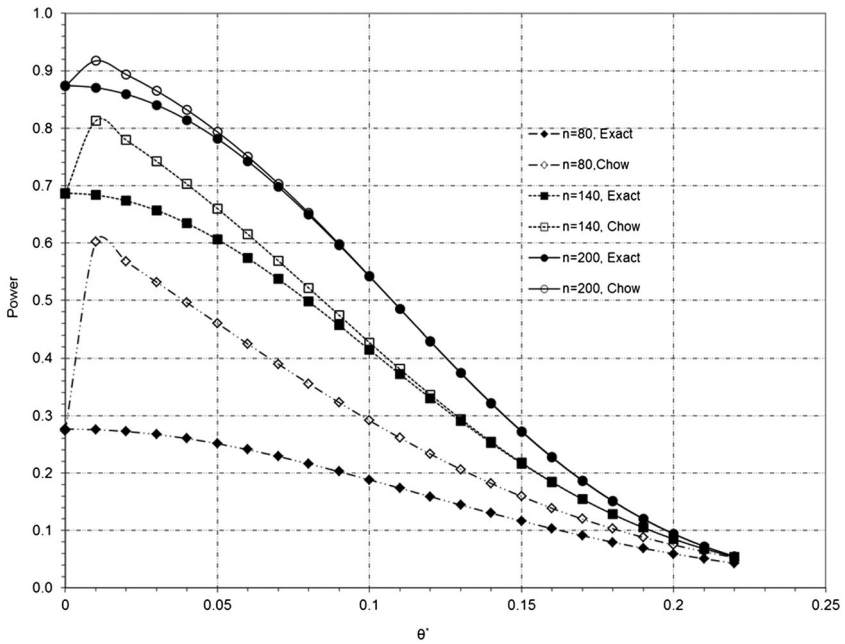

Figure 4. Approximate power of Chow and Wang versus exact power numerically calculated from Equation (6) against true mean difference θ^* when $\sigma = 0.7$, at different total sample sizes, *n*.

same σ and *n*, and also decreases as the true difference (θ^*) increases. When σ increases from 0.2 to 0.7, there are more cases when the approximate power of Chow and Wang exceeds the exact power.

Thus, sample sizes are underestimated using the approximate power of Chow and Wang for the combination of small *n* and small θ^*. When studies are underpowered, they may fail to meet the study objectives. For a combination of large θ^* and small *n*, the approximate power of Chow and Wang can underestimate the true power slightly, which results in having a few more subjects than necessary.

Previous authors [3,4,6] often focused on $CV<0.3$. However, errors of approximation may also be important for the large σ values of highly variable drugs.

From Figure 1, it is seen that the approximate power of Chow and Wang [6] is very close to the exact power for $n \geqslant 18$ and $\theta^*>0.04$ when $\sigma = 0.2$. From Figures 2–4, it is seen that the approximate power of Chow and Wang overestimates the exact

Table I. Comparison of sample size from exact power given by Equation (6) and approximate power of Chow and Wang for achieving 80% power at different combinations of σ and θ^*.

	$\theta^* = 0.01$		$\theta^* = 0.02$		$\theta^* = 0.03$		$\theta^* = 0.04$	
σ	Exact	Chow and Wang	Exact	Chow and Wang	Exact	Chow and Wang	Exact	Chow and Wang
0.1	6	6	6	6	6	6	6	6
0.2	16	14	16	14	18	16	18	18
0.3	34	28	34	30	36	32	38	36
0.4	58	46	60	50	62	56	66	62
0.5	90	70	92	78	94	86	100	94
0.6	128	100	130	110	136	122	144	136
0.7	172	136	176	150	184	164	194	184

Table II. Comparison of the exact power given by Equation (6) and power values from Figure 1(c) in Diletti and colleagues' paper when $CV = 20\%$ and $n = 24$.

	$\theta^* = \mu_T - \mu_R$		
Method	0	log(1.05)=0.04879	log(1.1)=0.09531
Diletti and colleagues' power	0.97	0.9	0.7
Exact power, $P_1(\theta^*, n, \sigma)$	0.9679	0.902	0.696

power for many more combinations of θ^* and n when $\sigma \geqslant 0.3$ than when $\sigma = 0.2$. For example, for $n = 50$, $\theta^* = 0.02$ and $\sigma = 0.4$, the approximate power of Chow and Wang is 80.35%, while the exact power is 71.63%.

4.1.3. Comparison of exact sample size with approximation of Chow and Wang. In Table I, we compare the total sample sizes from the exact power as in Equation (6) and the approximate power of Chow and Wang [6] for achieving 80% power. The total sample size is rounded up to the next even number. Table I shows that the difference in the total sample sizes increases as σ increases, and decreases as θ^* increases, for each given θ^*. For example, the total sample size from the Chow–Wang approximate power is about 10% less than that from exact power when $\theta^* = 0.03$ and $\sigma = 0.2$, about 15% less when $\theta^* = 0.02$ and $\sigma = 0.4$, and about 20% less than that from exact power when $\theta^* = 0.01$ and $\sigma = 0.3$.

4.2. Comparison of the exact power and Diletti and colleagues' power

Because Diletti *et al.* [4] did not provide the explicit mathematical formula for their power calculation, we read three power values from Figure 1(c) of [4]. The exact power values versus those read from Figure 1(c) of [4] for these parameters are listed in Table II.

From Table II, the power in [4] is very close to the exact value. However, the graphs and tables in [4] present a limited number of cases. For instance, [4] did not provide any power value for CV>30%, as would be the case with highly variable drugs. Exact and explicit power functions will allow such situations to be considered.

4.3. Comparison of the exact power and PhillipsŠ power

Because Phillips [3] did not provide the explicit mathematical formula for his power calculation, we also read power values for three sets of parameter values from Figure 3 of [3]. We then compared the exact power values by Equation (6) and the power values from Figure 3 [3] for these three sets of parameters and list these in Table III. In order to compare the exact power with Phillips' values, the exact values were computed using $\theta_2 = -\theta_1 = 0.20$.

Table III. Comparison of the exact power given by Equation (6) and power values from Figure 3 [3] when $CV = 20\%$ and $n = 24$.

	$\theta^* = \mu_T - \mu_R$		
Method	0.015	0.05	0.1
Phillips' power	0.9	0.8	0.5
Exact power, $P_1(\theta^*, n, \sigma)$	0.909	0.809	0.517

From Table III, Phillips' power is quite close to the exact power. However, the graphs and tables present a limited number of cases as Phillips [3] did. Exact and explicit power formulas, such as P_1 in Equations (5) and (6), allow any situation to be considered.

5. DISCUSSION AND CONCLUSIONS

The exact power has been derived from the joint density function of the two correlated test statistics (T_1 and T_2) used in the two one-sided tests procedure. The exact power numerically integrated from Equation (5) is corroborated by the results of the Monte Carlo simulations and is readily available using numerical integration, and **R** code for this is provided in Appendix A. Exact sample size calculation is then easy using the exact power for any parameter combinations in bioequivalence studies based on two one-sided tests. The Chow–Wang values are often used in standard software packages, for example, nQuery [14]. The fact that the approximate power of Chow and Wang [6] markedly overestimates the exact power for many combinations of θ^* and n as shown in Figures 1–4 and Table I demonstrates the preferability

of the exact power function $P_1(\theta^*, n, \sigma)$ in planning of bioequivalence studies, especially when θ^* is small relative to σ. Mean differences in log(*AUC*) or log(C_{max}) as small as 0.03 are of practical interest based on the review by Davit *et al.* [1] of 12 years of bioequivalence studies from the FDA, which found that more than 50% of studies have mean differences of less than 0.05 between generic product and innovator (these are mean differences in log(*AUC*) or log(C_{max}), when the corresponding mean differences in *AUC* and C_{max} between generic and innovator products are respectively 3.56% and 4.35%).

While the exact power agrees closely with [3,4] in the few cases displayed in those papers, it is important to have exact powers and sample sizes in all bioequivalence study settings, including those for highly variable drugs with $CV > 30\%$.

Acknowledgements

This research forms part of the PhD dissertation of the first author in the Statistics Program at the University of Maryland, under the joint guidance of the other authors. The first author is grateful for numerous discussions with Dr. Yi Tsong and thanks Dr. Yi Tsong and Dr. Stella Machado for their constant encouragement and support during her graduate studies while she has worked as a full time statistician at the FDA. Finally, the authors thank the anonymous reviewers for their comments.

REFERENCES

[1] Davit B, Nwakama P, Buehler G, Conner D, Haidar S, Patel D, Yang Y, Lawrence Y, Woodcock J. Comparing generic and innovator drugs: a review of 12 years of bioequivalence data from the United States Food and Drug Administration. *Annals of Pharmacotherapy* 2009; **43**:1583–1597.

[2] Schuirmann D. 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* 1987; **15**(6):657–680.

[3] Phillips KE. Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 1990; **18**:137–143.

[4] Diletti E, Hauschke D, Steinijans VW. Sample size determination for bioequivalence assessment by means of confidence intervals. *International Journal of Clinical Pharmacology, Therapy, and Toxicology* 1991; **29**:1–8.

[5] Kieser M, Hauschke D. Statistical methods for demonstrating equivalence in crossover trials based on the ratio of two location parameters. *Drug Information Journal* 2000; **34**:563–568.

[6] Chow SC, Wang H. On sample size calculation in bioequivalence trials. *Journal of Pharmacokinetics and Pharmacodynamics* 2001; **28**:155–169.

[7] Owen DB. A special case of a bivariate non-central t-distribution. *Biometrika* 1965; **52**:437–446.

[8] U.S. Food and Drug Administration Center for Drug Evaluation and Research. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, January 2001. Available at: http://www.fda.gov/downloads/Drugs/Guidances/ucm070244.pdf (accessed 25.11.2014).

[9] Liu F, Li Q. Exact Sequential Test of Equivalence Hypothesis Based on Bivariate Non-central T-statistics. *Computational Statistics and Data Analysis* 2014; **77**:14–24, Available at: http://dx.doi.org/10.1016/j.csda.2014.02.007.

[10] Casella G, Berger RL. *Statistical inference* (2nd edn). The Wadsworth Group, 2002.

[11] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2013. Available at: http://www.R-project.org/ (accessed 01.10.2014).

[12] Blume H, Midha KK. Report of Consensus Meeting: Bio-international '92. Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. Bad Homburg, Germany, 20–22 May, 1992. Eur. J. Pharm. Sci. 1993; 1: 165–171.

[13] Blume H, McGilveray I, Midha K. Report of Consensus Meeting: Bio-international '94, Conf. on Bioavailability, Bioequivalence and Pharmacokinetic Studies. Munich, 14–18 June, 1994. Eur. J. Pharm. Sci. 1995; 3: 113–124.

[14] Elashoff JD. *nQuery advisor, version 5.0 user's guide*. Statistical Solutions Ltd: Los Angeles, 2002.

APPENDIX A: R CODE FOR EXACT POWER AND SAMPLE SIZE

The power $P_1(\theta^*, n_1, n_2, \sigma_T, \sigma_R)$ in (5) is calculated by the following lines of R code. Here, the arguments n1 and n2 respectively represent the sample sizes $n_1, n_2$of sequences 1 and 2; sigT , sigR represent σ_T and σ_R; th1 and th2 respectively represent θ_1 and θ_2; and tstar represents θ^*. In fact, $(n_1, n_2, \sigma_T, \sigma_R)$ enters formula (5) only through the quantities $\sigma_{\hat{D}}^2 = \left(\sigma_R^2 + \sigma_T^2\right)(1/n_1 + 1/n_2)/4$ and $\nu = n_1 + n_2 - 2$, and arguments sigD and nu represent $\sigma_{\hat{D}}$ and ν . Pow1$value is the power in formula (5) (with Pow1$abs.error the estimated absolute error of integration), and SSiz denotes the sample size obtained by equating formula (6) to 1-beta. We use two standard R functions: **integrate** to perform univariate numerical integration, and *uniroot* to perform root-finding or inversion.

```
Pow1 = function(sigD, nu, th2,th1,tstar, alpha=.05){
  Integrand = function(x)    {
    (pnorm((th2-tstar)/sigD-qt(1-alpha, nu)*sqrt(x/nu))-
    pnorm((th1-tstar)/sigD+qt(1-alpha,
nu)*sqrt(x/nu)))*dchisq(x,nu)
  }
  integrate(Integrand, lower = 0, upper =
      nu*((th2-th1)/(2*qt(1-alpha, nu)*sigD))^2)[1:2]
}
  SSiz = function(beta, sigT,sigR,th2,th1,tstar, alpha=.05,
uppern=300){
    Pow2 = function(n) {
      Pow1(sqrt((sigT^2+sigR^2)/n), n-2, th2, th1, tstar,
alpha)$value-1+beta }
    uniroot(Pow2,c(3,uppern))$root
  }
```

APPENDIX B: PROOF OF MONOTONICITY OF $\sqrt{\nu}/t_\alpha(\nu)$

Define $t(\nu)/\sqrt{\nu} = Z/\sqrt{Y(\nu)}$, where Z is the standard normal random variable and $Y(\nu)$ is the chi-squared random variable with degrees of freedom. Because $Y(\nu) = \sum_{i=1}^{\nu} Z_i^2$, where $Z_1, Z_2, \ldots, Z_\nu$ are independent and identical standard normal random variables, then obviously $Y(\nu+1)$ is stochastically larger than $Y(\nu)$, which means for any real value x, $P(Y(\nu+1) \geq x) > P(Y(\nu) \geq x)$. From this, it follows that $t(\nu+1)/\sqrt{\nu+1}$ is stochastically smaller than $t(\nu)/\sqrt{\nu}$, from which it follows that $t_\alpha(\nu+1)/\sqrt{\nu+1} < t_\alpha(\nu)/\sqrt{\nu}$. Therefore, $\sqrt{\nu}/t_\alpha(\nu)$ is a monotonic increasing function of ν.