The International Journal of Biostatistics

Volume 5, Issue 1

2009

Article 15

Power for Testing Multiple Instances of the Two One-Sided Tests Procedure

Kem F. Phillips, *Sourland Biostatistics*

Recommended Citation:

Phillips, Kem F. (2009) "Power for Testing Multiple Instances of the Two One-Sided Tests Procedure," *The International Journal of Biostatistics*: Vol. 5: Iss. 1, Article 15.

DOI: 10.2202/1557-4679.1169

Power for Testing Multiple Instances of the Two One-Sided Tests Procedure

Kem F. Phillips

Abstract

The two one-sided tests procedure is used to test the equivalence of two measurements taken under different conditions. For example, two formulations of a drug are said to be bioequivalent if the average blood levels of the drug over time (AUC) are similar for the two formulations. In some studies there may be more than one parameter to test, such as a drug's AUC and maximum concentration, C_{max}, or AUCs from a parent drug and a metabolite. The power of testing two or more equivalence hypotheses simultaneously is less than the power to test any one hypothesis separately, and depends on the correlations of the measurements. This paper develops an exact mathematical formula for the power for two or more simultaneous comparisons for normally distributed variables when several comparisons are evaluated separately. The formula requires numerical integration with respect to the variance-covariance terms. These terms are distributed according to the Wishart distribution, and are integrated over a subset of positive-definite matrices defined by the equivalence criteria. An R program for the case of two comparisons is included.

KEYWORDS: power, TOST, simultaneous testing

1 Introduction

The two one-sided tests procedure (TOST) was formulated by Schuirmann (1986) as a test of bioequivalence. However, it is a general test of equivalence that can be used in other areas. For example, Barker et al (2002) advocated its use in testing "disparities in vaccination coverage", and Robinson and Froeseb (2004) used it for "equivalence testing to model validation using an empirical forest growth model." As an example of a multivariate test, Chervoneva et al (2007) analyzed spray pattern data with two or four equivalence criteria.

In an equivalence experiment, observations are obtained for a test condition and for a reference condition. TOST tests whether the difference of the means under these two conditions, $\mu = \mu_2 - \mu_1$, lies within a specified interval, (θ_1, θ_2) . In a bioequivalence experiment, the observations are usually AUCs, areas under blood concentration versus time curves, determined after administration of two formulations of a single drug. In bioequivalence, the equivalence interval is typically set to $(\theta_1, \theta_2) = (0.80, 1.25)$ when the data are analyzed after logarithmic transformation. The statistical analysis of data from bioequivalence experiments is discussed in detail by Patterson and Jones (2006).

The observations from the experiment would be analyzed by a suitable analysis of variance, from which we would obtain two random variables, the estimated mean difference, D , and the error variance, $\hat{\sigma}^2$. The error variance has ν degrees of freedom, which is based on the number of observations and the experimental design. We assume here that there are n observations for each condition. Equivalence, the alternative hypothesis, is declared if the 2α level confidence interval on the mean difference is contained in the equivalence interval:

$$\theta_1 \leq D - \hat{\sigma} \sqrt{2/n} t_{1-\alpha, \nu} \quad \text{and} \quad D + \hat{\sigma} \sqrt{2/n} t_{1-\alpha, \nu} \leq \theta_2 \quad (1)$$

The type 1 error of this test, that is, the probability of declaring equivalence when the true difference of the means is outside the equivalence interval, is less than α .

The power of TOST is the probability of declaring equivalence when equivalence holds, that is, the true difference of the means lies within (θ_1, θ_2) . Phillips (1990) used Owen's special case of the bivariate t-distribution (1965) to compute the power to demonstrate bioequivalence. Diletti et al (1991) extended this result to log-normally distributed data. Sample sizes can be determined from the power curves, based on a set criterion. For example, it may be required that there is 80% power to show bioequivalence when the true mean AUCs differ by 5%.

In some experiments there may be more than one parameter of importance, so that there are two equivalence hypotheses to test. For example, for bioequivalence it may be required to show that both the AUC and the maximum concentration, C_{max} , are equivalent between formulations. If the parent drug has an active metabolite, bioequivalence of both drugs may be required. An important aspect of any statistical test is its power against specified hypotheses. When parameters are correlated, as in the examples given above, the power to simultaneously test the parameters is not simply the product of their individual powers. This paper develops an exact formula for the power for any correlation structure based on separate, independent tests of the parameters.

Berger and Hsu reviewed methods for simultaneous tests of bioequivalence [1996], noting that Hauck et al [1995] had discussed testing multiple bioequivalence hypotheses independently. Berger and Hsu's interest was in developing uniformly most powerful tests, which are different from the simpler equivalence procedure of testing hypotheses on mean differences separately. They do not give formulae for power. For certain cases of more two and four equivalence comparisons, Chervoneva et al computed the power for their multivariate procedure using simulation. As these authors state, their equivalence procedure "combines a comparison of means and a comparison of variances into a single aggregate criterion," and is also different from procedure discussed here. However, since these procedures are based on the normal distribution, the current procedure might be extended to these cases if the region of integration, which will be seen below to be relatively simple, is determined.

2 Power for One Equivalence Test

We first calculate the power of one TOST using a new method that will be extended to two comparisons. The random variables D and $\hat{\sigma}^2$ are distributed as $D \sim N(\mu, \sqrt{\frac{2}{n}}\sigma^2)$ and $\frac{\nu}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(\nu)$, and they are stochastically independent. Therefore, their joint density, $f(D, \hat{\sigma} \mid \mu, \sigma)$, is the product of their densities.

Equivalence is declared if the 2α level confidence interval on the mean difference is contained in the equivalence interval, as formulated in inequalities 1. Those inequalities can be rewritten as

$$\hat{\sigma} \leq \frac{D - \theta_1}{t_{1-\alpha, \nu} \sqrt{2/n}} \quad \text{and} \quad \hat{\sigma} \leq -\frac{D - \theta_2}{t_{1-\alpha, \nu} \sqrt{2/n}} \quad (2)$$

Inequalities 2 define regions R_1 and R_2 in the $(D, \hat{\sigma})$ -plane whose triangular intersection leads to equivalence. This triangle is bounded by the D -axis

and two half-lines emanating from it. The peak occurs where

$$\hat{\sigma} = \frac{\theta_2 - \theta_1}{2t_{1-\alpha,\nu} \sqrt{2/n}} = L \quad (3)$$

A graph of this triangle is shown as Figure 1. Note that Berger and Hsu derived this triangle as part of a comparison of their procedure to TOST, but did not use it in an explicit power calculation.

Figure 1: Region Defining Equivalence for One Comparison

The power to demonstrate equivalence is the integral of the joint density of $(D, \hat{\sigma})$ over this triangle, that is,

$$\int_0^L \int_{a(\hat{\sigma})}^{b(\hat{\sigma})} f(D | \mu, \sigma) f(\hat{\sigma} | \sigma) dD d\hat{\sigma} \quad (4)$$

where $a(\hat{\sigma}) = \hat{\sigma} t_{1-\alpha,\nu} \sqrt{2/n} + \theta_1$ and $b(\hat{\sigma}) = -\hat{\sigma} t_{1-\alpha,\nu} \sqrt{2/n} + \theta_2$. Since D is normally distributed, this can be written as

$$\int_0^L \left(\int_{a(\hat{\sigma})}^{b(\hat{\sigma})} f(D | \mu, \sigma) dD \right) f(\hat{\sigma} | \sigma) d\hat{\sigma} \quad (5)$$

For a given $\hat{\sigma}$, the inner integral is just

$$\Delta(\hat{\sigma} | \mu, \sigma, n, \nu) = \Phi\left(\frac{b(\hat{\sigma}) - \mu}{\sqrt{2/n} \sigma}\right) - \Phi\left(\frac{a(\hat{\sigma}) - \mu}{\sqrt{2/n} \sigma}\right)$$

where Φ is the cumulative distribution of the standard normal distribution.

The density of $\hat{\sigma}$ is

$$f(\hat{\sigma} \mid \sigma, \nu) = \frac{\left(\frac{\sqrt{\nu}}{\sigma}\hat{\sigma}\right)^{\nu-1} e^{-(\frac{\sqrt{\nu}}{\sigma}\hat{\sigma})^2/2}}{2^{\nu/2-1}\Gamma(\frac{\nu}{2})} \frac{\sqrt{\nu}}{\sigma} \quad (6)$$

So the power is

$$\int_0^L \Delta(\hat{\sigma} \mid \mu, \sigma, n, \nu) \frac{\left(\frac{\sqrt{\nu}}{\sigma}\hat{\sigma}\right)^{\nu-1} e^{-(\frac{\sqrt{\nu}}{\sigma}\hat{\sigma})^2/2}}{2^{\nu/2-1}\Gamma(\frac{\nu}{2})} \frac{\sqrt{\nu}}{\sigma} d\hat{\sigma} \quad (7)$$

This formula has been implemented by the current author in the R computing language (2008) in the *MBESS* package (Kelly, 2008). The two one-sided tests procedure is implemented in the R *equivalence* package (Robinson, 2008).

3 Power for Two Equivalence Tests

As in the case of one comparison, we reject the j^{th} hypothesis of non-equivalence if the j^{th} confidence interval is contained in the j^{th} equivalence interval, $(\theta_{1j}, \theta_{2j})$. The calculations for two comparisons closely follow those for one comparison. We assume that the equivalence limits and the α -levels may be different for the two comparisons. We also assume that each subject would receive two formulations of the drug, $i = 1, 2$, and for each formulation, observations are obtained on two parameters, $j = 1, 2$. The data would be

$$\{d_i = \begin{pmatrix} d_{i1} \\ d_{i2} \end{pmatrix}, i = 1, \dots, n\} \quad (8)$$

where d_{ij} is the difference between formulations in parameter j for subject i . The mean difference vector, μ , is estimated by $D = \frac{1}{n} \sum_{i=1}^n d_i$. The variance-covariance matrix Σ is estimated by $\frac{1}{n-1}S$, where

$$S = \sum_{i=1}^n (d_i - D)(d_i - D)^T \quad (9)$$

We assume $d_i \sim N(\mu, \Sigma)$. The mean $(D_1, D_2)^T$ then has a multivariate normal distribution, $N(\mu, \Sigma/n)$ (Muirhead, 1982, page 80). S has the Wishart distribution, $W(n-1, \Sigma)$ with density $f_W(S \mid n-1, \Sigma)$ (Muirhead, page 85).

The standard deviation of d_{ij} is estimated by $\sqrt{\frac{1}{n-1}S_{jj}}$, and the standard error of μ_j is $\sqrt{\frac{1}{n-1}S_{jj}/n}$. Therefore, the α -level confidence interval on μ_j is

$$D_j \pm \sqrt{S_{jj}/(n-1)} \sqrt{1/n} t_{1-\alpha_j, \nu} \quad (10)$$

As with the inequalities 2, we can rearrange the confidence intervals to obtain

$$\begin{aligned} \sqrt{S_{11}/(n-1)} t_{1-\alpha_1, \nu} \sqrt{1/n} + \theta_{11} &\leq D_1 \\ &\leq -\sqrt{S_{11}/(n-1)} t_{1-\alpha_1, \nu} \sqrt{1/n} + \theta_{12} \end{aligned} \quad (11)$$

and

$$\begin{aligned} \sqrt{S_{22}/(n-1)} t_{1-\alpha_2, \nu} \sqrt{1/n} + \theta_{21} &\leq D_2 \\ &\leq -\sqrt{S_{22}/(n-1)} t_{1-\alpha_2, \nu} \sqrt{1/n} + \theta_{22} \end{aligned} \quad (12)$$

or, with the obvious notation, $b_{11} \leq D_1 \leq b_{12}$ and $b_{21} \leq D_2 \leq b_{22}$. That is, the vector $(D_1, D_2)^T$ must fall into a 2-dimensional rectangle defined by these inequalities.

As with one comparison, with $\hat{\sigma}$ replaced by S_{11} and S_{22} , we can rearrange inequalities 11 and 12 to show that S_{11} and S_{22} are constrained by

$$\sqrt{S_{jj}} \leq L_j = \sqrt{n-1} \frac{\theta_{j2} - \theta_{j1}}{2t_{1-\alpha_j, \nu} \sqrt{1/n}} \quad (13)$$

or $S_{11} \leq L_1^2$ and $S_{22} \leq L_2^2$. Muirhead (page 61, and Problem 2.8, page 77) states that integrals using the Wishart distribution can be considered as integrals with respect to Lebesgue measure over their free elements. In the case of a 2×2 matrix, there are three free elements, S_{11} , S_{12} , and S_{22} , and they are constrained by conditions guaranteeing that S is positive-definite. In particular, the determinant of S , $S_{11}S_{22} - S_{12}^2$, must be > 0 . This will be the case if and only if $S_{11} > 0$, $S_{22} > 0$, and $|S_{12}| < \sqrt{S_{11}S_{22}}$.

The probability of demonstrating equivalence is the integral of the joint density of D and S over the volume defined by the inequalities above. Since D and W are independent, the probability is

$$\begin{aligned} \int_0^{L_2^2} \int_0^{L_1^2} \int_{-\sqrt{S_{11}S_{22}}}^{\sqrt{S_{11}S_{22}}} \int_{b_{12}}^{b_{22}} \int_{b_{11}}^{b_{21}} f(D \mid \mu, n, \Sigma) \times \\ f_W(S \mid n-1, \Sigma) dD_1 dD_2 dS_{12} dS_{11} dS_{22} \end{aligned} \quad (14)$$

Since $f_W(S \mid n-1, \Sigma)$ does not contain D , this can be written as

$$\int_0^{L_2^2} \int_0^{L_1^2} \int_{-\sqrt{S_{11}S_{22}}}^{\sqrt{S_{11}S_{22}}} \left(\int_{b_{12}}^{b_{22}} \int_{b_{11}}^{b_{21}} f(D \mid \mu, n, \Sigma) dD_1 dD_2 \right) \times \\ f_W(S \mid n-1, \Sigma) dS_{12} dS_{11} dS_{22} \quad (15)$$

Let $\Delta(S \mid \mu, \Sigma, n)$ be the integral with respect to $(D_1, D_2)^T$. Then $\Delta(S \mid \mu, \Sigma, n)$ is simply the probability determined by the 2-dimensional normal distribution, $N(\mu, \Sigma/n)$, over the rectangle defined by the limits of integration.

Substituting $\Delta(S \mid \mu, \Sigma, n)$, the expression becomes

$$\int_0^{L_2^2} \int_0^{L_1^2} \int_{-\sqrt{S_{11}S_{22}}}^{\sqrt{S_{11}S_{22}}} \Delta(S \mid \mu, \Sigma, n) f_W(S \mid n-1, \Sigma) dS_{12} dS_{11} dS_{22} \quad (16)$$

Since $S_{12} < \sqrt{S_{11}S_{22}}$, it is also true that in this integral, $S_{12} < L_1L_2$, so that we can replace $\sqrt{S_{11}S_{22}}$ by L_1L_2 if we account for the values of S_{12} for which S is not positive-definite. This can be done using the indicator function $I()$, which is 1 if its argument is true, 0 otherwise. Then the integral can be written as

$$\int_0^{L_2^2} \int_0^{L_1^2} \int_{-L_1L_2}^{L_1L_2} I(|S_{12}| < \sqrt{S_{11}S_{22}}) \Delta(S \mid \mu, \Sigma, n) \times \\ f_W(S \mid n-1, \Sigma) dS_{12} dS_{11} dS_{22} \quad (17)$$

Since all the integration limits in this expression are constants, a general multidimensional integration routine can be used without evaluating the Wishart distribution when the matrix argument is not positive-definite.

The power function has several obvious characteristics. The power will be less than the power for either comparison alone. Power increases with n and with the correlation between the differences, and decreases with the mean difference between treatments, μ , and the variability. Also, the power for two comparisons is at most the minimum of the power of either comparison, and if the differences are uncorrelated, the power for two comparisons is just the square of the power for one comparison.

Using an iterative procedure, the power function could be used to set sample sizes with criteria similar to those used for one comparison. For example, on the logarithmic scale we may want 80% power to show that the ratios of the means of both parameters are in the interval (0.80, 1.25), given that the true ratios are 1.05. As usual, the variances of the differences for each parameter must be estimated, but, in addition, we must now specify the correlation of the differences. The correlation will usually be difficult to estimate, unless a similar experiment has been conducted. If there is no justifiable estimate

available, the best procedure is to set the correlation to a conservative, that is, low value that is consistent with the purpose of the study and available resources.

4 Implementation of the Formula for Two Comparisons

For two comparisons, functions in the R language were written to carry out the integration using the *adapt* function from the *adapt* package (Genz, 2008). These functions are given in the appendix. The adaptive algorithm can produce inaccurate results, especially for cases with small variances, large sample sizes, or high correlations. In such cases, adjustment of *adapt*'s convergence parameters increases the accuracy. Note also, that the functions are slow, sometimes requiring over one minute for one power calculation on a late-model PC. Kuonen (2003) discusses problems that may be encountered when using adaptive integration techniques.

Power values shown in Table 1 were computed for combinations of the parameters that are typical in bioequivalence studies. The values are tabled according to $\sqrt{\Sigma_{11}}/2 = \sqrt{\Sigma_{22}}/2$, the correlation, ρ , of d_{1i} and d_{2i} , the mean difference, $\mu_1 = \mu_2$, and the sample size. The corresponding σ is shown to put the variability on the usual scale. The calculations were done on the logarithmic scale. The bioequivalence interval was taken to be (0.80, 1.25) for both comparisons, and $\alpha_1 = \alpha_2 = 0.05$. The power values were validated using simulations and internal consistency as described above. The power for one comparison based on the root-mean-square error is also shown. To relate the formulae developed here to the power calculations for a single bioequivalence comparison, observe that data obtained from a two-period crossover study or other paired design could be analyzed by computing a 90% confidence interval on the mean difference using the observed mean and standard deviation of the differences. The standard deviation equals $\sqrt{2}$ times the root-mean-square error, $\hat{\sigma}$, from an analysis of variance with terms for treatment and subject. In the context of this article, the standard deviation of the difference corresponds to $\sqrt{\Sigma_{jj}}$.

These examples show that the power to test two simultaneous equivalence hypotheses can be accurately calculated using the exact formula for the power. However, the calculations must be done carefully, with due consideration of the convergence properties of the numerical integration technique that is chosen.

Table 1: Power for Demonstrating Bioequivalence for Various Parameters

		$\sigma = \sqrt{\Sigma_{11}/2} = \sqrt{\Sigma_{22}/2}, n$								
		0.15 ($\Sigma_{11} = 0.045$)			0.225 ($\Sigma_{11} = 0.010125$)			0.3 ($\Sigma_{11} = 0.18$)		
ρ	μ	12	24	36	12	24	36	12	24	36
0	1.00	0.850	0.998	>0.999	0.228	0.825	0.974	0.021	0.369	0.727
	1.05	0.706	0.976	0.999	0.184	0.678	0.888	0.018	0.292	0.583
	1.10	0.387	0.789	0.943	0.100	0.364	0.583	0.012	0.151	0.300
0.25	1.00	0.853	0.998	>0.999	0.236	0.828	0.974	0.022	0.376	0.730
	1.05	0.719	0.977	0.998	0.197	0.693	0.892	0.020	0.309	0.602
	1.10	0.420	0.800	0.945	0.118	0.400	0.608	0.014	0.178	0.337
0.50	1.00	0.860	0.998	0.999	0.260	0.833	0.974	0.029	0.401	0.746
	1.05	0.738	0.977	0.998	0.223	0.710	0.898	0.026	0.339	0.628
	1.10	0.458	0.814	0.947	0.144	0.437	0.637	0.020	0.211	0.378
0.75	1.00	0.873	0.999	>0.999	0.310	0.855	0.977	0.047	0.451	0.775
	1.05	0.764	0.980	0.998	0.271	0.745	0.910	0.043	0.390	0.667
	1.10	0.505	0.834	0.953	0.187	0.489	0.675	0.034	0.259	0.430
		Power for One Bioequivalence Test								
	1.00	0.922	0.999	>0.999	0.478	0.908	0.987	0.143	0.607	0.853
	1.05	0.840	0.988	0.999	0.429	0.823	0.942	0.133	0.541	0.763
	1.10	0.622	0.889	0.971	0.316	0.604	0.764	0.109	0.389	0.548

5 More than Two Simultaneous Equivalence Tests

Equation 17 can be extended to three or more simultaneous equivalence tests. The expression for the integral with respect to the treatment differences, Δ , carries over easily. For this case, Muirhead notes that to be positive-definite, for each pair of comparisons, (i, j) , the corresponding elements of S must satisfy $S_{ii}S_{jj} - S_{ij}^2 > 0$, and that the determinant of S itself must be > 0 . If there are c comparisons, denoting a positive determinant by $|S| > 0$ and extending the definitions of L_i and $\Delta(S | \mu, \Sigma, n)$ to c comparisons, equation 17 becomes

$$\int_0^{L_c^2} \cdots \int_0^{L_1^2} \int_{-L_{c-1}L_c}^{L_{c-1}L_c} \cdots \int_{-L_1L_2}^{L_1L_2} F(S) \times f_W(S | n-1, \Sigma) dS_{12} \cdots dS_{c-1,c} dS_{11} \cdots dS_{c,c} \quad (18)$$

where

$$F(S) \equiv I(|S_{12}| < \sqrt{S_{11}S_{22}}) \cdots I(|S_{c-1,c}| < \sqrt{S_{c-1,c-1}S_{c,c}}) \times I(|S| > 0) \Delta(S | \mu, \Sigma, n) \quad (19)$$

and the indicator functions and integrals with respect to the off-diagonal elements include all $c(c-1)/2$ such elements.

Note that there are $c(c+1)/2$ free elements in the Wishart matrix which require integration. That is, the number of repeated integrals with respect to S increases with the square of the number of comparisons, and is, for example, ten for four comparisons. The resulting computational burden quickly becomes prohibitive.

To determine the power when there are c comparisons, $c(c-1)/2$ correlations must be estimated. If each comparison has power p , then the power is at least p^c , the power under independence of the parameters. For example, for four comparisons, each with power 80%, the minimum would be 41%, and the maximum 80%. Since there will usually be non-zero, possibly negative correlations between the parameters, reasonable choices of the correlations will lead to lower sample sizes than imposed by independence. If there are results from earlier, appropriately designed experiments, they would be used to specify at least some of the correlations. Usually, however, these correlations would not be known accurately. In this respect they resemble the within-subject variability that must be specified in an ordinary cross-over experiment when no cross-over experiments have been conducted previously. Of course, the specified correlations must lead to a nonsingular Σ , that is, $|\Sigma| > 0$. Having to estimate these additional parameters increases the difficulty in accurately determining power and sample size. To set the sample size realistically, power calculations should be done with several choices of Σ .

A R Functions to Compute Power for Two Comparisons

These functions implement the formulas developed above for two comparisons. They require functions from the libraries *adapt*, *mvtnorm*, and *MCMCpack*.

Function: power_two_comparisons

```
function(alpha,logscale,ltheta,lmu,Sigma,n){

# power for two comparisons of the two one-sided tests procedure
# in equivalence on either the regular or the logarithmic scale

# The parameters are:
# alpha (vector):  alpha levels for the 2 t-tests
#                 Confidence level for full test = 1 - 2 alpha
# logscale:       FALSE=use regular scale;
```

```
# TRUE=use logarithmic scale
# ltheta: limits of equivalence interval (matrix)
# lmu (vector): treatment difference or ratio (for log scale)
# Sigma: covariance matrix of estimates of differences
# n: number of subjects
# power: output power for testing equivalence

nu<- n-1 # degrees of freedom
theta <- ltheta # limits and differences on regular scale
mu <- lmu
if (logscale){ # limits and differences on logarithmic scale
  theta <- log(ltheta)
  mu <- log(lmu)}

limits<-c(0,0,0) # compute limits of integration
power_t <- qt(1.0-alpha[1],nu,lower.tail=TRUE, log.p = FALSE)
limits[1] <- (theta[1,2] - theta[1,1])/(2.0*power_t*sqrt(1/n))
power_t <- qt(1.0-alpha[2],nu,lower.tail=TRUE, log.p = FALSE)
limits[3] <- (theta[2,2] - theta[2,1])/(2.0*power_t*sqrt(1/n))
limits<- nu*limits^2
limits[2]<-sqrt(limits[1]*limits[3]) # limit for covariance
upper<-limits
lower<-limits*c(0,-1,0)

power<-adapt(ndim=3,lower=lower,upper=upper,minpts=12000,
maxpts=NULL,wishart_marginal_density_by_delta_3dim,
eps=0.000001,alpha=alpha, theta=theta,Sigma=Sigma,
mu=mu,n=n,nu=nu)
# function adapt from package adapt
return(power)}
```

Function: wishart_marginal_density_by_delta

```
function (input_wishart,Sigma,alpha,theta,mu,n,nu){

# Compute the Wishart distribution and multiply it by Delta
# input_wishart is a 3-element vector for S[1,1],S[1,2],S[2,2]
# It is converted to a 2X2 matrix for input into dwish

wishmat<-matrix(c(input_wishart[1],input_wishart[2],
input_wishart[2],input_wishart[3]),nrow=2,ncol=2)
integrand<-0
if ( abs(wishmat[1,2]) < sqrt(wishmat[1,1]*wishmat[2,2]) )
{ integrand<-dwish(wishmat,nu,Sigma)
# Wishart density from package MCMCpack
lower<-c(0,0)
upper<-c(0,0)
lower[1]<- sqrt(input_wishart[1]/(n-1))*qt(1.0-alpha[1],
nu,lower.tail=TRUE, log.p = FALSE)*sqrt(1/n) + theta[1,1]
lower[2]<- sqrt(input_wishart[3]/(n-1))*qt(1.0-alpha[2],
```

Phillips: Power for Multiple Instances of TOST

```

    nu,lower.tail=TRUE, log.p = FALSE)*sqrt(1/n) + theta[2,1]
upper[1]<- -sqrt(input_wishart[1]/(n-1))*qt(1.0-alpha[1],
    nu,lower.tail=TRUE, log.p = FALSE)*sqrt(1/n) + theta[1,2]
upper[2]<- -sqrt(input_wishart[3]/(n-1))*qt(1.0-alpha[2],
    nu,lower.tail=TRUE, log.p = FALSE)*sqrt(1/n) + theta[2,2]

if ((lower[1]<upper[1]) & (lower[2]<upper[2]))
{ XSigma<-Sigma/n
  delta<- pmvnorm(lower=lower,upper=upper,
    mean=mu,sigma=XSigma)[1]}
# function pmvnorm from package \it mvtnorm \rm
else {delta<-0.0}
integrand<-integrand*delta}

return(integrand)}

# Example:

power_two_comparisons(c(0.05,0.05),TRUE,
  matrix(c(0.8,0.8,1.25,1.25),nrow=2,ncol=2),c(1.05,1.05),
  matrix(c(0.045,0.0,0.0,0.045),nrow=2,ncol=2),24)
Power value = 0.9762721

```

References

- [1] Barker, L.E., Luman, E.T., McCauley, M.M., and Chu, S.Y. (2002) Assessing Equivalence: An Alternative to the Use of Difference Tests for Measuring Disparities in Vaccination Coverage. *American Journal of Epidemiology* **156:11**, 1056-1061.
- [2] Berger, R.L. and Hsu, C.J. (1996). Bioequivalence Trials, Intersection-Union Tests, and Equivalence Confidence Sets. *Statistical Science* **11**, 283-319.
- [3] Chervoneva, I., Hyslop, T., and Hauck, W.W. (2007). A multivariate test for population bioequivalence. *Statistics in Medicine* **26**, 1208-1223.
- [4] Genz, A., Meyer, M., Lumley, T., and Maechler, M. (2008). adapt: Adaptive Quadrature in up to 20 dimensions. See <http://www.R-project.org>.
- [5] Hauck, W.W., Hyslop, T., Anderson, S., Bois, F.Y., and Tozer, T.N. (1995). Statistical and regulatory considerations for multiple measures in bioequivalence studies. *Clinical Research and Regulatory Affairs* **12**, 249-265.

- [6] Kelley, K. and Lai, K. (2008). MBESS: Methods for the Behavioral, Educational, and Social Sciences, 2008-2-14. URL <http://www.indiana.edu/~kenkel>, [//wiki.r-project.org/rwiki/doku.php?id=packages:cran:mbess](http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:mbess).
- [7] Kuonen, D. (2003). Numerical Integration in S-PLUS or R: A Survey. *Journal of Statistical Software*, **8:13**, 1-14, (<http://www.jstatsoft.org/v08/i13>).
- [8] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.
- [9] Owen, D.B. (1965). A special case of a non-central t-distribution. *Biometrika* **52**, 437-446.
- [10] Patterson, S. and Jones, B. (2006). *Bioequivalence and Statistics in Clinical Pharmacology*. Chapman & Hall/CRC.
- [11] Phillips, K.F. (1990). Power of the two one-sided tests procedure in bioequivalence. *J. of Pharmacokinetics and Biopharmaceutics* **18**, 137-144.
- [12] R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.
- [13] Robinson, A. (2008). R equivalence package. 2008-4-16. See <http://www.R-project.org>.
- [14] Robinson, A.P. and Froeseb, R.E. (2004). Model validation using equivalence tests. *Ecological Modelling* **176:3-4**, 349-358.
- [15] Schuirmann D.J. (1986). A new statistical procedure for testing equivalence in two-group bioavailability trials. *J. of Pharmacokinetics and Biopharmaceutics* **15**, 657-680.