

## Chapter 545

# Equivalence Tests for the Ratio of Two Means in a Higher-Order Cross-Over Design (Log-Normal Data)

---

## Introduction

This procedure calculates power and sample size of statistical tests of equivalence of two means of higher-order cross-over designs when the analysis uses a t-test or equivalent. The parameter of interest is the ratio of the two means. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen, Chow, and Li (1997). The designs covered in this chapter are analyzed using what is called the 'multiplicative model' in Chen et al (1997). The 'additive model' is covered in the procedure that uses differences. Note that the multiplicative model is simply an additive model applied to the logarithmic transform of the original data.

Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. This *cross-over* design may be analyzed by a TOST equivalence test to show that the two means do not differ by more than a small amount, called the margin of equivalence.

---

## Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to another. It is assumed that there is a *washout* period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design refers to two treatments (periods) and two *sequences* (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do not carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

## Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

### Balaam's Design

Balaam's design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<b><u>Sequence</u></b>	<b><u>Period 1</u></b>	<b><u>Period 2</u></b>
1	A	A
2	B	B
3	A	B
4	B	A

### Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<b><u>Sequence</u></b>	<b><u>Period 1</u></b>	<b><u>Period 2</u></b>	<b><u>Period 3</u></b>
1	A	B	B
2	B	A	A

### Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

<b><u>Sequence</u></b>	<b><u>Period 1</u></b>	<b><u>Period 2</u></b>	<b><u>Period 3</u></b>	<b><u>Period 4</u></b>
1	A	B	B	A
2	B	A	A	B

### Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

<b><u>Sequence</u></b>	<b><u>Period 1</u></b>	<b><u>Period 2</u></b>	<b><u>Period 3</u></b>	<b><u>Period 4</u></b>
1	A	A	B	B
2	B	B	A	A
1	A	B	B	A
2	B	A	A	B

## Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.

## Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

## Outline of an Equivalence Test

**PASS** follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialized notation for the discussion of these tests.

<b>Parameter</b>	<b>PASS Input/Output</b>	<b>Interpretation</b>
$\mu_T$	Not used	<i>Treatment mean.</i> This is the treatment mean.
$\mu_R$	Not used	<i>Reference mean.</i> This is the mean of a reference population.
$\phi_L, \phi_U$	RL, RU	<i>Margin of equivalence.</i> These limits define an interval of the ratio of the means in which their difference is so small that it may be ignored.
$\phi$	R1	<i>True ratio.</i> This is the value of $\phi = \mu_T / \mu_R$ at which the power is calculated.

Note that the actual values of  $\mu_T$  and  $\mu_R$  are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of non-equivalence is

$$H_0: \phi \leq \phi_L \text{ or } \phi \geq \phi_U, \text{ where } \phi_L < 1, \phi_U > 1.$$

The alternative hypothesis of equivalence is

$$H_1: \phi_L < \phi < \phi_U$$

## Log-Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as percentages, but differences must be interpreted as actual amounts in their original scale. Also, the distributions of variables such as CMax or AUC are often skewed. Hence, a common practice is to take the following steps in hypothesis testing.

1. State the statistical hypothesis in terms of a ratio.
2. Transform this into a hypothesis about the difference by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in the original ratio scale.

The details of step 2 for the alternative hypothesis are as follows:

$$\phi_L < \phi < \phi_U$$

$$\Rightarrow \phi_L < \left\{ \frac{\mu_T}{\mu_R} \right\} < \phi_U$$

$$\Rightarrow \ln(\phi_L) < \{\ln(\mu_T) - \ln(\mu_R)\} < \ln(\phi_U)$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

When performing an equivalence test on the difference between means, the usual procedure is to set the equivalence limits symmetrically above and below zero. Thus, the equivalence limits will be plus or minus an appropriate amount. The common practice is to do the same when the data are being analyzed on the log scale. However, when symmetric limits are set on the log scale, they do not translate to symmetric limits on the original scale. Instead, they translate to limits that are the inverses of each other.

Perhaps these concepts can best be understood by considering an example. Suppose the researchers have determined that the lower equivalence limit should be 80% on the original scale. Since they are planning to use a log scale for their analysis, they transform this limit into the log scale by taking the logarithm of 0.80. The result is -0.223144. Wanting symmetric limits, they set the upper equivalence limit to 0.223144. Exponentiating this value, they find that  $\exp(0.223144) = 1.25$ . Note that  $1/(0.80) = 1.25$ . Thus, the limits on the original scale are 80% and 125%, not 80% and 120%.

Using this procedure, appropriate equivalence limits for the ratio of two means can be easily determined. Here are a few sets of equivalence limits for ratios:

Specified Percent Change	Lower Limit Original Scale	Upper Limit Original Scale	Lower Limit Log Scale	Upper Limit Log Scale
-25%	75.0%	133.3%	-0.287682	0.287682
+25%	80.0%	125.0%	-0.223144	0.223144
-20%	80.0%	125.0%	-0.223144	0.223144
+20%	83.3%	120.0%	-0.182322	0.182322
-10%	90.0%	111.1%	-0.105361	0.105361
+10%	90.9%	110.0%	-0.095310	0.095310

## Equivalence Tests for the Ratio of Two Means in a Higher-Order Cross-Over Design (Log-Normal Data)

Note that negative percent-change values specify the lower limit first, while positive percent-change values specify the upper limit first. After the first limit is found, the other limit is calculated as its inverse.

## Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable  $X$  is the logarithm of the original variable  $Y$ . That is,  $X = \ln(Y)$  and  $Y = \exp(X)$ . Label the mean and variance of  $X$  as  $\mu_X$  and  $\sigma_X^2$ , respectively. Similarly, label the mean and variance of  $Y$  as  $\mu_Y$  and  $\sigma_Y^2$ , respectively. If  $X$  is normally distributed, then  $Y$  is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

$$\mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}}$$

$$\sigma_Y^2 = \mu_Y^2 (e^{\sigma_X^2} - 1)$$

From this relationship, the coefficient of variation of  $Y$  can be found to be

$$\begin{aligned} COV_Y &= \frac{\sqrt{\mu_Y^2 (e^{\sigma_X^2} - 1)}}{\mu_Y} \\ &= \sqrt{e^{\sigma_X^2} - 1} \\ &= \sqrt{e^{\sigma_W^2} - 1} \end{aligned}$$

where  $\sigma_W^2$  is the within mean square error from the analysis of variance of the logged data. Solving this relationship for  $\sigma_X^2$ , the standard deviation of  $X$  can be stated in terms of the coefficient of variation of  $Y$  as

$$\sigma_X = \sqrt{\ln(COV_Y^2 + 1)}$$

Similarly, the mean of  $X$  is

$$\mu_X = \ln\left(\frac{\mu_Y}{\sqrt{COV_Y^2 + 1}}\right)$$

Thus, the hypotheses can be stated in the original ( $Y$ ) scale and then power can be analyzed in the transformed ( $X$ ) scale.

## Test Statistics

The analysis for assessing equivalence using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. These confidence limits can then be compared to the equivalence limits to test for equivalence. We refer you to their book for details.

## Power Calculation

The power is given by

$$Power(\phi) = T_V \left( \frac{\ln(\phi_U) - |\ln(\phi)|}{CV_m \sqrt{b/n}} - t_{V,1-\alpha} \right) - T_V \left( t_{V,1-\alpha} - \frac{|\ln(\phi)| - \ln(\phi_L)}{CV_m \sqrt{b/n}} \right)$$

where

$$CV_m = \sqrt{\exp(MSE_{\text{logged}}) - 1},$$

and  $T$  represents the cumulative  $t$  distribution,  $V$  and  $b$  depend on the design, and  $n$  is the average number of subjects per sequence.

Note that the constants  $V$  and  $b$  depend on the design as follows:

<b><u>Design Type</u></b>	<b><u>Parameters (<math>V, b</math>)</u></b>
Balaam's Design	$V = 4n - 3, b = 2.$
Two-Sequence Dual Design	$V = 4n - 4, b = 3/4.$
Four-Period Design with Two Sequences	$V = 6n - 5, b = 11/20.$
Four-Period Design with Four Sequences	$V = 12n - 5, b = 1/4.$

## Example 1 – Finding Power

A company has opened a new manufacturing plant and wants to show that the drug produced in the new plant is equivalent to that produced in the older plant. A two-sequence, dual cross-over design will be used to test the equivalence of drugs produced at the two plants.

Researchers have decided to set the equivalence limits for the ratio at 0.80 and 1.25. Past experience leads the researchers to set the COV to 0.40. The significance level is 0.05. The power will be computed assuming that the true ratio is 0.96. Sample sizes between 10 and 80 will be included in the analysis.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Power</b>
Design Type.....	<b>3x2 (Three-Period, Two-Sequence Dual: ABB BAA)</b>
Alpha.....	<b>0.05</b>
N (Total Sample Size).....	<b>10 20 30 40 60 80</b>
RU (Upper Equivalence Limit) .....	<b>1.25</b>
RL (Lower Equivalence Limit) .....	<b>1/RU</b>
R1 (Mean Ratio) .....	<b>0.96</b>
COV (Coefficient of Variation).....	<b>0.40</b>

## Output

Click the Calculate button to perform the calculations and generate the following output.

## Numeric Reports

### Numeric Results

Solve For: [Power](#)  
 Design Type: Three-Period, Two-Sequence Dual  
 Treatment Sequences: ABB | BAA  
 Ratio:  $R = \mu_T / \mu_R = \text{Treatment Mean} / \text{Reference Mean}$

Power	Sample Size N	Equivalence Limits		Actual Ratio R1	Coefficient of Variation COV	Alpha
		Lower RL	Upper RU			
<a href="#">0.0000</a>	10	0.8	1.25	0.96	0.4	0.05
<a href="#">0.3051</a>	20	0.8	1.25	0.96	0.4	0.05
<a href="#">0.5858</a>	30	0.8	1.25	0.96	0.4	0.05
<a href="#">0.7483</a>	40	0.8	1.25	0.96	0.4	0.05
<a href="#">0.9035</a>	60	0.8	1.25	0.96	0.4	0.05
<a href="#">0.9627</a>	80	0.8	1.25	0.96	0.4	0.05

Power The probability of rejecting non-equivalence when the means are equivalent.  
 N The total number of subjects. They are divided evenly among all sequences.  
 RU and RL The upper and lower equivalence limits. Ratios between these limits are equivalent.  
 R1 The ratio of the means at which the power is computed.  
 COV The coefficient of variation on the original scale.  
 Alpha The probability of rejecting non-equivalence when the means are non-equivalent.

### Summary Statements

A three-period, two-sequence dual cross-over design (ABB | BAA) will be used to test whether the treatment mean ( $\mu_T$ ) is equivalent to the reference mean ( $\mu_R$ ), with mean ratio equivalence limits of 0.8 and 1.25 ( $H_0: R \leq 0.8$  or  $R \geq 1.25$  versus  $H_1: 0.8 < R < 1.25$ ,  $R = \mu_T / \mu_R$ ). The comparison will be made using two one-sided t-tests using a log-transformation, with an overall Type I error rate ( $\alpha$ ) of 0.05. The coefficient of variation on the original scale is assumed to be 0.4. To detect a ratio of means ( $\mu_T / \mu_R$ ) of 0.96, with a total sample size of 10 (allocated equally to the 2 sequences), the power is 0.



## Equivalence Tests for the Ratio of Two Means in a Higher-Order Cross-Over Design (Log-Normal Data)

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	10	13	3
20%	20	25	5
20%	30	38	8
20%	40	50	10
20%	60	75	15
20%	80	100	20

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$ , with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. $D = N' - N$ .

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 13 subjects should be enrolled to obtain a final sample size of 10 subjects.

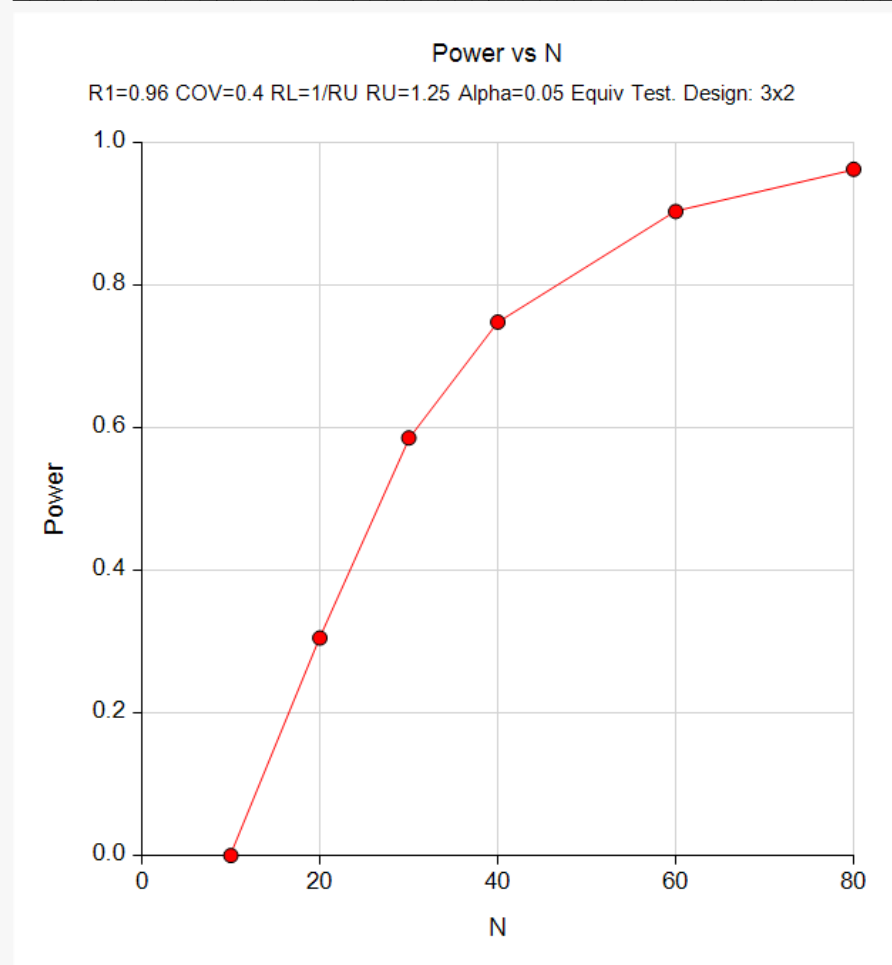
**References**

- Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York
- Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
- Chen, K.W., Chow, S.C., and Li, G. 1997. 'A Note on Sample Size Determination for Bioequivalence Studies with Higher-Order Crossover Designs.' Journal of Pharmacokinetics and Biopharmaceutics, Volume 25, No. 6, pages 753-765.

This report shows the power for the indicated scenarios. Note that 60 subjects will yield a power of just over 90%.

## Plots Section

### Plots



This plot shows the power versus the sample size.

## Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 90% power.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For ..... **Sample Size (Exact)**  
 Design Type..... **3x2 (Three-Period, Two-Sequence Dual: ABB|BAA)**  
 Power..... **0.80 0.90**  
 Alpha..... **0.05**  
 RU (Upper Equivalence Limit) ..... **1.25**  
 RL (Lower Equivalence Limit) ..... **1/RU**  
 R1 (Mean Ratio) ..... **0.96**  
 COV (Coefficient of Variation)..... **0.40**

### Output

Click the Calculate button to perform the calculations and generate the following output.

#### Numeric Results

Solve For: [Sample Size \(Exact\)](#)  
 Design Type: Three-Period, Two-Sequence Dual  
 Treatment Sequences: ABB | BAA  
 Ratio:  $R = \mu_T / \mu_R = \text{Treatment Mean} / \text{Reference Mean}$

Power	Sample Size N	Equivalence Limits		Actual Ratio R1	Coefficient of Variation COV	Alpha
		Lower RL	Upper RU			
0.8026	45	0.8	1.25	0.96	0.4	0.05
0.9035	60	0.8	1.25	0.96	0.4	0.05

We note that 60 subjects are needed to achieve 90% power and 45 subjects are needed to achieve at least 80% power.

## Example 3 – Validation using Chen et al. (1997)

Chen et al. (1997) page 761 presents a table of sample sizes for various parameter values for Balaam's design. We will reproduce entries from the first and seventeenth lines of the table in which the COV is 10%, the equivalence limits are 0.8 and 1.25, the actual ratio of between the means is 1, the power values are 80% and 90%, and the significance level is 0.05. Chen reports total sample sizes of 16 and 20. We will now setup this example in **PASS**.

The COV entered by Chen is the COV of the logged data. Since **PASS** requires the COV of the original data, we must use the relationship

$$\begin{aligned}
 COV_Y &= \sqrt{e^{\sigma_w^2} - 1} \\
 &= \sqrt{e^{0.1^2} - 1} \\
 &= \sqrt{e^{0.01} - 1} \\
 &= 0.10025
 \end{aligned}$$

to obtain the appropriate value of COV.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Sample Size (Equal Per Sequence)</b>
Design Type.....	<b>2x4 (Balaam: AA BB AB BA)</b>
Power.....	<b>0.80 0.90</b>
Alpha.....	<b>0.05</b>
RU (Upper Equivalence Limit) .....	<b>1.25</b>
RL (Lower Equivalence Limit) .....	<b>1/RU</b>
R1 (Mean Ratio) .....	<b>1.0</b>
COV (Coefficient of Variation).....	<b>0.10025</b>

## Output

Click the Calculate button to perform the calculations and generate the following output.

### Numeric Results

Solve For:	Sample Size (Equal per Sequence)
Design Type:	Two-Period, Four-Sequence (Balaam)
Treatment Sequences:	AA   BB   AB   BA
Ratio:	$R = \mu_T / \mu_R = \text{Treatment Mean} / \text{Reference Mean}$

Power	Sample Size N	Equivalence Limits		Actual Ratio R1	Coefficient of Variation COV	Alpha
		Lower RL	Upper RU			
0.8106	16	0.8	1.25	1	0.1003	0.05
0.9085	20	0.8	1.25	1	0.1003	0.05

Note that **PASS** has obtained the same samples sizes as Chen et al. (1997).