Equivalence and Noninferiority Tests

(2x2 Crossover Study)

statgraphics 18®
centurion

Revised: 10/10/2017

Summary

This procedure is used to demonstrate the equivalence of 2 treatments based on a 2x2 crossover study. In such a study, subjects are randomly assigned to 2 sequences. In one sequence, treatment #1 is applied first, followed by treatment #2. In the other sequence, treatment #2 is applied first followed by treatment #1. We wish to demonstrate equivalence between the means of the 2 treatments.

Sample StatFolio: *crossover.sgp*

Sample Data:

The file *crossover.sgd* contains the results of a crossover study published in Chow and Liu (2009). 24 patients were each given two drugs: a reference formulation and a test formulation. The file contains measurements made on each patient after receiving each of the drugs:

C:\DocData18\crossover.sgd

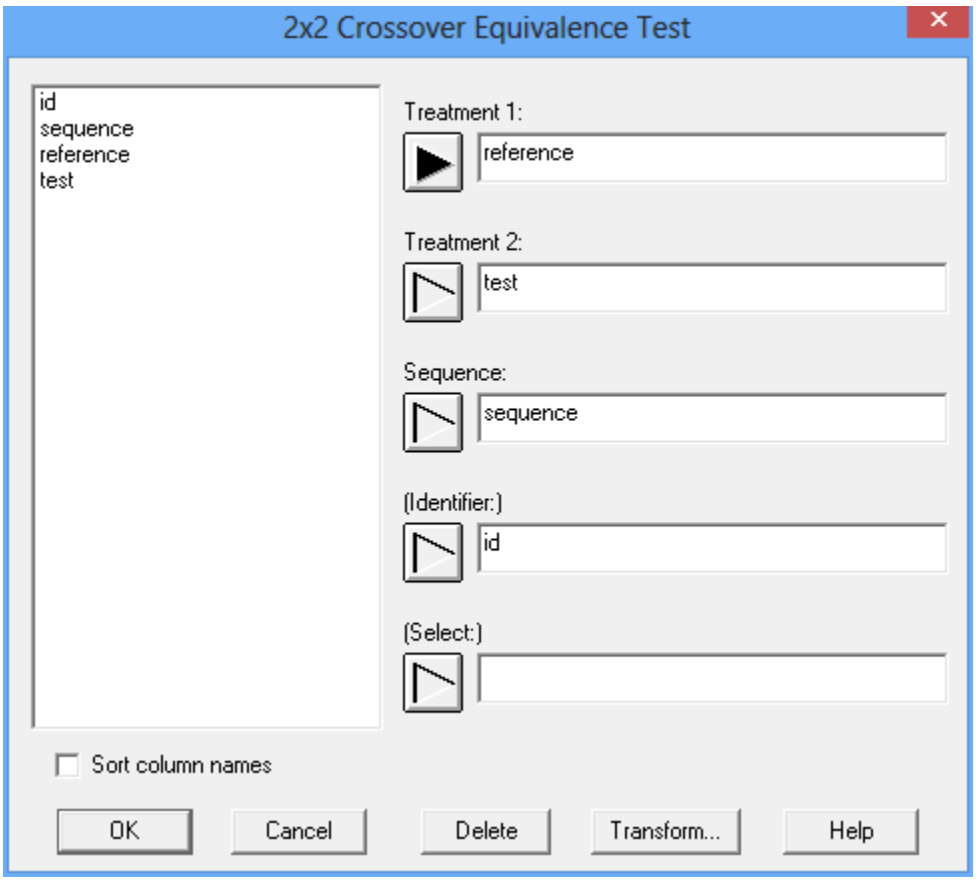
	id	sequence	reference	test
	Numeric	Character	Numeric	Numeric
1	1	RT	74.675	73.675
2	4	RT	96.4	93.25
3	5	RT	101.95	102.125
4	6	RT	79.05	69.45
5	11	RT	79.05	69.025
6	12	RT	85.95	68.7
7	15	RT	69.725	59.425
8	16	RT	86.275	76.125
9	19	RT	112.675	114.875
10	20	RT	99.525	116.25
11	23	RT	89.425	64.175
12	24	RT	55.175	74.575
13	2	TR	37.35	74.825
14	3	TR	51.925	86.875
15	7	TR	72.175	81.675
16	8	TR	77.5	92.7
17	9	TR	71.875	50.45
18	10	TR	94.025	66.125
19	13	TR	124.975	122.45
20	14	TR	85.225	99.075
21	17	TR	95.925	86.35
22	18	TR	67.1	49.925
23	21	TR	59.425	42.7
24	22	TR	114.05	91.725

crossover

Patients in sequence RT received the reference formulation first, while patients in sequence TR received the test formulation first.

Data Input

To perform the desired equivalence tests, select **Compare - Equivalence and Noninferiority Tests - 2x2 Crossover Study** from the main menu. The data input dialog box requests the names of columns containing the results of the 2 treatments, together with the name of a column identifying the sequences:



- **Treatment 1:** results for each subject when applying treatment #1.
- **Treatment 2:** results for each subject when applying treatment #2.
- **Sequence:** column containing 2 unique values identifying the 2 treatment sequences. The first unique value found in the column is assumed to be the sequence in which treatment 1 is applied before treatment 2. The second unique value found in the column is assumed to be the sequence in which treatment 2 is applied before treatment 1.
- **(Identifier):** optional identifier for each participant in the study.
- **(Select):** optional subset selection.

Note: When comparing a test formulation to a reference formulation, assign the reference formulation to treatment #1. This simplifies the interpretation of the calculated statistics.

Statistical Model

Let Y_{ijk} be the observed value for the i^{th} subject assigned to sequence k during period j, where i = 1, 2, …, n_k, j = 1,2, and k = 1,2. The total number of subjects in the study is the sum of the numbers of subjects assigned to each sequence

$$n = n_1 + n_2 \tag{1}$$

n_1 and n_2 may or may not be equal.

The general linear model for the 2x2 crossover study is

$$Y_{ijk} = \mu + S_{ik} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk} \tag{2}$$

where μ is a constant, S_{ik} is the random effect of the i^{th} patient assigned to sequence k, P_j is the effect of period j, $F_{(j,k)}$ is the effect of the treatment (formulation) applied during period j in sequence k, $C_{(j-1,k)}$ is the carryover effect of the treatment applied during period $j-1$ in sequence k, and e_{ijk} is a random error.

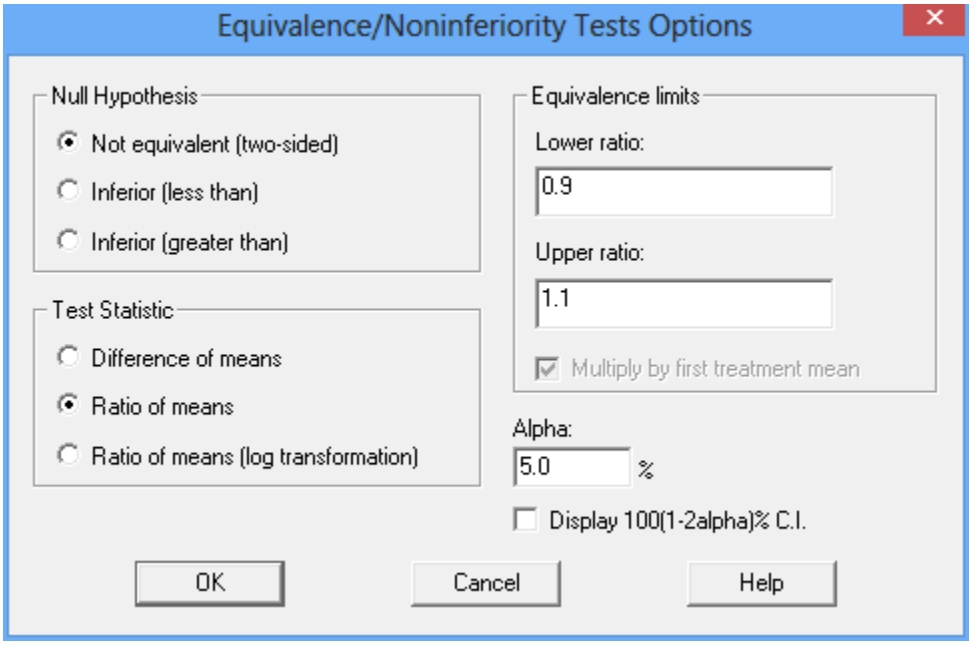
The model is best understood by examining the table below, which shows the means for each sequence during each of the 2 periods:

Sequence	Period 1	Period 2
RT	$\mu_{R1} = \mu + P_1 + F_R$	$\mu_{T2} = \mu + P_2 + F_T + C_R$
TR	$\mu_{T1} = \mu + P_1 + F_T$	$\mu_{R2} = \mu + P_2 + F_R + C_T$

where the subscripts R and T refer to the reference and test treatments, respectively. As with most linear models, we assume that the factor effects sum to 0, so $P_1 + P_2 = 0$, $F_T + F_R = 0$, and $C_T + C_R = 0$.

Analysis Options

Once the data are specified, a second dialog box is displayed on which to specify the hypothesis to be tested.



The most common type of test is a two-sided test of equivalence. In such a test, the null hypothesis is that the means of the two treatments being compared, μ_1 and μ_2, are not equivalent. By not equivalent, we mean that the difference between the means μ_2-μ_1 is either less than some lower differential Δ_L, or greater than some upper differential Δ_U.

Null hypothesis: $\mu_2 - \mu_1 < \Delta_L$ or $\mu_2 - \mu_1 > \Delta_U$

If this hypothesis is rejected, then we will have demonstrated that the difference between the means satisfies $\Delta_L \leq \mu_2 - \mu_1 \leq \Delta_U$, which is our definition of equivalence.

To demonstrate equivalence, Statgraphics uses the TOST procedure of Schuirman (1987). This procedure consists of two one-sided tests: an upper-tailed test used to demonstrate that $\mu_2 - \mu_1 \geq \Delta_L$ and a lower-tailed test used to demonstrate that $\mu_2 - \mu_1 \leq \Delta_U$. Obtaining significant results on both tests allows an assertion of equivalence between the means.

The fields on the *Analysis Options* dialog box specify:

- **Null hypothesis:** whether to perform a two-tailed test of equivalence as described above or a one-tailed test of noninferiority. In the latter case, the null hypothesis is one of the following:

  "Less than" null hypothesis: $\mu_2 - \mu_1 < \Delta_L$

"Greater than" null hypothesis: $\mu_2 - \mu_1 > \Delta_U$

- **Test Statistic:** whether to base the equivalence test on the difference between the test and reference means, the ratio of the test and reference means, or the ratio using a log transformation of the data values. If a test involving the ratio is selected, then the null hypothesis is changed to:

  Null hypothesis: $\mu_2 / \mu_1 < \Delta_L$ or $\mu_2 / \mu_1 > \Delta_U$

  with similar changes to the alternative hypothesis.

- **Equivalence limits:** the value of the lower differential Δ_L and the upper differential Δ_U.

- **Multiply by second treatment mean:** if checked, the lower and upper differentials are calculating by multiplying the equivalence limits specified above by the first treatment mean.

- **Alpha:** the significance level at which the tests will be performed.

- **Display 100(1-2alpha)% C.I.:** when displaying confidence intervals, use $(1-2\alpha)$ instead of $(1-\alpha)$.

Analysis Summary

The *Analysis Summary* displays sample statistics, estimated effects, and a test of equivalence or noninferiority. The top part of the output when applied to the sample data is shown below:

2x2 Crossover Equivalence Analysis
Treatment 1: reference
Treatment 2: test
Sequence: sequence
Identifier: id

Comparison method: reference mean - test mean
Sequence RT: reference, test
Sequence TR: test, reference

Sample Statistics

Treatment	*Sequence*	*n*	*Minimum*	*Maximum*	*Mean*	*Std. deviation*
reference	RT	12	55.175	112.675	85.8229	15.6913
reference	TR	12	37.35	124.975	79.2958	25.1979
reference	Pooled	24	37.35	124.975	82.5594	20.9899
test	RT	12	59.425	116.25	81.8042	19.7116
test	TR	12	42.7	122.45	78.7396	23.2071
test	Pooled	24	42.7	122.45	80.2719	21.5304

Standard deviation within subjects: 6.4662

Effects

	Effect	*Stnd. error*	*Df*	*t statistic*	*P-Value*
Carryover	-9.59167	15.6725	22	-0.6120	0.5468
Treatment	-2.2875	3.73326	22	-0.6127	0.5463
Period	-1.73125	3.73326	22	-0.4637	0.6474

The *Sample Statistics* table displays summary statistics for each of the 2 treatments, for each sequence separately and for the 2 sequences combined. For example, the test treatment averaged 78.7396 when applied first and 81.8042 when applied after the reference treatment. Note: the standard deviation for the *Combined* sequences is the pooled standard deviation of the data in the 2 sequences, not the standard deviation of the combined samples.

The *Effects* table displays statistical tests for the presence of carryover, treatment and period effects. The first line of the table tests the null hypothesis that $C_T = C_R = 0$, i.e., that there are no carryover effects. This is a very critical test, since it tests whether there was a long enough gap between the administration of the 2 treatments so that the effect of the first treatment vanished before the second treatment was applied. If carryover effects are present, the equivalence tests may not be valid. For the sample data, the P-value for the t-test is well above $\alpha = 0.05$, so the carryover effect is not statistically significant at the 5% significance level.

The second and third lines of the table perform t-tests to determine whether or not there are significant treatment and period effects, respectively. A significant period effect would be problematic, since it would imply that the effect of a treatment depended on whether it was

applied first or second. In this case, neither treatment nor period effects are statistically significant.

The second half of the output shows the results of an equivalence analysis comparing the 2 treatments.

Equivalence Analysis
Null hypothesis: Not equivalent (two-sided)
Lower equivalence ratio: 0.9
Upper equivalence ratio: 1.1

Comparison	*Ratio*	*Lower 90% CL*	*Upper 90% CL*
test / reference	0.972293	0.897871	1.05193

Comparison	Lower t-value	Upper t-value	Lower P-value	Upper P-value
test / reference	1.66674	-2.68508	0.0549	0.0068

Comparison	Maximum P-value	Conclusion (alpha=5%)
test / reference	0.0549	Equivalence has not been demonstrated.

It shows:

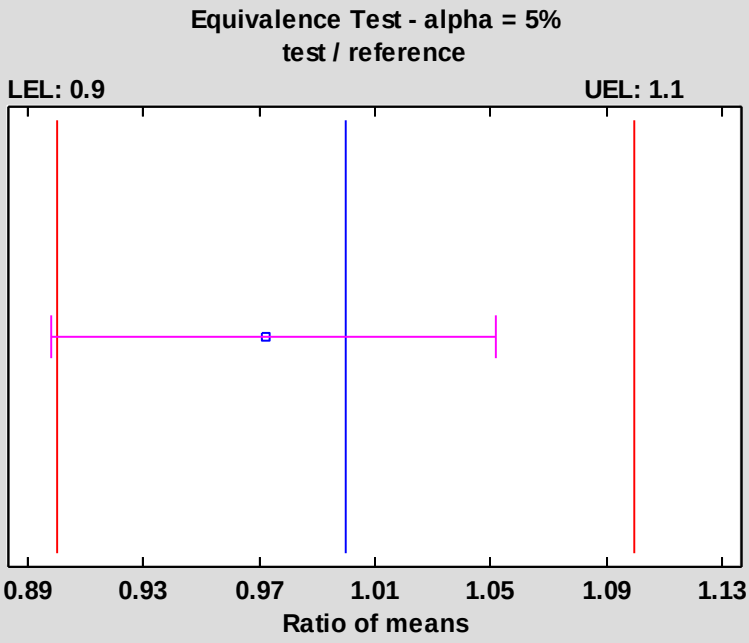
- *Null hypothesis:* the selected null hypothesis to be tested.
- *Lower and upper equivalence ratio:* if the ratio of the 2 treatment means is within this interval, the treatments will be considered to be equivalent.
- *Ratio:* the estimated ratio of the test mean to the reference mean:

$$\hat{\Delta} = \hat{\mu}_2 / \hat{\mu}_1 \qquad (3)$$

- *Lower and Upper CL:* lower and upper confidence limits for the ratio of the means. If the confidence interval is entirely within the equivalence range, then equivalence can be asserted. Otherwise, it cannot. For the sample data, the lower confidence limit is less than the lower equivalence ratio, so equivalence between the treatments cannot be claimed.
- *Lower and upper t-value:* the results of 2 one-sided t-tests, one comparing the estimated ratio to the lower equivalence ratio and the other comparing it to the upper equivalence ratio. If both P-values are less α, then equivalence can be asserted.
- *Maximum P-value:* the larger of the 2 P-values calculated above and a statement about whether equivalence between the treatment means has been demonstrated.

Equivalence Plot

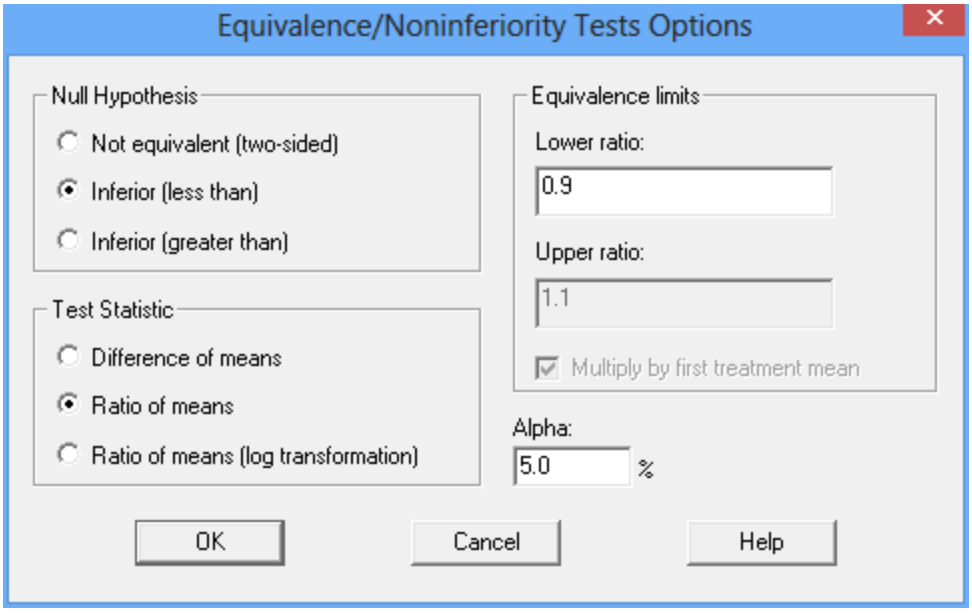
This plot shows the confidence interval for the ratio of the 2 treatment means. If the interval is contained entirely in the region between the lower and upper equivalence limits, then the means may be asserted to be equivalent.



One-Sided Noninferiority Tests

In some circumstances, the desired goal is not one of showing that the difference or ratio of 2 means is within some specified range. Instead, the goal is either to show that the difference or ratio is no bigger than some value Δ_U or to show that the difference or ratio is no smaller than some value Δ_L. Rejection of a null hypothesis in such a one-sided situation leads to the assertion that one treatment is not inferior to the other (it might be either equivalent or superior).

For example, suppose it was desired to show that the mean of the test treatment was no less than 90% of the reference treatment. In such a case, the *Analysis Options* dialog box would be completed as shown below:



In this case, the null hypothesis is $\mu_2 / \mu_1 < 0.9$. If this hypothesis can be rejected, then we can claim that the test treatment mean is not inferior to the reference treatment mean.

For the sample data, the relevant section of the *Analysis Summary* is shown below:

Equivalence Analysis
Null hypothesis: Inferior (less than)
Lower equivalence ratio: 0.9

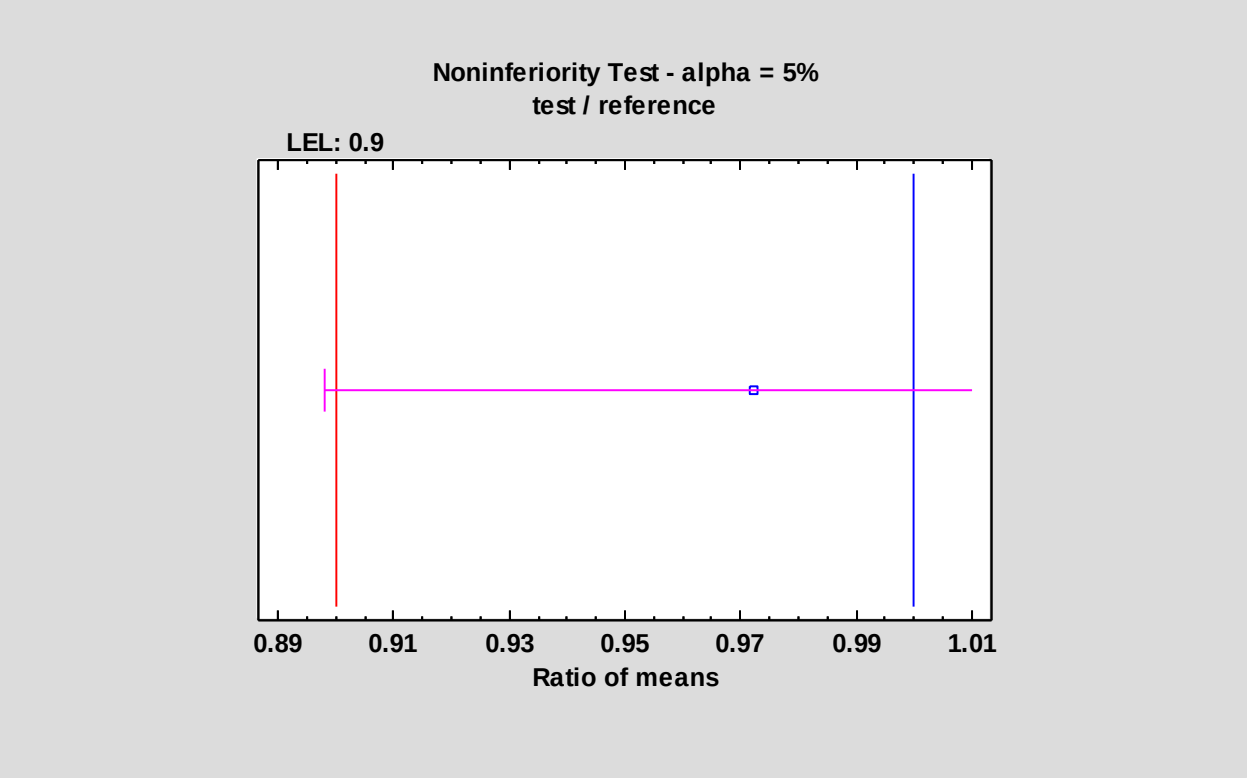
Comparison	*Ratio*	*Lower 95% CL*
test / reference	0.972293	0.897871

Comparison	Lower t-value	Lower P-value
test / reference	1.66674	0.0549

Comparison	Maximum P-value	Conclusion (alpha=5%)
test / reference	0.0549	Noninferiority has not been demonstrated.

The output shows a 95% lower confidence bound for the ratio of the means. Since that confidence bound is not greater than the lower equivalence ratio, noninferiority has not been demonstrated. A one-sided is also performed comparing the ratio to the lower equivalence ratio. Since the P-value is less than α=0.05, we have not demonstrated at the 5% significance level that the ratio of the test and reference means is greater than 0.9.

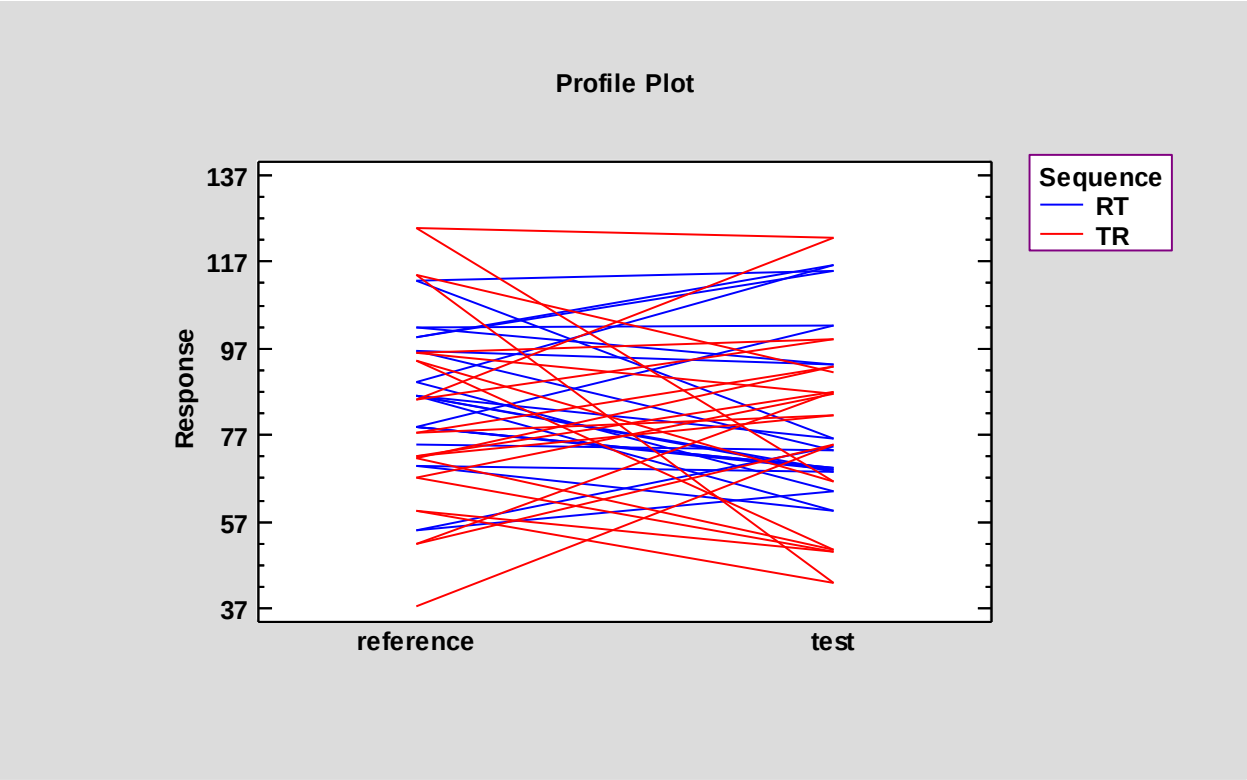
The *Equivalence Plot* displays the one-sided confidence bound for the ratio:



Noninferiority may be not asserted in this case since the lower confidence bound if not greater than the lower equivalence ratio.

Profile Plot

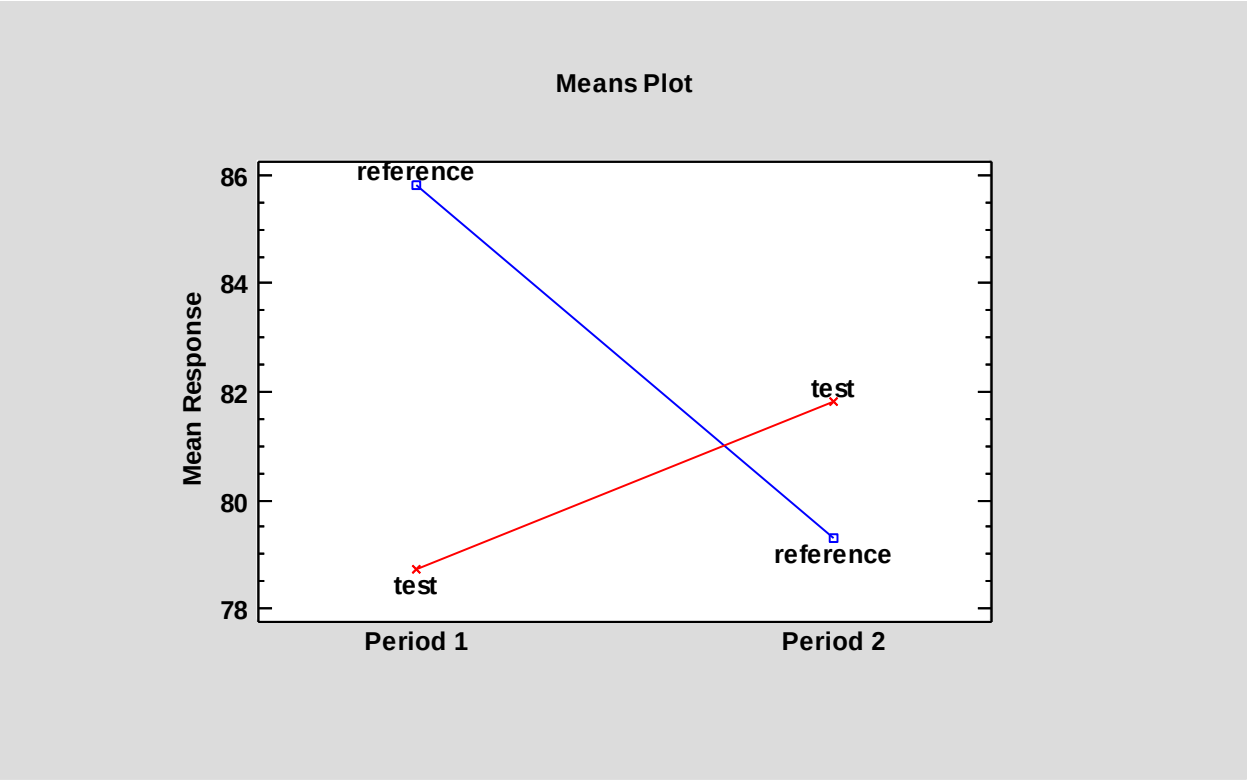
This plot draws a line connecting the 2 results for each subject.



If the lines associated with the 2 sequences go in predominantly different directions, the order of application of the treatments may be affecting the results. The plot can also illustrate differences in variability between the 2 sequences.

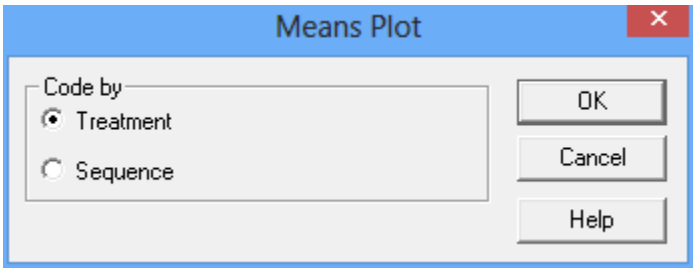
Means Plot

This plot displays the means for each combination of treatment and period.

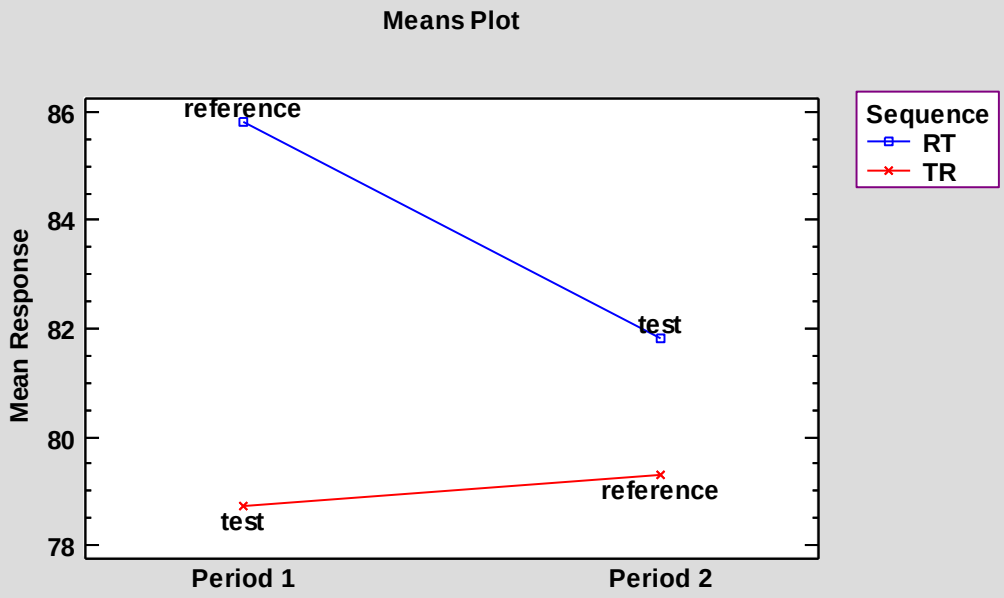


It illustrates well any treatment or period effect.

Pane Options

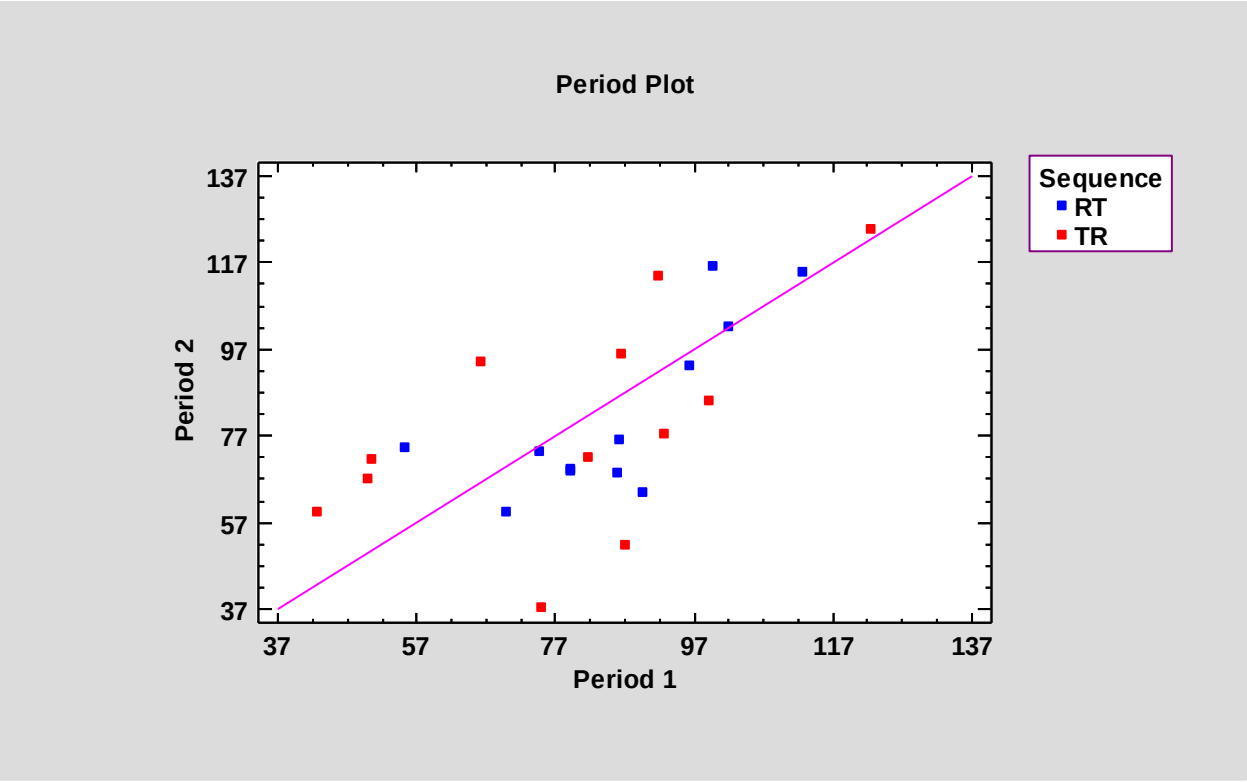


Specify the factor to be used to define the line connectors. Selecting *Sequence* creates a plot similar to that shown below.


Means Plot
Mean Response
86
84
82
80
78
reference
test
test
reference
Period 1
Period 2
Sequence
RT
TR

Period Plot

This plot contains a point for each subject in the study. It shows the values obtained during the 2 periods:



The plot may be used to identify outliers and other unusual effects.

References

Berger, R.L. and Hsu, J.C. (1995). "Bioequivalence trials, intersection-union tests, and equivalence confidence sets." Institute of Statistics Mimeo Series Number 2279.

Chow, S.C., and J.P. Liu. (2009). Design and Analysis of Bioavailability and Bioequivalence Studies. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.

Chow, S.-H. and Shao, J. (2002). Statistics in Drug Research: Methodologies and Recent Developments. New York: Marcel-Dekker.

Hsu, J.C., Hwang, J.T.G., Liu, H.-K., and Ruberg, S.J. (1994). "Confidence intervals associated with tests for bioequivalence." Biometrika 81: 103-114.

Locke, C.S. (1984). "An exact confidence interval for untransformed data for the ratio of two formulation means." J Pharmacokinet Biopharm 12: 649-655.

Schuirmann, D.J. (1987). "A comparison of the one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability." J Pharmacokinet Biopharm 15: 657-680.

Calculations

Carryover effect

Calculate the sum of the observed values for each subject in each sequence:

$$u_{ik} = Y_{i1k} + Y_{i2k} \tag{4}$$

Next calculate the average of the sums for each sequence $\bar{u}_k$ and the standard deviation s_k. The estimated carryover effect is given by

$$\hat{C} = \bar{u}_2 - \bar{u}_1 \tag{5}$$

The estimated standard error of the carryover effect is given by

$$\hat{\sigma}_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{6}$$

where $\hat{\sigma}_u^2$ is the pooled variance for the sums in the 2 sequences. Confidence intervals and hypothesis tests for the carryover effect are constructed using a t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Treatment effect

Calculate one-half of the within-subject differences for each subject in each sequence:

$$d_{ik} = (Y_{i2k} - Y_{i1k}) / 2 \quad (7)$$

Next calculate the average of the differences for each sequence $\bar{d}_k$ and the standard deviation s_k. The estimated treatment or formulation effect is given by

$$\hat{F} = \bar{d}_1 - \bar{d}_2 \quad (8)$$

The estimated standard error of the treatment effect is given by

$$\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (9)$$

where $\hat{\sigma}_d^2$ is the pooled variance for the differences in the 2 sequences. Confidence intervals and hypothesis tests for the treatment effect are constructed using a t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Period Effect

Using the differences calculated above, the estimated period effect is given by

$$\hat{P} = \bar{d}_1 + \bar{d}_2 \quad (10)$$

The estimated standard error of the period effect is given by

$$\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (11)$$

where $\hat{\sigma}_d^2$ is the pooled variance for the differences in the 2 sequences. Confidence intervals and hypothesis tests for the period effect are constructed using a t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

Confidence Interval

A *100(1-2α)*% confidence interval is calculated for either the difference between the test and reference means or the ratio of the test and reference means. When estimating the difference, the confidence interval is given by

$$\left[min\left(0, \hat{\Delta} - t_{\propto,\upsilon} SE\right), max\left(0, \hat{\Delta} + t_{\propto,\upsilon} SE\right)\right] \quad (12)$$

where

$$\upsilon = n_1 + n_2 - 2 \text{ `} \quad (13)$$

$$SE = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (14)$$

and s_p equals the pooled standard deviation of the within-subject differences in the 2 sequences. When estimating the ratio based on a log transformation of the data, confidence limits are calculated for the difference of the mean logs and inverted to create confidence limits for the ratio. When estimating the ratio directly, Fieller's theorem is applied using the method outlined by Locke (1984). Note that there are some data sets for which this latter approach cannot be used.

t-tests

Two one-sided t tests are performed to test for equivalence using Schuirmann's procedure. The t statistics are given by

$$t_L = \frac{\hat{\Delta} - LEL}{SE} \quad (15)$$

$$t_U = \frac{\hat{\Delta} - UEL}{SE} \quad (16)$$

They are compared to a t distribution with *v* degrees of freedom. When estimating the ratio based on a log transformation of the data, the t tests are performed in the log scale.

