Chapter 528

Superiority by a Margin Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Introduction

This procedure calculates power and sample size for non-zero null tests which use the difference in the means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

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 no 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

Sequence	Period 1	Period 2
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

Sequence	Period 1	Period 2	Period 3
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

Sequence	Period 1	Period 2	Period 3	Period 4
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

Sequence	Period 1	Period 2	Period 3	Period 4
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.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

$$H_0: \delta \leq A \quad \text{versus} \quad H_1: \delta > A$$

Rejecting H0 implies that the mean is larger than the value *A*. This test is called an *upper-tailed test* because H0 is rejected only in samples in which the difference in sample means is larger than *A.*

Following is an example of a *lower-tailed test*.

$$H_0: \delta \geq A \quad \text{versus} \quad H_1: \delta < A$$

Non-inferiority and *superiority* tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
μ_T	Not used	*Treatment mean*. This is the treatment mean.
μ_R	Not used	*Reference mean*. This is the mean of a reference population.
M_S	SM	*Margin of superiority.* This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.
δ	D	*True difference*. This is the value of $\mu_T - \mu_R$, the difference between the treatment and reference means. This is the value at which the power is calculated.

Note that the actual values of μ_T and μ_R are not needed. Only their difference is needed for power and sample size calculations.

Superiority by a Margin Tests

A *superiority test* tests that the treatment mean is better than the reference mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The value of δ must be greater than $|M_S|$. The following are equivalent sets of hypotheses.

$H_0: \mu_1 \leq \mu_2 + |M_S|$ versus $H_1: \mu_1 > \mu_2 + |M_S|$

$H_0: \mu_1 - \mu_2 \leq |M_S|$ versus $H_1: \mu_1 - \mu_2 > |M_S|$

$H_0: \delta \leq |M_S|$ versus $H_1: \delta > |M_S|$

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The value of δ must be less than $-|M_S|$. The following are equivalent sets of hypotheses.

$H_0: \mu_1 \geq \mu_2 - |M_S|$ versus $H_1: \mu_1 < \mu_2 - |M_S|$

$H_0: \mu_1 - \mu_2 \geq -|M_S|$ versus $H_1: \mu_1 - \mu_2 < -|M_S|$

$H_0: \delta \geq -|M_S|$ versus $H_1: \delta < -|M_S|$

Test Statistics

The analysis for assessing non-inferiority and non-zero null tests 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. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

$$Power = T_V\left(\left(\frac{\delta - \varepsilon}{\sigma_W\sqrt{b\,/\,n}}\right) - t_{V,1-\alpha}\right)$$

where *T* represents the cumulative *t* distribution, *V* and *b* depend on the design, σ_W is the square root of the within mean square error from the ANOVA table used to analyze the cross-over design, and *n* is the average number of subjects per sequence. Note that the constants *V* and *b* depend on the design as follows.

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

$$Power = 1 - T_V\left(t_{V,1-\alpha} - \left(\frac{\varepsilon - \delta}{\sigma_W\sqrt{b\,/\,n}}\right)\right)$$

The constants *V* and *b* depend on the design as follows:

Design Type	**Parameters (*V*,*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

Researchers want to calculate the power of a non-zero null test using data from a two-sequence, dual cross-over design. The margin of superiority is either 5 or 10 at several sample sizes between 6 and 66. The true difference between the means under is assumed to be 15. Similar experiments have had a standard deviation (σw) of 10. The significance level is 0.025.

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	**Power**
Design Type	**3x2 (Three-Period, Two-Sequence Dual: ABB|BAA)**
Higher Means Are	**Better**
Alpha	**0.025**
N (Total Sample Size)	**6 to 66 by 10**
SM (Superiority Margin)	**5 10**
D (True Difference)	**15**
Specify σ as σw or σb and ρ	**σw (Within Std Dev)**
σw (Within Std Dev)	**10**

Output

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

Numeric Reports

Numeric Results for a Superiority Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design

Solve For:	Power
Treatment Sequences:	ABB | BAA
Higher Means Are:	Better
Hypotheses:	H0: μT - μR ≤ SM vs. H1: μT - μR > SM

Power	Total Sample Size N	Superiority Margin SM	Mean Difference D (μT - μR)	Within Standard Deviation σw	Alpha
0.38371	6	5	15	10	0.025
0.11393	6	10	15	10	0.025
0.88323	16	5	15	10	0.025
0.34050	16	10	15	10	0.025
0.98180	26	5	15	10	0.025
0.52817	26	10	15	10	0.025
0.99751	36	5	15	10	0.025
0.67437	36	10	15	10	0.025
0.99969	46	5	15	10	0.025
0.78172	46	10	15	10	0.025
0.99996	56	5	15	10	0.025
0.85714	56	10	15	10	0.025
1.00000	66	5	15	10	0.025
0.90836	66	10	15	10	0.025

Power	The probability of rejecting H0 (concluding non-inferiority) when H0 is false.
N	The total number of subjects. They are divided evenly among all sequences.
μT	The treatment mean. It is usually associated with the letter "A" in the design.
μR	The reference mean. It is usually associated with the letter "B" in the design.
SM	The magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference mean that is required to be considered superior.
D (μT - μR)	The mean difference at which the power is computed. D = treatment mean - reference mean.
σw	The square root of the within mean square error from the ANOVA table.
Alpha	The probability of falsely rejecting H0 (falsely concluding superiority).

Summary Statements

A three-period, two-sequence dual cross-over (ABB | BAA) design (where higher means are considered to be better) will be used to test whether the treatment mean (μT) is superior to the reference mean (μR) by a margin, by testing whether the difference in means (μT - μR) is greater than the superiority margin 5 (H0: μT - μR ≤ 5 versus H1: μT - μR > 5). The comparison will be made using a one-sided t-test, with a Type I error rate (α) of 0.025. The within-subject standard deviation is assumed to be 10. To detect a difference in means (μT - μR) of 15, with a total sample size of 6 (allocated equally to the 2 sequences), the power is 0.38371.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	6	8	2
20%	16	20	4
20%	26	33	7
20%	36	45	9
20%	46	58	12
20%	56	70	14
20%	66	83	17

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 Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 8 subjects should be enrolled to obtain a final sample size of 6 subjects.

References

Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York

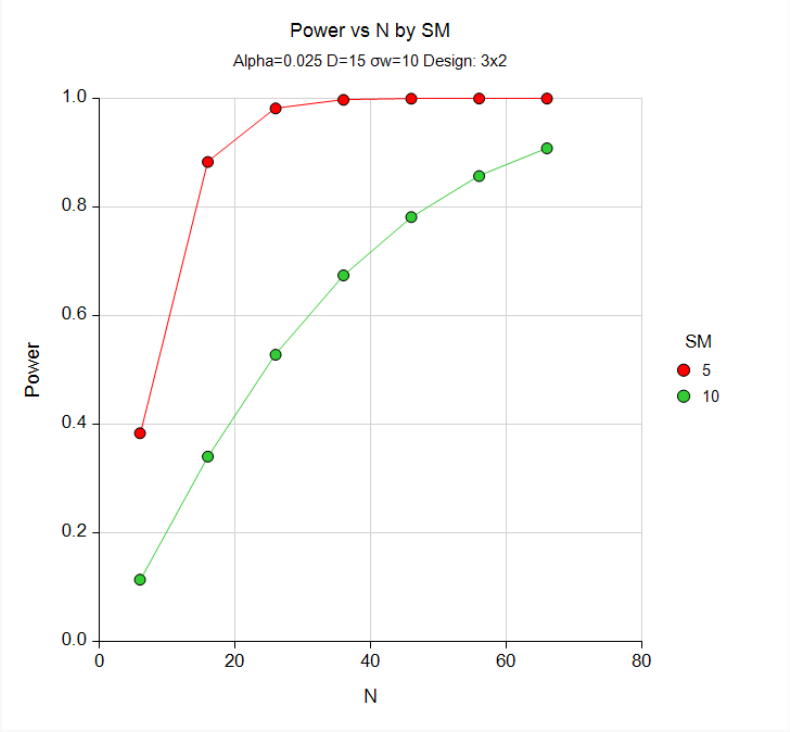
Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

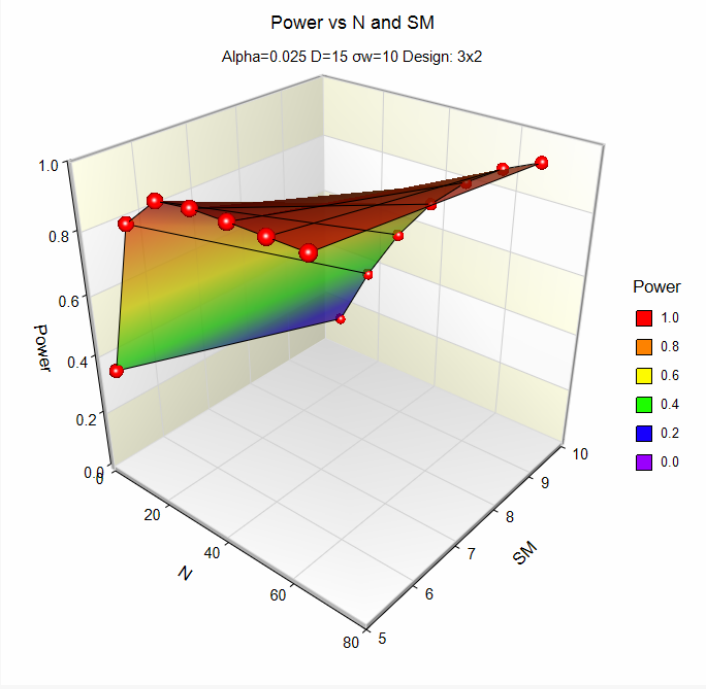
Chen, K.W., Chow, S.C., and Li, G. 1997. 'A Note on Sample Size Determination for Bioequivalence Studies with Higher-Order Crossover Designs.' Journal of Pharmacokinetics and Biopharmaceutics, Volume 25, No. 6, pages 753-765.

This report shows the power for the indicated scenarios.

Plots Section

Plots





These plots show 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)**
Higher Means Are	**Better**
Power	**0.80 0.90**
Alpha	**0.025**
SM (Superiority Margin)	**5 10**
D (True Difference)	**15**
Specify σ as σw or σb and ρ	**σw (Within Std Dev)**
σw (Within Std Dev)	**10**

Output

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

Numeric Results for a Superiority Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design

Solve For:	Sample Size (Exact)
Treatment Sequences:	ABB | BAA
Higher Means Are:	Better
Hypotheses:	H0: μT - μR ≤ SM vs. H1: μT - μR > SM

Power	Total Sample Size N	Superiority Margin SM	Mean Difference D (μT - μR)	Within Standard Deviation σw	Alpha
0.80317	13	5	15	10	0.025
0.80734	49	10	15	10	0.025
0.90229	17	5	15	10	0.025
0.90412	65	10	15	10	0.025

When the superiority margin is set to 10, 65 subjects are needed to achieve 90% power and 49 subjects are needed to achieve at least 80% power.

Example 3 – Validation

This procedure uses the same mechanics as the *Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design* procedure. We refer the user to Example 3 of Chapter 530 for the validation.