

Chapter 172

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

Introduction

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments, and the objective 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 the other. It is assumed that there is a *washout* period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a *carry-over* effect.

A 2×2 cross-over design contains two *sequences* (treatment orderings) and two time periods (occasions). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do not carry over to the second. Thus, the groups in this design are defined by the sequence in which the drugs are administered, not by the treatments they receive. Indeed, higher-order cross-over designs have been used in which the same treatment is used on both occasions.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

The sample size calculations in the procedure are based on the formulas presented in Lui (2016).

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 into the study may be easier because each patient will receive both treatments. Finally, it is often more difficult to obtain a subject than to obtain a measurement.

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. It may be difficult to separate the treatment effect from the period effect, the carry-over effect of the previous treatment, and the interaction between period and treatment.

The design cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it should not be used to compare treatments that are intended to provide a cure.

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

Because subjects must be measured at least twice, it is often more difficult to keep patients enrolled in the study. It is arguably simpler to measure a subject once than to obtain their measurement twice. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

Technical Details

The 2x2 crossover design may be described as follows. Randomly assign the subjects to one of two sequence groups so that there are n_1 subjects in sequence one and n_2 subjects in sequence two. In order to achieve design balance, the sample sizes n_1 and n_2 are assumed to be equal so that $n_1 = n_2 = n = N/2$.

Sequence one is given the control (A) followed by the treatment (B). Sequence two is given the treatment (B) followed by the control (A).

The design can be analyzed using a simple z-test if we ignore period and sequence effects or using a more complex random effects logistic regression model that adjusts for period and sequence effects. The sample size calculations herein ignore period and sequence effects. Julious (2010) suggests on page 175 that the bias due to ignoring period effects if a period-adjusted analysis is planned is not great and that sample size calculations that ignore period effects are adequate.

Cross-Over Design

The discussions that follow summarize the results in Lui (2016). Consider a 2x2 cross-over design and let $x_{ij}^{(g)}$ represent the binary response (0 or 1) from the j^{th} subject, $j = 1, \dots, n_g$, in the i^{th} period ($i = 1, 2$), in sequence g ($g = 1, 2$). Let $n_{rc}^{(g)}$ represent the number of subjects among n_g subjects in sequence g with the response vector ($x_{1j} = r, x_{2j} = c$). We can then summarize the results in terms of counts from a cross-over design with the following table for sequences 1 and 2 as

SEQUENCE 1 (Control → Treatment)

		Period 2 (Treatment)		
		Yes	No	Total
Period 1 (Control)	Yes	$n_{11}^{(1)}$	$n_{10}^{(1)}$	$n_{1\cdot}^{(1)}$
	No	$n_{01}^{(1)}$	$n_{00}^{(1)}$	$n_{0\cdot}^{(1)}$
Total		$n_{\cdot 1}^{(1)}$	$n_{\cdot 0}^{(1)}$	n_1

SEQUENCE 2 (Treatment → Control)

		Period 2 (Control)		
		Yes	No	Total
Period 1 (Treatment)	Yes	$n_{11}^{(2)}$	$n_{10}^{(2)}$	$n_{1\cdot}^{(2)}$
	No	$n_{01}^{(2)}$	$n_{00}^{(2)}$	$n_{0\cdot}^{(2)}$
Total		$n_{\cdot 1}^{(2)}$	$n_{\cdot 0}^{(2)}$	n_2

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

In terms of proportions, the 2×2 cross-over design tables can be summarized as

SEQUENCE 1 (Control → Treatment)					SEQUENCE 2 (Treatment → Control)				
		Period 2 (Treatment)					Period 2 (Control)		
		Yes	No	Total			Yes	No	Total
Period 1 (Control)	Yes	$P_{11}^{(1)}$	$P_{10}^{(1)}$	$P_{1\cdot}^{(1)}$	Period 1 (Treatment)	Yes	$P_{11}^{(2)}$	$P_{10}^{(2)}$	$P_{1\cdot}^{(2)}$
	No	$P_{01}^{(1)}$	$P_{00}^{(1)}$	$P_{0\cdot}^{(1)}$		No	$P_{01}^{(2)}$	$P_{00}^{(2)}$	$P_{0\cdot}^{(2)}$
Total		$P_{\cdot 1}^{(1)}$	$P_{\cdot 0}^{(1)}$	1	Total		$P_{\cdot 1}^{(2)}$	$P_{\cdot 0}^{(2)}$	1

with the individual proportions estimated as

$$\hat{P}_{rc}^{(g)} = \frac{n_{rc}^{(g)}}{n_g}.$$

Lui (2016) indicates on pages 32-42 that the odds ratio for the treatment versus the control (O_T/O_C) is defined for a 2×2 cross-over design as

$$OR = \sqrt{\frac{P_{01}^{(1)} P_{10}^{(2)}}{P_{10}^{(1)} P_{01}^{(2)}}}$$

with estimate

$$\widehat{OR} = \sqrt{\frac{\hat{P}_{01}^{(1)} \hat{P}_{10}^{(2)}}{\hat{P}_{10}^{(1)} \hat{P}_{01}^{(2)}}}.$$

The estimated log odds ratio, $\log(\widehat{OR})$, has asymptotic variance σ^2/n with

$$\sigma^2 = \frac{1}{4} \left(\frac{1}{P_{01}^{(1)}} + \frac{1}{P_{10}^{(1)}} + \frac{1}{P_{01}^{(2)}} + \frac{1}{P_{10}^{(2)}} \right)$$

which can be estimated as

$$\hat{\sigma}^2 = \frac{1}{4} \left(\frac{1}{\hat{P}_{01}^{(1)}} + \frac{1}{\hat{P}_{10}^{(1)}} + \frac{1}{\hat{P}_{01}^{(2)}} + \frac{1}{\hat{P}_{10}^{(2)}} \right).$$

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

The standard deviation, then, is

$$SD = \sigma = \sqrt{\sigma^2}$$

with estimate

$$\widehat{SD} = \hat{\sigma} = \sqrt{\hat{\sigma}^2}.$$

Superiority by a Margin Test Statistics

Higher Proportions Better

When higher proportions are better, the null and alternative hypotheses for a one-sided superiority test are

$$H_0: OR \leq OR_0 \text{ vs. } H_A: OR > OR_0$$

where OR_0 is the superiority bound (i.e., the smallest odds ratio (Ot/Oc) for which the treatment will be considered superior to the standard or control). When higher proportions are better, OR_0 should be greater than one.

The power and sample size calculations are based on the test statistic

$$Z = \frac{\log(\widehat{OR}) - \log(OR_0)}{\frac{\widehat{SD}}{\sqrt{n}}}$$

which is asymptotically distributed as standard normal under the null hypothesis. The null hypothesis is rejected in favor of the alternative at level α if

$$\frac{\log(\widehat{OR}) - \log(OR_0)}{\frac{\widehat{SD}}{\sqrt{n}}} > Z_{1-\alpha}$$

where $Z_{1-\alpha}$ is the upper $1 - \alpha$ percentile of the standard normal distribution.

Higher Proportions Worse

When higher proportions are worse, the null and alternative hypotheses for a one-sided superiority test are

$$H_0: OR \geq OR_0 \text{ vs. } H_A: OR < OR_0$$

where OR_0 is the superiority bound (i.e., the largest odds ratio (O_t/O_c) for which the treatment will be considered superior to the standard or control). When higher proportions are worse, OR_0 should be less than one.

The power and sample size calculations are based on the test statistic

$$Z = \frac{\log(\widehat{OR}) - \log(OR_0)}{\frac{SD}{\sqrt{n}}}$$

which is asymptotically distributed as standard normal under the null hypothesis. The null hypothesis is rejected in favor of the alternative at level α if

$$\frac{\log(\widehat{OR}) - \log(OR_0)}{\frac{SD}{\sqrt{n}}} < Z_\alpha$$

where Z_α is the lower α percentile of the standard normal distribution.

Superiority by a Margin Power Calculations

Higher Proportions Better

Derived from the sample size formula given in Lui (2016) on pages 42 and 43, the power for the one-sided superiority test of $H_0: OR \leq OR_0$ versus $H_A: OR > OR_0$ is

$$\Phi \left(\frac{\log(OR_1) - \log(OR_0)}{\frac{SD}{\sqrt{n}}} - Z_{1-\alpha} \right)$$

where $\Phi()$ is the standard normal distribution function, OR_1 is the actual value of the odds ratio under the alternative hypothesis, and $Z_{1-\alpha}$ is the upper $1 - \alpha$ percentile of the standard normal distribution. The sample size calculation formula is

$$n = \text{Ceiling} \left\{ \left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{\log(OR_1) - \log(OR_0)} \right)^2 \right\}$$

Higher Proportions Worse

Derived from the sample size formula given in Lui (2016) on pages 42 and 43, the power for the one-sided superiority test of $H_0: OR \geq OR_0$ versus $H_A: OR < OR_0$ is

$$\Phi\left(\frac{\log(OR_0) - \log(OR_1)}{\frac{SD}{\sqrt{n}}} + Z_\alpha\right)$$

where $\Phi()$ is the standard normal distribution function, OR_1 is the actual value of the odds ratio under the alternative hypothesis, and Z_α is the lower α percentile of the standard normal distribution. The sample size calculation formula is

$$n = \text{Ceiling}\left\{\left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{\log(OR_0) - \log(OR_1)}\right)^2\right\}.$$

Example 1 – Power Analysis

Suppose you want to consider the power of a superiority by a margin test of the hypotheses $H_0: OR \leq 1.2214$ versus $H_A: OR > 1.2214$ in a balanced cross-over design with a binary endpoint where the test is computed based on the odds ratio for sample sizes between 75 and 175. Let's assume that the actual odds ratio is 2 and the estimated standard deviation of the log odds ratio is 2.5. The significance level is 0.05.

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**
 Higher Proportions Are **Better**
 Alpha..... **0.05**
 n (Sample Size per Sequence) **75 to 175 by 25**
 OR0 (Superiority Odds Ratio) **1.2214**
 OR1 (Actual Odds Ratio) **2**
 Estimation Method **Enter SD Directly**
 Standard Deviation (SD)..... **2.5**

Output

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

Numeric Results

Solve For: **Power**
 Higher Proportions Are: **Better**
 Hypotheses: **H0: $OR \leq OR_0$ vs. H1: $OR > OR_0$**

Power	Sample Size		Odds Ratio		Standard Deviation SD	Alpha
	Sequence n	Total N	Superiority OR0	Actual OR1		
0.52530	75	150	1.221	2	2.5	0.05
0.62845	100	200	1.221	2	2.5	0.05
0.71246	125	250	1.221	2	2.5	0.05
0.77967	150	300	1.221	2	2.5	0.05
0.83264	175	350	1.221	2	2.5	0.05

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
 n The sample size in each sequence (or group).
 N The total combined sample size from both sequences.
 OR0 The superiority odds ratio used to specify the hypothesis test.
 OR1 The actual odds ratio at which power is calculated.
 SD The user-entered standard deviation. This is estimated from a previous study.
 Alpha The probability of rejecting a true null hypothesis.

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

Summary Statements

A 2x2 cross-over design will be used to test whether the treatment proportion is superior to the standard proportion by a margin, with a superiority odds ratio ($OR = O_t / O_c$) of 1.221 ($H_0: OR \leq 1.221$ versus $H_1: OR > 1.221$). The comparison will be made using a one-sided log odds ratio Z-test, with a Type I error rate (α) of 0.05. The standard deviation of the log odds ratio is assumed to be 2.5. To detect an odds ratio of 2 with a sample size of 75 in each sequence (totaling 150 subjects), the power is 0.5253.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size		Dropout-Inflated Enrollment Sample Size		Expected Number of Dropouts	
	n	N	n'	N'	d	D
20%	75	150	94	188	19	38
20%	100	200	125	250	25	50
20%	125	250	157	314	32	64
20%	150	300	188	376	38	76
20%	175	350	219	438	44	88

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 and N	The evaluable group and total sample sizes, respectively, at which power is computed (as entered by the user). If n subjects from each group are evaluated out of the n' subjects that are enrolled in the study, the design will achieve the stated power. $N = 2n$.
n' and N'	The number of subjects that should be enrolled in the study in order to obtain n and 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.). $N' = 2n'$.
d and D	The expected number of group and total dropouts, respectively. $d = n' - n$ and $D = 2d$.

Dropout Summary Statements

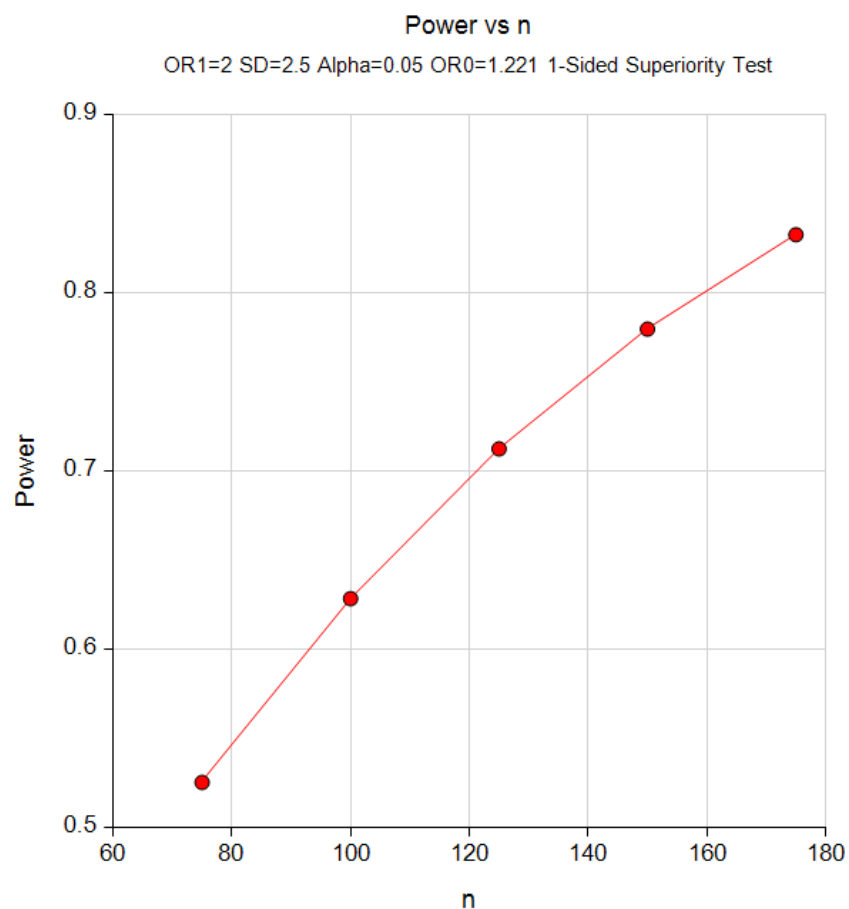
Anticipating a 20% dropout rate, 94 subjects should be enrolled in each group to obtain final sample sizes of 75 subjects per group.

References

Lui, Kung-Jong. 2016. Crossover Designs: Testing, Estimation, and Sample Size. John Wiley & Sons Ltd. Chichester, West Sussex, England.

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

Plots



This report shows the values of each of the parameters, one scenario per row. This plot shows the relationship between sample size and power. We see that a sample size of about 160 per sequence is needed detect an odds ratio of 2 with 80% power.

Example 2 – Calculating Sample Size when Estimating the Standard Deviation from a Previous Study (Validation using Hand Calculations)

This example demonstrates how to calculate the sample size when estimating the standard deviation of the log odds ratio from data in a previous study using the method in Lui (2016) on page 42. In this example we'll find the sample size required to detect an odds ratio of 2 with 80% power at a significance level of 0.05 in a superiority by a margin test against a lower odds ratio bound of 1.2214. The SD is estimated from discordant proportions in a previous study.

Table 3.2 of Lui (2016) on page 36 presents the following results from 279 subjects a simple 2x2 cross-over trial comparing two inhalation devices, A and B.

SEQUENCE 1 (Control (A) → Treatment (B))					SEQUENCE 2 (Treatment (B) → Control (A))				
		Period 2 (B)					Period 2 (A)		
		Yes	No	Total			Yes	No	Total
Period 1 (A)	Yes	26	41	67	Period 1 (B)	Yes	38	16	54
	No	15	57	72		No	32	54	86
Total		41	98	139	Total		70	70	140

The discordant proportions are estimated as

$$\hat{p}_{01}^{(1)} = \frac{15}{139} = 0.1079$$

$$\hat{p}_{10}^{(1)} = \frac{41}{139} = 0.2950$$

$$\hat{p}_{01}^{(2)} = \frac{32}{140} = 0.2286$$

$$\hat{p}_{10}^{(2)} = \frac{16}{140} = 0.1143$$

Superiority by a Margin Tests for the Odds Ratio of Two Proportions in a 2x2 Cross-Over Design

Using the method of Lui (2016) on page 42, SD is estimated as

$$\begin{aligned}
 \widehat{SD} &= \sqrt{\widehat{\sigma}^2} \\
 &= \sqrt{\frac{1}{4} \left(\frac{1}{\widehat{p}_{01}^{(1)}} + \frac{1}{\widehat{p}_{10}^{(1)}} + \frac{1}{\widehat{p}_{01}^{(2)}} + \frac{1}{\widehat{p}_{10}^{(2)}} \right)} \\
 &= \sqrt{\frac{1}{4} \left(\frac{1}{0.1079} + \frac{1}{0.2950} + \frac{1}{0.2286} + \frac{1}{0.1143} \right)} \\
 &= \mathbf{2.5388}
 \end{aligned}$$

PASS will calculate this SD value for you automatically when you input the discordant cell proportions.

Since there is no example given for this calculation in the book, we'll validate this procedure using hand calculations. The sample size calculated by hand using the equation on Lui (2016) page 43 is

$$\begin{aligned}
 n &= \text{Ceiling} \left\{ \left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{\log(OR_1) - \log(OR_0)} \right)^2 \right\} = \text{Ceiling} \left\{ \left(\frac{(1.64 + 0.84)2.5388}{\log(2) - \log(1.2214)} \right)^2 \right\} \\
 &= \text{Ceiling}\{163.0059\} \\
 &= 164
 \end{aligned}$$

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
Higher Proportions Are	Better
Power.....	0.80
Alpha.....	0.05
OR0 (Superiority Odds Ratio)	1.2214
OR1 (Actual Odds Ratio)	2
Estimation Method	Use Estimated Discordant Cell Proportions
p01(1)	0.1079
p10(1)	0.2950
p01(2)	0.2286
p10(2)	0.1143

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Higher Proportions Are: Better
 Hypotheses: $H_0: OR \leq OR_0$ vs. $H_1: OR > OR_0$

Power	Sample Size		Odds Ratio		Standard Deviation SD*	Alpha
	Sequence n	Total N	Superiority OR0	Actual OR1		
0.80031	164	328	1.221	2	2.539	0.05

* SD Estimated using Previously Estimated Discordant Cell Proportions:
 $p_{01}(1) = 0.1079$, $p_{10}(1) = 0.295$, $p_{01}(2) = 0.2286$, $p_{10}(2) = 0.1143$

This report indicates that the estimated standard deviation using the method of Lui (2016) is 2.539 and the required sample size is 164 per sequence. The discordant cell proportions are also listed. These values for SD and n computed by **PASS** match our hand calculations above.