Chapter 179

Non-Inferiority Tests for the Generalized Odds Ratio for Ordinal Data 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.

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 2×2 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).

Cross-Over Design

The discussions that follow summarize the results in Lui (2016) on pages 57-70. Consider a 2×2 cross-over design and let $Y_{ij}^{(g)}$ represent the response category for the j^{th} subject, j = 1, ..., n_g, in the i^{th} period (i = 1, 2), in sequence g (g = 1, 2) on an ordinal scale with L possible categories, $C_1 < C_2 < C_3 < \cdots < C_L$. Assume that subjects in sequence 1 are given the control or standard in period 1 and then the experimental treatment in period 2. Assume that subjects in sequence 2 are given the experimental treatment in period 1 and then the control or standard in period 2. If we define η as the relative effect of the treatment to the control, then $\eta > 0$ indicates that the response category, C_k, of a subject receiving the treatment tends to be greater than the response category of a subject receiving the control or standard. Similarly, $\eta < 0$ indicates that the response category, C_k, of a subject receiving the treatment tends to be less than the response category of a subject receiving the control or standard.

For a randomly selected subject from sequence g (g = 1, 2), the probability that the measured response in period 1, $Y_{1j}^{(g)}$, is less than the measured response in period 2, $Y_{2j}^{(g)}$, is

$$\Pi_C^{(g)} = P\left(Y_{1j}^{(g)} < Y_{2j}^{(g)}\right) = \sum_{r=1}^{L-1}\sum_{s=r+1}^{L} \pi_{rs}^{(g)}$$

where

$$\pi_{rs}^{(g)} = P\left(Y_{1j}^{(g)} = C_r, Y_{2j}^{(g)} = C_s\right) \text{ with } r = 1, 2, \ldots, L \text{ and } s = 1, 2, \ldots, L.$$

Similarly, the probability that the measured response in period 1, $Y_{1j}^{(g)}$, is greater than the measured response in period 2, $Y_{2j}^{(g)}$, is

$$\Pi_D^{(g)} = P\left(Y_{1j}^{(g)} > Y_{2j}^{(g)}\right) = \sum_{r=2}^{L}\sum_{s=1}^{r-1} \pi_{rs}^{(g)}.$$

The generalized odds ratio, $GOR^{(g)}$, of subject responses between periods 1 and 2 for a randomly selected subject from group g is defined as

$$GOR^{(g)} = \frac{\Pi_C^{(g)}}{\Pi_D^{(g)}}.$$

The generalized odds ratio, GOR, for the treatment versus the control is then

$$GOR = e^{\eta} = \sqrt{\frac{GOR^{(1)}}{GOR^{(2)}}}.$$

Non-Inferiority Test Statistics

Higher Values Better

When higher values are better, the null and alternative hypotheses for a one-sided non-inferiority test are

$$H_0: GOR \leq GOR_0 \quad \text{vs.} \quad H_A: GOR > GOR_0$$

where GOR_0 is the lower non-inferiority bound (i.e., the smallest generalized odds ratio for which the treatment will still be considered non-inferior to the standard or control). When higher values are better, GOR_0 should be less than one.

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

$$Z = \frac{\log\left(\widehat{GOR}\right) - \log(GOR_0)}{\sqrt{\widehat{Var}\left(\log\left(\widehat{GOR}\right)\right)}}$$

which is asymptotically distributed as standard normal under the null hypothesis with

$$\widehat{Var}\left(\log\left(\widehat{GOR}\right)\right) = \frac{1}{4}\sum_{g=1}^{2}\frac{\widehat{\Pi}_C^{(g)} + \widehat{\Pi}_D^{(g)}}{n_g \widehat{\Pi}_C^{(g)} \widehat{\Pi}_D^{(g)}}.$$

The null hypothesis is rejected in favor of the alternative at level α if

$$Z > Z_{1-\alpha}$$

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

Higher Values Worse

When higher values are worse, the null and alternative hypotheses for a one-sided non-inferiority test are

$$H_0: GOR \geq GOR_0 \quad \text{vs.} \quad H_A: GOR < GOR_0$$

where GOR_0 is the upper non-inferiority bound (i.e., the largest generalized odds ratio for which the treatment will still be considered non-inferior to the standard or control). When higher values are worse, GOR_0 should be greater than one.

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

$$Z = \frac{\log\left(\widehat{GOR}\right) - \log(GOR_0)}{\sqrt{\widehat{Var}\left(\log\left(\widehat{GOR}\right)\right)}}$$

which is asymptotically distributed as standard normal under the null hypothesis with

$$\widehat{Var}\left(\log\left(\widehat{GOR}\right)\right) = \frac{1}{4}\sum_{g=1}^{2}\frac{\hat{\Pi}_C^{(g)} + \hat{\Pi}_D^{(g)}}{n_g \hat{\Pi}_C^{(g)} \hat{\Pi}_D^{(g)}}.$$

The null hypothesis is rejected in favor of the alternative at level α if

$$Z < Z_\alpha$$

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

Non-Inferiority Power Calculations

If $\widehat{GOR}$ is the estimate of the generalized odds ratio, then $\hat{\eta} = \log\left(\widehat{GOR}\right)$ has asymptotic variance

$$Var\left(\log\left(\widehat{GOR}\right)\right) = \frac{V\left(\log\left(\widehat{GOR}\right)\right)}{n},$$

where

$$V\left(\log\left(\widehat{GOR}\right)\right) = \frac{1}{4}\sum_{g=1}^{2}\frac{\Pi_C^{(g)} + \Pi_D^{(g)}}{\Pi_C^{(g)} \Pi_D^{(g)}}.$$

Also define the standard deviation of log(generalized odds ratio) as

$$SD = \sqrt{V\left(\log\left(\widehat{GOR}\right)\right)}$$

Higher Values Better

Derived from the sample size formula given in Lui (2016) on page 68, the power for the one-sided non-inferiority test of $H_0: GOR \leq GOR_0$ versus $H_A: GOR > GOR_0$ is

$$\Phi\left(\frac{\log(GOR_1) - \log(GOR_0)}{\frac{SD}{\sqrt{n}}} - Z_{1-\alpha}\right)$$

where $\Phi()$ is the standard normal distribution function, GOR_1 is the actual value of the generalized 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

$$n = \text{Ceiling}\left\{\left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{\log(GOR_1) - \log(GOR_0)}\right)^2\right\}.$$

Higher Values Worse

Derived from the sample size formula given in Lui (2016) on page 68, the power for the one-sided non-inferiority test of $H_0: GOR \geq GOR_0$ versus $H_A: GOR < GOR_0$ is

$$\Phi\left(\frac{\log(GOR_0) - \log(GOR_1)}{\frac{SD}{\sqrt{n}}} + Z_{\alpha}\right)$$

where $\Phi()$ is the standard normal distribution function, GOR_1 is the actual value of the generalized 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

$$n = \text{Ceiling}\left\{\left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{\log(GOR_0) - \log(GOR_1)}\right)^2\right\}.$$

Example 1 – Power Analysis

Suppose you want to consider the power of a non-inferiority test of the hypotheses $H_0: GOR \leq 0.8$ versus $H_A: GOR > 0.8$ in a balanced cross-over design with ordinal data where the test is computed based on the generalized odds ratio for sequence sample sizes between 25 and 125. Let's assume that the actual generalized odds ratio to detect is 2 and the estimated standard deviation of the log generalized 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 Values Are	**Better**
Alpha	**0.05**
n (Sample Size per Sequence)	**25 to 125 by 25**
GOR0 (Non-Inferiority Gen. Odds Ratio)	**0.8**
GOR1 (Actual Gen. 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: GOR ≤ GOR0 vs. H1: GOR > GOR0

	Sample Size		Generalized Odds Ratio			
Power	Sequence n	Total N	Non-Inferiority GOR0	Actual GOR1	Standard Deviation SD	Alpha
0.57445	25	50	0.8	2	2.5	0.05
0.82813	50	100	0.8	2	2.5	0.05
0.93690	75	150	0.8	2	2.5	0.05
0.97832	100	200	0.8	2	2.5	0.05
0.99291	125	250	0.8	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.
GOR0	The non-inferiority generalized odds ratio used to specify the hypothesis test.
GOR1	The actual generalized odds ratio at which power is calculated.
SD	The user-entered standard deviation of the log generalized odds ratio. This is estimated from a previous study.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A 2×2 cross-over design will be used to test whether the (ordinal) treatment response is non-inferior to the standard response, with a non-inferiority generalized odds ratio (GOR) of 0.8 (H0: GOR ≤ 0.8 versus H1: GOR > 0.8). The comparison will be made using a one-sided log generalized odds ratio Z-test, with a Type I error rate (α) of 0.05. The standard deviation of the log generalized odds ratio is assumed to be 2.5. To detect a generalized odds ratio of 2 with a sample size of 25 in each sequence (totaling 50 subjects), the power is 0.57445.

Dropout-Inflated Sample Size

	Sample Size		Dropout-Inflated Enrollment Sample Size		Expected Number of Dropouts	
Dropout Rate	n	N	n'	N'	d	D
20%	25	50	32	64	7	14
20%	50	100	63	126	13	26
20%	75	150	94	188	19	38
20%	100	200	125	250	25	50
20%	125	250	157	314	32	64

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.

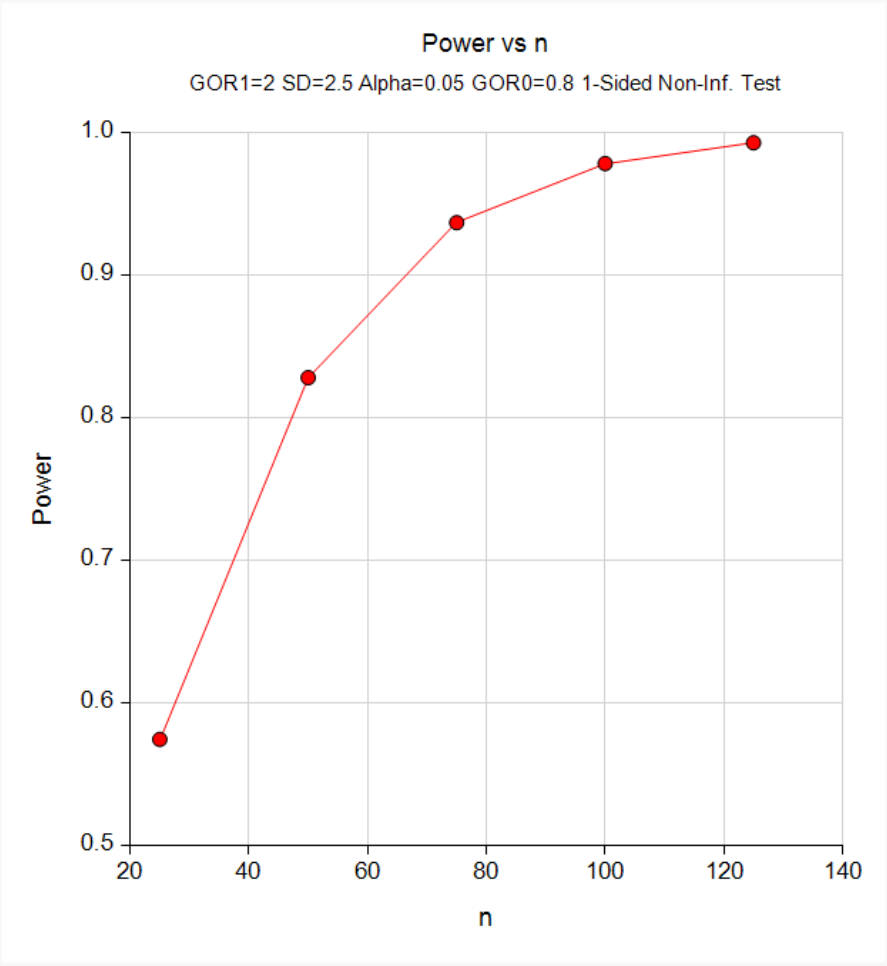
Dropout Summary Statements

Anticipating a 20% dropout rate, 32 subjects should be enrolled in each group to obtain final sample sizes of 25 subjects per group.

References

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

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 just under 50 per sequence is required to detect a generalized 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 generalized odds ratio from data in a previous study using the method in Lui (2016) on page 67. In this example we'll find the sample size required to detect a generalized odds ratio of 2 with 80% power at a significance level of 0.05 in a non-inferiority test against a lower generalized odds ratio bound of 0.8. The SD is estimated from cumulative proportions from a previous study.

Suppose the cumulative proportions are

$$\Pi_C^{(1)} = 0.11$$

$$\Pi_D^{(1)} = 0.29$$

$$\Pi_C^{(2)} = 0.23$$

$$\Pi_D^{(2)} = 0.11$$

Using the method of Lui (2016) on page 67, SD is calculated as

$$SD = \sqrt{\frac{1}{4}\sum_{g=1}^{2}\frac{\Pi_C^{(g)} + \Pi_D^{(g)}}{\Pi_C^{(g)}\Pi_D^{(g)}}}$$

$$= \sqrt{\frac{1}{4}\left(\frac{0.11 + 0.29}{0.11(0.29)} + \frac{0.23 + 0.11}{0.23(0.11)}\right)}$$

$$= \mathbf{2.5484}$$

PASS will calculate this SD value for you automatically when you input the cumulative 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 68 is

$$n = \text{Ceiling}\left\{\left(\frac{(Z_{1-\alpha} + Z_{1-\beta})SD}{log(GOR_1) - log(GOR_0)}\right)^2\right\} = \text{Ceiling}\left\{\left(\frac{(1.645 + 0.842)2.5484}{log(2) - log(0.8)}\right)^2\right\}$$

$$= \text{Ceiling}\{47.8240\}$$

$$= 48$$

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 Values Are	**Better**
Power	**0.80**
Alpha	**0.05**
GOR0 (Non-Inferiority Gen. Odds Ratio)	**0.8**
GOR1 (Actual Gen. Odds Ratio)	**2**
Estimation Method	**Use Estimated Cumulative Proportions**
∏C(1)	**0.11**
∏D(1)	**0.29**
∏C(2)	**0.23**
∏D(2)	**0.11**

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:	H0: GOR ≤ GOR0 vs. H1: GOR > GOR0

	Sample Size		Generalized Odds Ratio			
Power	Sequence n	Total N	Non-Inferiority GOR0	Actual GOR1	Standard Deviation SD*	Alpha
0.80128	48	96	0.8	2	2.548	0.05

* SD Estimated using Previously Estimated Cumulative Proportions:
∏C(1) = 0.11, ∏D(1) = 0.29, ∏C(2) = 0.23, ∏D(2) = 0.11

This report indicates that the estimated standard deviation using the method of Lui (2016) is 2.548 and the required sample size is 48 per sequence. The cumulative proportions are also listed. These values for SD and *n* computed by **PASS** match our hand calculations above.