

Chapter 181

Equivalence 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.

Equivalence Tests for the Generalized Odds Ratio for Ordinal Data 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).

Cross-Over Design

The discussions that follow summarize the results in Lui (2016) on pages 57-70. Consider a 2x2 cross-over design and let $Y_{ij}^{(g)}$ represent the response category for the j^{th} subject, $j = 1, \dots, n_g$, in the i^{th} period ($i = 1, 2$), in sequence g ($g = 1, 2$) on an ordinal scale with L possible categories, $C_1 < C_2 < C_3 < \dots < C_L$. 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

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

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

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

Equivalence Tests for the Generalized Odds Ratio for Ordinal Data in a 2x2 Cross-Over Design

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

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

Equivalence Test Statistics

The null and alternative hypotheses for an equivalence test are

$$H_0: GOR \leq GOR_{0L} \text{ or } GOR \geq GOR_{0U} \quad \text{vs.} \quad H_A: GOR_{0L} < GOR < GOR_{0U}$$

where GOR_{0L} and GOR_{0U} are the lower and upper equivalence bounds, respectively (i.e., the smallest and largest generalized odds ratios for which the treatment and control will be considered equivalent).

The power and sample size calculations are based on the two one-sided test (TOST) statistics

$$Z_L = \frac{\log(\widehat{GOR}) - \log(GOR_{0L})}{\sqrt{\widehat{Var}(\log(\widehat{GOR}))}} \quad \text{and} \quad Z_U = \frac{\log(\widehat{GOR}) - \log(GOR_{0U})}{\sqrt{\widehat{Var}(\log(\widehat{GOR}))}}$$

which are each asymptotically distributed as standard normal under the null hypothesis, with

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

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

$$Z_L > Z_{1-\alpha} \quad \text{and} \quad Z_U < Z_\alpha$$

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

Equivalence Power Calculation

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

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

where

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

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

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

On page 68 of Lui (2016), the power for an equivalence test of $H_0: GOR \leq GOR_{0L}$ or $GOR \geq GOR_{0U}$ versus $H_A: GOR_{0L} < GOR < GOR_{0U}$ is given as

$$\Phi\left(\frac{\log(GOR_{0U}) - \log(GOR_1)}{\frac{SD}{\sqrt{n}}} - Z_{1-\alpha}\right) - \Phi\left(\frac{\log(GOR_{0L}) - \log(GOR_1)}{\frac{SD}{\sqrt{n}}} + Z_{1-\alpha}\right)$$

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 is determined using a binary search of possible values for n .

Example 1 – Power Analysis

Suppose you want to consider the power of an equivalence test of the hypotheses $H_0: OR \leq 0.667$ or $OR \geq 1.5$ versus $H_A: 0.667 < OR < 1.5$ 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 100 and 300. The equivalence bounds in this example are log-scale symmetric since $0.667 = 1/1.5$. Let's assume that the actual odds ratio is 1 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**
 Alpha..... **0.05**
 n (Sample Size per Sequence) **100 to 300 by 50**
 GOR0.U (Upper Equiv. Gen. Odds Ratio) **1.5**
 GOR0.L (Lower Equiv. Gen. Odds Ratio) **1/GOR0.U**
 GOR1 (Actual Gen. Odds Ratio)..... **1**
 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**
 Hypotheses: $H_0: GOR \leq GOR0.L$ or $GOR \geq GOR0.U$ vs. $H_1: GOR0.L < GOR < GOR0.U$

Power	Sample Size		Generalized Odds Ratio			Standard Deviation SD	Alpha
	Sequence n	Total N	Equivalence Limits				
			Lower GOR0.L	Upper GOR0.U	Actual GOR1		
0.00000	100	200	0.667	1.5	1	2.5	0.05
0.26728	150	300	0.667	1.5	1	2.5	0.05
0.48353	200	400	0.667	1.5	1	2.5	0.05
0.64218	250	500	0.667	1.5	1	2.5	0.05
0.75569	300	600	0.667	1.5	1	2.5	0.05

Equivalence Tests for the Generalized Odds Ratio for Ordinal Data in a 2x2 Cross-Over Design

- 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.L The lower equivalence generalized odds ratio used to specify the hypothesis test.
- GOR0.U The upper equivalence 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 2x2 cross-over design will be used to test whether the (ordinal) treatment response is equivalent to the standard response, with equivalence generalized odds ratio (GOR) bounds of 0.667 and 1.5 ($H_0: GOR \leq 0.667$ or $GOR \geq 1.5$ versus $H_1: 0.667 < GOR < 1.5$). The comparison will be made using two one-sided log generalized odds ratio Z-tests, with an overall 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 1 with a sample size of 100 in each sequence (totaling 200 subjects), the power is 0.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size		Dropout-Inflated Enrollment Sample Size		Expected Number of Dropouts	
	n	N	n'	N'	d	D
20%	100	200	125	250	25	50
20%	150	300	188	376	38	76
20%	200	400	250	500	50	100
20%	250	500	313	626	63	126
20%	300	600	375	750	75	150

- 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 Lohknygina, 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

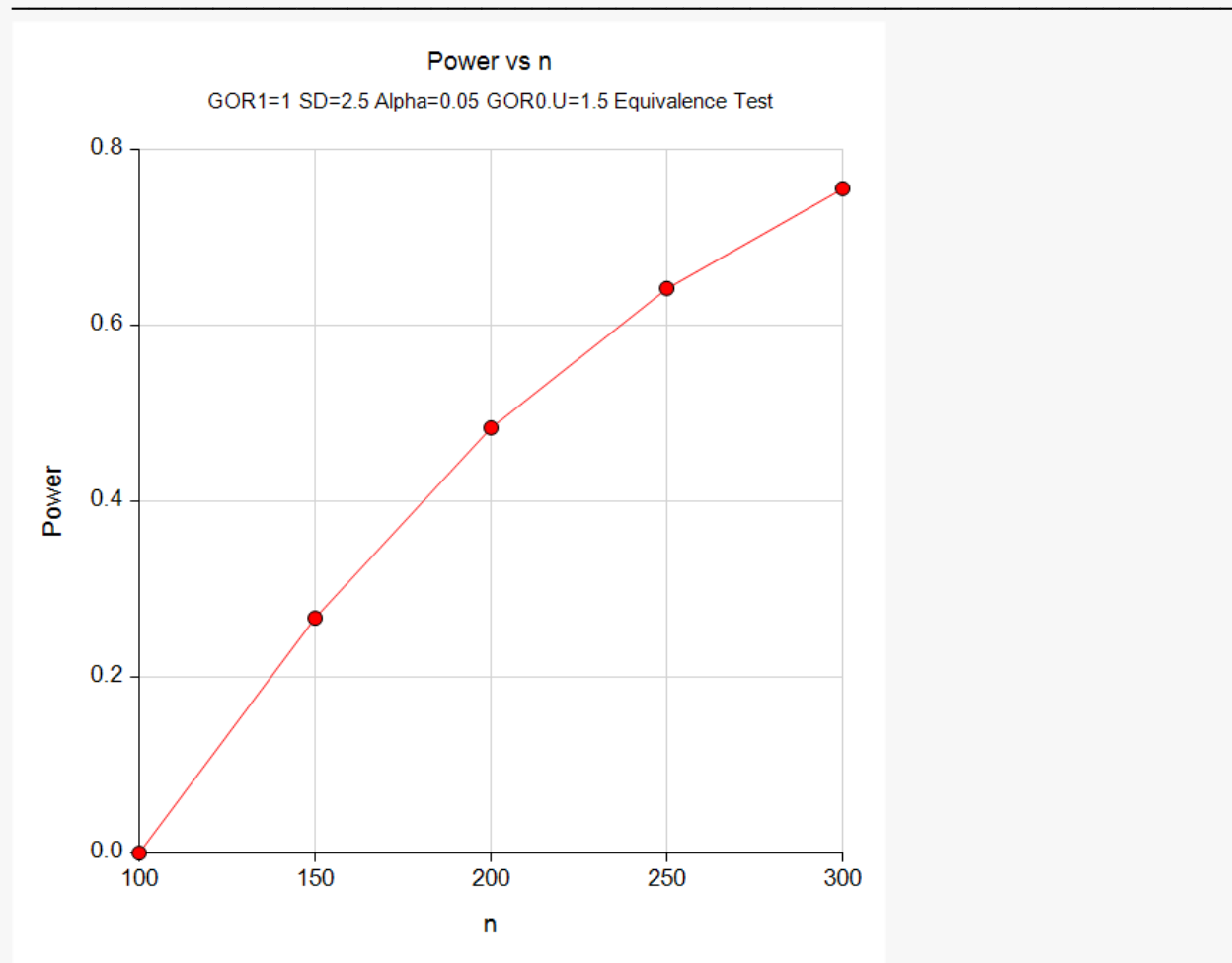
Anticipating a 20% dropout rate, 125 subjects should be enrolled in each group to obtain final sample sizes of 100 subjects per group.

References

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

Equivalence Tests for the Generalized Odds Ratio for Ordinal Data 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 greater than 300 per sequence is required for 80% power if the actual generalized odds ratio is 1.

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 an odds ratio of 1 with 80% power at a significance level of 0.05 in an equivalence test of the hypotheses $H_0: OR \leq 0.667$ or $OR \geq 1.5$ versus $H_A: 0.667 < OR < 1.5$. 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

$$\begin{aligned} SD &= \sqrt{\frac{1}{4} \sum_{g=1}^2 \frac{\Pi_C^{(g)} + \Pi_D^{(g)}}{\Pi_C^{(g)} \Pi_D^{(g)}}} \\ &= \sqrt{\frac{1}{4} \left(\frac{0.11 + 0.29}{0.11(0.29)} + \frac{0.23 + 0.11}{0.23(0.11)} \right)} \\ &= \mathbf{2.5484} \end{aligned}$$

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 power for per-sequence sample sizes of 338 and 339 calculated by hand using the equation on Lui (2016) page 68 is

$$\text{Power} = \Phi \left(\frac{\log(GOR_{0U}) - \log(GOR_1)}{\frac{SD}{\sqrt{n}}} - Z_{1-\alpha} \right) - \Phi \left(\frac{\log(GOR_{0L}) - \log(GOR_1)}{\frac{SD}{\sqrt{n}}} + Z_{1-\alpha} \right)$$

Equivalence Tests for the Generalized Odds Ratio for Ordinal Data in a 2x2 Cross-Over Design

$$\text{Power}_{(n=338)} = \Phi\left(\frac{\log(1.5) - \log(1)}{\frac{2.5484}{\sqrt{338}}} - 1.644854\right) - \Phi\left(\frac{\log\left(\frac{1}{1.5}\right) - \log(1)}{\frac{2.5484}{\sqrt{338}}} + 1.644854\right)$$

$$= 0.7995$$

$$\text{Power}_{(n=339)} = \Phi\left(\frac{\log(1.5) - \log(1)}{\frac{2.5484}{\sqrt{339}}} - 1.644854\right) - \Phi\left(\frac{\log\left(\frac{1}{1.5}\right) - \log(1)}{\frac{2.5484}{\sqrt{339}}} + 1.644854\right)$$

$$= 0.8011$$

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
Power.....	0.80
Alpha.....	0.05
GOR0.U (Upper Equiv. Gen. Odds Ratio)	1.5
GOR0.L (Lower Equiv. Gen. Odds Ratio)	1/GOR0.U
GOR1 (Actual Gen. Odds Ratio).....	1
Estimation Method	Use Estimated Cumulative Proportions
ΠC(1).....	0.11
ΠD(1).....	0.29
ΠC(2).....	0.23
ΠD(2).....	0.11

Output

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

Numeric Results

Solve For: [Sample Size](#)

Hypotheses: $H_0: GOR \leq GOR_{0.L} \text{ or } GOR \geq GOR_{0.U}$ vs. $H_1: GOR_{0.L} < GOR < GOR_{0.U}$

Power	Sample Size		Generalized Odds Ratio			Standard Deviation SD*	Alpha
	Sequence n	Total N	Equivalence Limits		Actual GOR1		
			Lower GOR0.L	Upper GOR0.U			
0.80105	339	678	0.667	1.5	1	2.548	0.05

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

This report indicates that the estimated standard deviation using the method of Lui (2016) is 2.548 and the required sample size is 339 per sequence with 80.105% power. The cumulative proportions are also listed. These values for power, SD, and n computed by **PASS** match our hand calculations above.