Kung-Jong Lui*

Testing Equality in Ordinal Data with Repeated Measurements: A Model-Free Approach

DOI 10.1515/ijb-2015-0075

Abstract: In randomized clinical trials, we often encounter ordinal categorical responses with repeated measurements. We propose a model-free approach with using the generalized odds ratio (GOR) to measure the relative treatment effect. We develop procedures for testing equality of treatment effects and derive interval estimators for the GOR. We further develop a simple procedure for testing the treatment-by-period interaction. To illustrate the use of test procedures and interval estimators developed here, we consider two real-life data sets, one studying the gender effect on pain scores on an ordinal scale after hip joint resurfacing surgeries, and the other investigating the effect of an active hypnotic drug in insomnia patients on ordinal categories of time to falling asleep.

Keywords: testing equality, interval estimators, generalized odds ratio, ordinal data, repeated measurements, model-free

1 Introduction

In clinical trials or health-related studies, we often encounter the patient response on an ordinal scale, for example, worse, same or better. Arbitrarily assigning scores (such as –1, 0, 1) to these ordinal categories for arithmetic operation can be inappropriate due to the fact that the relative distance between any two successive ordinal categories is not really equal or even comparable. One may also have difficulty in interpreting the mean of these arbitrary scores in terms of practical meaning. Grouping multiple categories into a single category to reduce ordinal outcomes into dichotomous responses can cause the loss of efficiency.

To reduce the number of patients in clinical trials, taking more than one measurement on each patient frequently arises in practice. Research on repeated ordinal responses has been intensive [1, 2, 3–5, 6, 7, 8, 9]. Francom et al. [2] applied a family of structural log-linear models and required investigators to assign the scores to represent the relative distance between ordinal categories. Agresti [1] addressed use of generalized linear models with different link functions and linear predictors. Ware et al. [9] as well as Kenward and Jones [4] discussed various approaches to analyze repeated categorical measurements. Parsons et al. [7] considered the proportional odds logistic regression with a range of working correlation models. All these publications focused discussions on model-based methods. By contrast, the methods proposed here is model-free and does not assume any parametric form for the data structure. Furthermore, there is no need to assume or specify any particular dependence structure between repeated measurements. Also, rather than concentrating attentions on testing equality between treatments in ordinal data with repeated measurements, there were publications [10, 11] discussing and deriving procedures for testing positive (quadrant) dependence under various parameter constraints and marginal modeling in two-way tables with ordinal categories. Agresti and Coull [12] focused discussion on model-based approach as well and addressed testing hypothesis against order-restricted alternatives. The purposes of these papers are different from what we focus here is to derive model-free procedures for testing equality of treatments, while the dependence between measurements taken within patients is nuisance effect.

Using the generalized odds ratio (GOR) [13], we develop in this paper model-free procedures for testing equality of treatments and derive interval estimators for the relative treatment effect in ordinal data with repeated measurements. We further develop a simple procedure for testing the treatment-by-period

^{*}Corresponding author: Kung-Jong Lui, San Diego State University, E-mail: kjl@rohan.sdsu.edu

interaction. We discuss the usefulness and limitations of test procedures developed here. To illustrate the use of these procedures, we consider two real-life data sets, one taken from a trial studying the gender effect on pain scores on an ordinal scale after hip joint resurfacing surgeries [7] and the other taken from a double-blind randomized trial comparing an active hypnotic drug with a placebo in insomnia patients with respect to the ordinal category of time to falling asleep [2].

2 Notation and Methods

Consider comparing two treatments in a randomized clinical trial, in which we randomly assign n_g patients to group g (=1, 2). For a randomly selected patient i(=1, 2,..., n_g) from group g, we let Y_{izg} denote the patient response at period z (=1, 2), where Y_{izg} takes one of L ordinal labels C_j (j=1, 2,..., L) with $C_1 < C_2 < C_3 < \cdots < C_L$. Let n_{rsg} denote the number of patients with $(Y_{i1g} = C_r, Y_{i2g} = C_s)$ among n_g patients in group *g*. The random vector $(n_{11g}, n_{12g}, \cdots n_{1Lg}, n_{21g}, n_{22g}, ..., n_{2Lg}, \cdots, n_{L1g}, n_{L2g}, ..., n_{LLg})'$ then follows the multinomial distribution with parameters n_g and $(\pi_{11g}, \pi_{12g}, \dots \pi_{1Lg}, \pi_{21g}, \pi_{21g}, \dots, \pi_{2Lg}, \dots, \pi_{L1g}, \dots, \pi_{L1g}, \dots)$ $\pi_{L2g},...,\pi_{Llg})'$, where π_{rsg} denotes the cell probability that a randomly selected patient from group g has the vector of responses $(Y_{i1g} = C_r, Y_{i2g} = C_s)$. We let "+" denote the summation over that particular subscript. For example, $\pi_{r+g}(=\sum_s \pi_{rsg})$ denotes the probability $P(Y_{ilg}=C_r)$ that a randomly selected patient from group g (=1, 2) has the response $Y_{i1g} = C_r$ at period 1. Following Agresti [13], we define $\Pi_{C1} =$ $\sum_{r=1}^{L-1} \sum_{r'=r+1}^{L} \pi_{r+1} \pi_{r'+2}$, denoting for period 1 the probability that the response of a randomly selected patient from group 2 is larger than that from group 1. We further define $\Pi_{D1} = \sum_{r=2}^{L} \sum_{r'=1}^{r-1} \pi_{r+1} \pi_{r'+2}$ denoting for period 1 the probability that the response of a randomly selected patient from group 1 is larger than that from group 2. The GOR of responses between groups 2 and 1 at period 1 is simply defined as $G_1 = \Pi_{C1}/\Pi_{D1}$. When there is no difference in treatment effects at period 1 between the two groups, $G_1 = 1$. If the treatment at period 1 in group 2 tends to increase the response as compared with that in group 1, $G_1 > 1$. If the former tends to decrease the response as compared with the latter, $G_1 < 1$. Similarly, we define the GOR of responses between groups 2 and 1 at period 2 as $G_2 = \Pi_{C2}/\Pi_{D2}$, where $\Pi_{C2} = \sum_{s=1}^{L-1} \sum_{s'=s+1}^{L} \pi_{+s1} \pi_{+s'2}$ and $\Pi_{D2} = \prod_{s=1}^{L} \pi_{+s1} \pi_{+s'2}$ $\sum_{s=2}^{L} \sum_{s'=1}^{s-1} \pi_{+s1} \pi_{+s'2}$. Note that when L=2, $G_z(z=1, 2)$ reduces to the regular odds ratio (OR) of responses in dichotomous data.

Note that we can estimate π_{rsg} by the unbiased consistent sample proportion estimator $\hat{\pi}_{rsg} = n_{rsg}/n_g$, and thereby, we can estimate G_1 at period z=1 between groups 2 and 1 by

$$\hat{G}_1 = \hat{\Pi}_{C1}/\hat{\Pi}_{D1},\tag{1}$$

where $\hat{\Pi}_{C1} = \sum_{r=1}^{L-1} \sum_{r'=r+1}^{L} \hat{\pi}_{r+1} \hat{\pi}_{r'+2}$ and $\hat{\Pi}_{D1} = \sum_{r=2}^{L} \sum_{r'=1}^{r-1} \hat{\pi}_{r+1} \hat{\pi}_{r'+2}$. Using the delta method [14], we can show that an estimated asymptotic variance of $\hat{G}_1(1)$ with the logarithmic transformation is given by [13, 15]

$$Var\left(\log(\hat{G}_{1})\right) = \frac{\sum_{r=1}^{L} \left[\sum_{r'=r+1}^{L} \hat{\pi}_{r'+2} - \hat{G}_{1} \sum_{r'=1}^{r-1} \hat{\pi}_{r'+2}\right]^{2} \hat{\pi}_{r+1}}{n_{1}(\hat{\Pi}_{G})^{2}} + \frac{\sum_{r'=1}^{L} \left[\sum_{r=1}^{r'-1} \hat{\pi}_{r+1} - \hat{G}_{1} \sum_{r=r'+1}^{L} \hat{\pi}_{r+1}\right]^{2} \hat{\pi}_{r'+2}}{n_{2}(\hat{\Pi}_{G})^{2}}.$$
 (2)

Note that we define $\sum_{r=L+1}^{L} \hat{\pi}_{r+1} = \sum_{r'=L+1}^{L} \hat{\pi}_{r'+2} = 0$. Similarly, we define $\sum_{r=1}^{0} \hat{\pi}_{r+1} = \sum_{r'=1}^{0} \hat{\pi}_{r'+2} = 0$.

Following the same arguments as for deriving \hat{G}_1 (1) and $\hat{Var}(\log(\hat{G}_1))$ (2), we obtain a consistent estimator for G_2 between groups 2 and 1 at period 2 as

$$\hat{G}_2 = \hat{\Pi}_{C2} / \hat{\Pi}_{D2},$$
 (3)

where $\hat{\Pi}_{C2} = \sum_{s=1}^{L-1} \sum_{s'=s+1}^{L} \hat{\pi}_{+s1} \hat{\pi}_{+s'2}$ and $\hat{\Pi}_{D1} = \sum_{s=2}^{L} \sum_{s'=1}^{s-1} \hat{\pi}_{+s1} \hat{\pi}_{+s'2}$. Again, using the

delta method, we can show that an estimated asymptotic variance of $\hat{G}_2(3)$ with the logarithmic transformation is given by

$$Var(\log(\hat{G}_{2})) = \frac{\sum_{s=1}^{L} \left[\sum_{s'=s+1}^{L} \hat{\pi}_{+s'2} - \hat{G}_{2} \sum_{s'=1}^{s-1} \hat{\pi}_{+s'2}\right]^{2} \hat{\pi}_{+s1}}{n_{1}(\hat{\prod}_{C2})^{2}} + \frac{\sum_{s'=1}^{L} \left[\sum_{s=1}^{s'-1} \hat{\pi}_{+s1} - \hat{G}_{2} \sum_{s=s'+1}^{L} \hat{\pi}_{+s1}\right]^{2} \hat{\pi}_{+s'2}}{n_{2}(\hat{\prod}_{C2})^{2}}.$$
 (4)

Also, we define $\sum_{s=L+1}^{L} \hat{\pi}_{+s1} = \sum_{s'=L+1}^{L} \hat{\pi}_{+s'2} = 0$ and $\sum_{s=1}^{0} \hat{\pi}_{+s1} = \sum_{s'=1}^{0} \hat{\pi}_{+s'2} = 0$ in (4).

Note that $\hat{G}_1(1)$ and $\hat{G}_2(3)$ are actually correlated. For clarity, we include the tedious details in derivation of $\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$ in **Appendix I**.

2.1 Test non-equality between treatments in the absence of interactions

When treatments received at the two periods in a group g are the same and there is no treatment-byperiod interaction (i. e., $G_1 = G_2$), we may apply a weighted average $w \log(\hat{G}_1) + (1 - w) \log(\hat{G}_2)$ to test $H_0:G_1=G_2=1$ versus $H_a:G_1\neq 1$ or $G_2\neq 1$, where $0\leq w\leq 1$ is the weight reflecting the relative importance of responses at period 1 to those at period 2, and can be assigned by clinicians based on their subjective knowledge. If we have no prior preference to assign the weight, we may simply set w equal to 0.50. This leads us to consider the following summary test procedure over the two periods. We will reject $H_0:G_1=G_2=1$ at the α -level if

$$|\log(\hat{G}_1) + \log(\hat{G}_2)|/\sqrt{\hat{Var}(\log(\hat{G}_1) + \log(\hat{G}_2))} > Z_{\alpha/2},$$
(5)

where $\hat{Var}(\log(\hat{G}_1) + \log(\hat{G}_2)) = \hat{Var}(\log(\hat{G}_1)) + \hat{Var}(\log(\hat{G}_2)) + 2\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2)), \hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$ is given in (A.3) with replacing parameters by their corresponding parameter estimators (**Appendix I**), and Z_{α} is the upper $100(\alpha)$ th percentile of the standard normal distribution. Note that because the sampling distribution of \hat{G}_z (z=1, 2), which is a ratio of two random variables, can be skewed, we use the logarithmic transformation to improve the normal approximation here.

2.2 Test non-equality between treatments in the presence of interactions

When treatments received at the two periods in a group are not the same or there is a treatment-by-period interaction (i. e., $G_1 \neq G_2$), the test procedure (5) can be meaningless or lose power. When there is a treatment-by-period interaction, the estimates \hat{G}_1 and \hat{G}_2 may fall in opposite relative directions. For example, consider the situation $G_1 = 1$ and $G_2 > 1$, in which $\log(\hat{G}_1)$ can be negative with a non-negligible probability, while $\log(\hat{G}_2)$ tends to be positive. In this case, $|\log(\hat{G}_1) + \log(\hat{G}_2)|$ can be small due to cancelation between values of $\log(\hat{G}_1)$ and $\log(\hat{G}_2)$. Therefore, the summary test procedure (5) can lack power in these cases. To alleviate this concern, we may consider the following bivariate test procedure. We will reject $H_0: G_1 = G_2 = 1$ at the α -level if

$$\left(\log(\hat{G}_1), \log(\hat{G}_2)\right) \underline{\hat{\Sigma}}^{-1} \left(\frac{\log(\hat{G}_1)}{\log(\hat{G}_2)}\right) > \chi_{\alpha}^2(2), \tag{6}$$

where $\hat{\underline{\Sigma}}$ is the estimated covariance matrix with diagonal elements equal to $\hat{Var}(\log(\hat{G}_1))$ and $\hat{Var}(\log(\hat{G}_2))$, and the off-diagonal element equal to $\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$, as well as χ^2_{α} is the upper $100(\alpha)$ th percentile of the central chi-squared distribution with 2 degrees of freedom.

2.3 Procedure for testing the group-by-period interaction

When wishing to study whether there is a group-by-period interaction, we may consider testing $H_0:G_1=G_2$ versus $H_a:G_1\neq G_2$. We will reject $H_0:G_1=G_2$ at the α -level if

$$|\log(\hat{G}_1) - \log(\hat{G}_2)|/\sqrt{\hat{Var}(\log(\hat{G}_1) - \log(\hat{G}_2))} > Z_{\alpha/2}, \tag{7}$$

where $\hat{Var}(\log(\hat{G}_1) - \log(\hat{G}_2)) = \hat{Var}(\log(\hat{G}_1)) + \hat{Var}(\log(\hat{G}_2)) - 2\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$. Note that when treatments received at two periods in a group are the same, the group-by-period interaction may actually represent the treatment-by-period interaction.

2.4 Interval estimation of the relative treatment effect

When treatments received at the two periods in a group are the same and there is no treatment-by-period interaction, we let G_0 denote the common value of G_1 and G_2 . On the basis of $\hat{G}_1(1)$, $\hat{Var}(\log(\hat{G}_1))(2)$, $\hat{G}_2(3)$, $\hat{Var}(\log(\hat{G}_2))(4)$ and $\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$ (A.3), we may obtain 100(1- α)% confidence interval for G_0 as

$$[\exp(LG_l), \exp(LG_u)],$$
 (8)

where
$$LG_l = \left(\log(\hat{G}_1) + \log(\hat{G}_2)\right)/2 - Z_{\alpha/2}\sqrt{\hat{Var}\left(\log(\hat{G}_1) + \log(\hat{G}_2)\right)}/2$$
, and $LG_u = \left(\log(\hat{G}_1) + \log(\hat{G}_2)\right)/2 + Z_{\alpha/2}\sqrt{\hat{Var}\left(\log(\hat{G}_1) + \log(\hat{G}_2)\right)}/2$.

When treatments received at the two periods are not the same or there is a treatment-by-period interaction, we may wish to obtain an interval estimator for $G_z(z=1, 2)$ separately. On the basis of \hat{G}_z and $\hat{Var}(\log(\hat{G}_z))$ for z=1, 2, we obtain $100(1-\alpha)\%$ confidence interval for $G_z(z=1, 2)$ as

$$\left[\hat{G}_z \exp\left(-Z_{\alpha/2} \sqrt{\hat{Var}(\log(\hat{G}_z))}\right), \quad \hat{G}_z \exp\left(Z_{\alpha/2} \sqrt{\hat{Var}(\log(\hat{G}_z))}\right)\right]. \tag{9}$$

3 Examples

To illustrate the use of point estimators \hat{G}_1 and \hat{G}_2 , test procedures and interval estimators developed here, we first consider the trial studying hip joint resurfacing surgeries [7]. To alleviate pain and debilitation caused by osteoarthritis, rheumatoid arthritis, fractures and other hip related problems, hip replacement surgery is a widely-used procedure. To assess the failure rate and prognosis after hip joint resurfacing, we may use Harris scores [16], of which one important component is the pain scores coded on an ordinal scale: none (1), slight (2), mild (3) and moderate or marked (4) pain in hip joint. We summarize in Table 1 the data regarding the frequency distribution according to pain scores taken at two and five years after hip joint resurfacing for 58 patients by gender published elsewhere [7]. Because there were very few patients with moderate or marked pain (4), we grouped these patients and patients with mild pain (3) into one category without loss of much information. It is of interest to study whether there is a difference in pain scores between genders over time after surgeries. In terms of our notation, we define "females" and "males" as group (g = 1) and 2, as well as "two years" and "five years" after the surgery as period (g = 1) and 2. From Table 1, the numbers of patients are g = 1 and g = 1 and g = 1 for the two comparison groups. We can calculate, for

Table 1: Frequency distribution of patients with pain scores coded as: none (1), slight (2), mild, moderate or marked pain (3) taken at two and five years post-surgery between females and males.

Gender	At Two Years			At Five Years		
		None	Slight	Mild, Moderate	Marginal Total	Marginal
				or Marked Pain		Percentage
Female	None	7	1	2	10	0.476
	Slight	3	1	3	7	0.333
	Mild, Moderate or Marked Pain	0	1	3	4	0.190
	Marginal Total	10	3	8	21	
	Marginal Percentage	0.476	0.143	0.381		1.000
		None	Slight	Mild, Moderate	Marginal Total	Marginal
				or Marked Pain		Percentage
Male	None	19	7	2	28	0.757
	Slight	1	3	1	5	0.135
	Mild, Moderate or Marked Pain	0	0	4	4	0.108
	Marginal Total	20	10	7	37	
	Marginal Percentage	0.541	0.270	0.189		1.000

example, the estimated cell proportion $\hat{\pi}_{111} = n_{111}/n_1 = 7/21$. The marginal percentage for each row (or column) is simply equal to the corresponding marginal total divided by the total number of patients $n_g(g=1, 2)$. For example, the marginal percentage $\hat{\pi}_{1+1}$ in the first row for females is 0.476 (=10/21). Using marginal percentages for the rows between the two sub-tables (Table 1), we can then calculate the GOR estimate G_1 (1) of pain scores at two years post surgeries as

 $[0.476 \times (0.135 + 0.108) + 0.333 \times 0.108]/[0.333 \times 0.757 + 0.190 \times (0.757 + 0.135)] = 0.360$. Similarly, using marginal percentages for columns between the two sub-tables, we obtain $\hat{G}_2 = 0.637$ at five years post surgeries. As an example, we can interpret $\hat{G}_1 = 0.360$ as that the odds of a randomly selected male with the pain score higher than a randomly selected female at two years post surgeries is 0.360, given no ties in pain scores between genders. Furthermore, when employing interval estimators (9), we obtain 95% confidence intervals [0.133, 0.974] and [0.250, 1.623] for G_1 and G_2 , respectively. Since the upper limit of the 95% confidence interval for G_1 falls below 1, there is significant evidence at the 5% level that males tend to have pain scores lower than females at two years after surgeries. However, since \hat{G}_7 (= 0.637) is closer to 1 than G_1 (= 0.360), this difference in pain scores between genders seems to decrease as the time increases from two years to five years after surgeries. Note that the above resulting 95% confidence interval for G_2 covers 1. Thus, the difference in pain scores between genders is no longer significant at the 5 % level after five years post surgeries, despite that males is still likely to have pain scores lower than females (because $G_2 < 1$). When using test procedures (5) and (6) for testing the overall equality of pain scores across two periods between genders, we obtain p-values as 0.086 and 0.131. On the basis of the above results, we may conclude that although males tend to fall, as compared with females, in categories with low pain scores especially at two years post surgeries, the difference in pain scores between genders seems to become smaller in the long-term results.

When assuming a normal random effects proportional odds model [17], we may employ *Proc Glimmix* in SAS [18] to study the difference in pain scores between genders on the basis of the model-based approach. Using the data in Table 1, we have obtained the parameter estimate (and its estimated standard error (SE)) for the relative gender effect of males versus females to be 0.961 (SE = 0.6148). This leads the p-value for testing the equality of pain scores across the two periods between genders is 0.124, which is similar to those obtained by use of test procedures (5) and (6). Also, note that the parameter estimate 0.961 is larger than 0. Thus, males tend to fall in categories with lower pain scores than females. This inference is identical to that obtained on the basis of the GOR focused here.

We may sometimes encounter a trial in which the patient response taken at period (z=) 1 actually represents the baseline response. In this case, we may apply the test procedure (7), developed for testing

Table 2: Frequency distribution of patients with time to falling asleep (in minutes) taken at the end of one-week washout period and at the conclusion of two-week treatment period.

Treatment	At One-week Period	At Two-week Treatment Period					
		<20	20-30	30-60	>60	Total	
Active	< 20	7	4	1	0	12	
	20-30	11	5	2	2	20	
	30-60	13	23	3	1	40	
	> 60	9	17	13	8	47	
	Total	40	49	19	11	119	
		<20	20-30	30-60	>60	Total	
Placebo	< 20	7	4	2	1	14	
	20-30	14	5	1	0	20	
	30-60	6	9	18	2	35	
	> 60	4	11	14	22	51	
	Total	31	29	35	25	120	

group-by-period interaction, to study whether there is a relative treatment effect. To illustrate this point, we consider the data (Table 2) taken from a double-blind randomized clinical trial comparing an active hypnotic drug (g=1) with a placebo (g=2) in patients with insomnia [1, 14, 2]. The outcome of interest is to respond the question "How quickly did you fall asleep after going to bed?" and is recorded on a fourpoint ordinal scale (< 20, 20–30, 30–60, > 60 in minutes). Each participated subject was asked this question twice, one after a one-week placebo washout period for both groups and the other at the conclusion of a two-week treatment period. Using the data in Table 2, we obtain $\hat{G}_1 = 1.031$ and $\hat{G}_2 = 1.883$. Because subjects received placebo at the first one-week washout period in both groups, the estimate for the GOR of responses is expected to be, as shown $G_1 = 1.031$, around 1 due to randomization. At the end of two-week treatment period, the estimate for the GOR of responses becomes $\hat{G}_2 = 1.883$ with 95% confidence interval using interval (9) given by [1.270, 2.792]. Because this lower confidence limit falls above 1, one may conclude that the hypnotic drug changes the GOR over periods and tends to reduce time to falling asleep. When using test procedures (6) and (7), we obtain p-value as 0.002 and 0.004; these small p-values also suggest that the hypnotic drug can significantly reduce the time to falling asleep. We note that all these test results are essentially similar to those based on the proportional odds model with the cumulative logits found elsewhere [1]. When using the summary test procedure (5) using data in Table 2, we obtain the p-value 0.055. Note that since the GOR at the initial period is around 1, there is non-negligible probability that the time to falling asleep for patients in the placebo group is smaller than those in the other group at the initial washout period. As noted previously, use of a summary test procedure (5) over two periods in this case may not only be senseless (due to different treatments received at two periods), but also lack power.

4 Discussion

We can employ, as demonstrated here, the GOR to measure the relative treatment effect without the need to assume any specific parametric model. Since the GOR has a simple interpretation and is easily understood, the GOR is of use in ordinal data. In fact, the GOR is closely related to the gamma correlation [19], a commonly-used measure of the strength of association between two ordinal variables. We refer readers to some publications on estimation of the GOR and its applications under other situations [20, 21, 22–24]. When repeated measurements are taken at the same time (or there are no period effects), one may employ the Dirichlet-multinomial distribution to model the intraclass correlation between repeated measurements [25]. As considered in the above two examples, however, repeated measurements on patients are often taken at different time intervals in clinical trials. The period effect is likely to exist and is required to be incorporated to avoid bias in

data analysis [26, 27]. The methods based on the Dirichlet-multinomial model for cluster sampling without accounting for the period effect would not be appropriate for use in situations focused here.

The proportional odds model is probably one of the most commonly-used models to analyze ordinal data. Just like all model-based approaches, we can easily extend the proportional odds model to account for confounders (if there were) or accommodate other general situations. However, the implicit assumption of the proportional odds model can be badly violated by many bivariate distributions [13, 28, 29]. Furthermore, when applying Proc Glimmix in SAS [18] based on the random effects proportional odds model to ordinal data with repeated measurements, we need to assume that the random effects (accounting for the intraclass correlation between repeated measurements) due to patients follow a normal distribution. This normal assumption for random effects can be difficult to be justified. By contrast, the proposed method is modelfree. It does not require the random effects due to patients to follow the normal distribution, not does assume any parametric models for the data structure. Thus, our methods are applicable despite of various parametric models for the underlying data structure and distribution assumptions for the patient random effects. Furthermore, the point estimators, test procedures and interval estimators developed here can all be expressed in closed forms. Readers may employ these test procedures and estimators by use of a pocket calculator even without knowledge of any statistical software. The interpretation of the GOR is, as illustrated in examples, easily understood. When there are confounders in a trial of a large size, we may extend the methods proposed here by use of stratified analysis with strata determined by the combined levels of confounders [23]. But we want to note that the model-based approach can be preferable to the model-free approach proposed here if there are many covariates to adjust for a trial of a small or moderate size.

Finally, we note that using similar arguments as above, it is straightforward to extend the results to accommodate the cases with three or more periods. We outlines the extension of results presented here to accommodate three periods in Appendix II.

In summary, we have developed model-free test procedures for testing equality of treatments in ordinal data with repeated measurements. We have further derived interval estimators for the relative treatment effect measured by the GOR. We recommend use of the summary test procedure to improve power when treatments received at two periods in a group are the same and there is no treatment-by-period interaction. However, we should cautiously employ this summary test procedure when there is a treatment-by-period interaction. The bivariate test procedure can be of use in this case. We further outline the extension of results to accommodate three periods. The results, findings and discussions should have use for biostatisticians and clinicians when they encounter ordinal responses with repeated measurements.

Acknowledgements: The author wishes to thank the associate editor and two reviewers for many valuable comments and suggestions to improve the clarity and contents of this article.

Funding: The research received no specific grant from any funding agency in the public, commercial, or not for-for-profit sectors.

Appendix I

Suppose that random vectors (X_1, X_2) and (Y_1, Y_2) have the joint probability density functions $f(X_1, X_2)$ and $f(Y_1, Y_2)$, respectively. Suppose further that $(X_1, X_2)'$ and $(Y_1, Y_2)'$ are mutually independent. We can show that the covariance

$$Cov(X_1Y_1, X_2Y_2) = Cov(X_1, X_2)Cov(Y_1, Y_2) + Cov(X_1, X_2)E(Y_1)E(Y_2) + Cov(Y_1, Y_2)E(X_1)E(X_2).$$
(A.1)

On the basis of (A.1), we have

$$Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s1}\hat{\pi}_{+s'2}) = [(\pi_{rs1} - \pi_{r+1}\pi_{+s1})/n_1][(\pi_{r's'2} - \pi_{r'+2}\pi_{+s'2})/n_2] + [(\pi_{rs1} - \pi_{r+1}\pi_{+s1})/n_1]\pi_{r'+2}\pi_{+s'2} + \pi_{r+1}\pi_{+s1}[(\pi_{r's'2} - \pi_{r'+2}\pi_{+s'2})/n_2].$$
(A.2)

Using the delta method, we may obtain the asymptotic covariance

$$\begin{aligned} Cov\Big(\log(\hat{G}_1),\ \log(\hat{G}_2)\Big) &= Cov\big(\log(\hat{\Pi}_{C1}),\ \log(\hat{\Pi}_{C2})\big) - Cov\big(\log(\hat{\Pi}_{C1}),\ \log(\hat{\Pi}_{D2})\big) \\ &- Cov\big(\log(\hat{\Pi}_{D1}),\ \log(\hat{\Pi}_{C2})\big) + Cov\big(\log(\hat{\Pi}_{D1}),\ \log(\hat{\Pi}_{D2})\big), \end{aligned} \tag{A.3}$$

where
$$Cov(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{C2})) = [\sum_{r=1}^{L-1} \sum_{s=1}^{L-1} \sum_{r'=r+1}^{L} \sum_{s'=s+1}^{L} Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s1}\hat{\pi}_{+s'2})]/(\Pi_{C1}\Pi_{C2}),$$

$$Cov(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{D2})) = [\sum_{r=1}^{L-1} \sum_{s=2}^{L} \sum_{r'=r+1}^{L} \sum_{s'=1}^{s-1} Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s1}\hat{\pi}_{+s'2})]/(\Pi_{C1}\Pi_{D2}),$$

$$Cov(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{C2})) = [\sum_{r=2}^{L} \sum_{s=1}^{L-1} \sum_{r'=1}^{r-1} \sum_{s'=s+1}^{L} Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s1}\hat{\pi}_{+s'2})]/(\Pi_{D1}\Pi_{C2}), \text{ and }$$

$$Cov(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{D2})) = [\sum_{r=2}^{L} \sum_{s=2}^{L} \sum_{r'=1}^{r-1} \sum_{s'=1}^{s-1} Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s1}\hat{\pi}_{+s'2})]/(\Pi_{D1}\Pi_{D2}).$$

We can obtain an estimated asymptotic covariance $Cov \log(\hat{G}_1)$, $\log(\hat{G}_2)$) with replacing the parameters in (A.3) by their corresponding parameter estimators.

Appendix II

For a randomly selected patient $i(=1, 2, ..., n_g)$ from group g, we let Y_{izg} denote the patient response at period z (=1, 2, 3). Let n_{rstg} denote the number of patients with $(Y_{i1g} = C_r, Y_{i2g} = C_s, Y_{i3g} = C_t)$ among n_g patients in group g. The random vector $(n_{111g}, n_{112g}, \cdots, n_{11Lg}, n_{121g}, n_{122g}, ..., n_{12Lg}, \cdots, n_{L11g}, n_{L12g}, ..., n_{L1Lg}, \cdots, n_{L1Lg}, \dots, n_{L1Lg}, \cdots, n_{L1Lg}, \dots, n_{L1$

$$\hat{G}_{z} = \hat{\Pi}_{Cz}/\hat{\Pi}_{Dz}, \tag{A.4}$$
 where $\hat{\Pi}_{C1} = \sum_{r=1}^{L-1} \sum_{r'=r+1}^{L} \hat{\pi}_{r++1} \hat{\pi}_{r'++2}$ and $\hat{\Pi}_{D1} = \sum_{r=2}^{L} \sum_{r'=1}^{r-1} \hat{\pi}_{r++1} \hat{\pi}_{r'++2};$
$$\hat{\Pi}_{C2} = \sum_{s=1}^{L-1} \sum_{s'=s+1}^{L} \hat{\pi}_{+s+1} \hat{\pi}_{+s'+2} \text{ and } \hat{\Pi}_{D2} = \sum_{s=2}^{L} \sum_{s'=1}^{s-1} \pi_{+s+1} \hat{\pi}_{+s'+2}; \hat{\Pi}_{C3} = \sum_{t=1}^{L-1} \sum_{t'=t+1}^{L} \hat{\pi}_{++t'1} \hat{\pi}_{++t'2} \text{ and } \hat{\Pi}_{D3} = \sum_{t=2}^{L} \sum_{t'=1}^{t-1} \hat{\pi}_{++t'1} \hat{\pi}_{++t'2}.$$

Using the delta method [14], we obtain an estimated asymptotic variance for $\log(\hat{G}_1)$ with the logarithmic transformation as

$$Var\left(\log(\hat{G}_{1})\right) = \frac{\sum_{r=1}^{L} \left[\sum_{r'=r+1}^{L} \hat{\pi}_{r'++2} - \hat{G}_{1} \sum_{r'=1}^{r-1} \hat{\pi}_{r'++2}\right]^{2} \hat{\pi}_{r++1}}{n_{1}(\hat{\prod}_{C1})^{2}} + \frac{\sum_{r'=1}^{L} \left[\sum_{r=1}^{r'-1} \hat{\pi}_{r++1} - \hat{G}_{1} \sum_{r=r'+1}^{L} \hat{\pi}_{r++1}\right]^{2} \hat{\pi}_{r'++2}}{n_{2}(\hat{\prod}_{C1})^{2}},$$
(A.5)

where $\sum_{r=L+1}^{L} \hat{\pi}_{r++1} = \sum_{r'=L+1}^{L} \hat{\pi}_{r'++2} = \sum_{r=1}^{0} \hat{\pi}_{r++1} = \sum_{r'=1}^{0} \hat{\pi}_{r'++2} = 0$. Similarly, we obtain the estimated asymptotic variances for $\log(\hat{G}_2)$ as given by

$$\hat{Var}\left(\log(\hat{G}_{2})\right) = \frac{\sum_{s=1}^{L} \left[\sum_{s'=s+1}^{L} \hat{\pi}_{+s'+2} - \hat{G}_{2} \sum_{s'=1}^{s-1} \hat{\pi}_{+s'+2}\right]^{2} \hat{\pi}_{+s+1}}{n_{2}(\hat{\Pi}_{+})^{2}} + \frac{\sum_{s'=1}^{L} \left[\sum_{s=1}^{s'-1} \hat{\pi}_{+s+1} - \hat{G}_{2} \sum_{s=s'+1}^{L} \hat{\pi}_{+s+1}\right]^{2} \hat{\pi}_{+s'+2}}{n_{2}(\hat{\Pi}_{+})^{2}}, \quad (A.6)$$

where $\sum_{s=L+1}^{L} \hat{\pi}_{+s+1} = \sum_{s'=L+1}^{L} \hat{\pi}_{+s'+2} = \sum_{s=1}^{0} \hat{\pi}_{+s+1} = \sum_{s'=1}^{0} \hat{\pi}_{+s'+2} = 0$. Also, we obtain an estimated asymptotic variances for $\log(\hat{G}_3)$ as

$$\widehat{Var}\left(\log(\hat{G}_{3})\right) = \frac{\sum_{t=1}^{L} \left[\sum_{t'=t+1}^{L} \hat{\pi}_{++t'2} - \hat{G}_{3} \sum_{t'=1}^{t-1} \hat{\pi}_{++t'2}\right]^{2} \hat{\pi}_{++t1}}{n_{1}(\hat{\prod}_{C3})^{2}} + \frac{\sum_{t'=1}^{L} \left[\sum_{t=1}^{t'-1} \hat{\pi}_{++t1} - \hat{G}_{3} \sum_{t=t'+1}^{L} \hat{\pi}_{++t1}\right]^{2} \hat{\pi}_{++t'2}}{n_{2}(\hat{\prod}_{C3})^{2}}, \quad (A.7)$$

where $\sum_{t=L+1}^{L} \hat{\pi}_{++t1} = \sum_{t'=L+1}^{L} \hat{\pi}_{++t'2} = \sum_{t=1}^{0} \hat{\pi}_{++t1} = \sum_{t'=1}^{0} \hat{\pi}_{++t'2} = 0$. Note that the covariance

$$\begin{aligned} Cov(\hat{\pi}_{r++g}, \hat{\pi}_{+s+g}) &= \sum_{k} \sum_{k'} Cov(\hat{\pi}_{r+kg}, \hat{\pi}_{+sk'g}) \\ &= \sum_{k} Cov(\hat{\pi}_{r+kg}, \hat{\pi}_{+skg}) + \sum_{k} \sum_{k' \neq k'} Cov(\hat{\pi}_{r+kg}, \hat{\pi}_{+sk'g}) \\ &= \sum_{k} (\pi_{rskg} - \pi_{r+kg} \pi_{+skg}) - \sum_{k} \sum_{k' \neq k'} (\hat{\pi}_{r+kg} \hat{\pi}_{+sk'g}) \\ &= \pi_{rs+g} - \pi_{r++g} \pi_{+s+g} \end{aligned}$$
(A.8)

Thus, from (A.1) we have

$$Cov(\hat{\pi}_{r+1}\hat{\pi}_{r'+2}, \hat{\pi}_{+s+1}\hat{\pi}_{+s'+2}) = [(\pi_{rs+1} - \pi_{r+1}\pi_{+s+1})/n_1][(\pi_{r's'+2} - \pi_{r'+2}\pi_{+s'+2})/n_2] + [(\pi_{rs+1} - \pi_{r+1}\pi_{+s+1})/n_1]\pi_{r'+2}\pi_{+s'+2} + \pi_{r+1}\pi_{+s+1}[(\pi_{r's'+2} - \pi_{r'+2}\pi_{+s'+2})/n_2].$$
(A.9)

Using the delta method, we obtain the asymptotic covariance

$$Cov\left(\log(\hat{G}_{1}), \log(\hat{G}_{2})\right) = Cov\left(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{C2})\right) - Cov\left(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{D2})\right) - Cov\left(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{D2})\right) + Cov\left(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{D2})\right),$$
(A.10)

where
$$Cov(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{C2})) = [\sum_{r=1}^{L-1} \sum_{s=1}^{L-1} \sum_{r'=r+1}^{L} \sum_{s'=s+1}^{L} Cov(\hat{\pi}_{r++1}\hat{\pi}_{r'++2}, \hat{\pi}_{+s+1}\hat{\pi}_{+s'+2})]/(\Pi_{C1}\Pi_{C2}),$$

$$Cov(\log(\hat{\Pi}_{C1}), \log(\hat{\Pi}_{D2})) = [\sum_{r=1}^{L-1} \sum_{s=2}^{L} \sum_{r'=r+1}^{L} \sum_{s'=1}^{s-1} Cov(\hat{\pi}_{r++1}\hat{\pi}_{r'++2}, \hat{\pi}_{+s+1}\hat{\pi}_{+s'+2})]/(\Pi_{C1}\Pi_{D2}),$$

$$Cov(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{C2})) = [\sum_{r=2}^{L} \sum_{s=1}^{L-1} \sum_{r'=1}^{r-1} \sum_{s'=s+1}^{L} Cov(\hat{\pi}_{r++1}\hat{\pi}_{r'++2}, \hat{\pi}_{+s+1}\hat{\pi}_{+s'+2})]/(\Pi_{D1}\Pi_{C2}), \text{ and }$$

$$Cov(\log(\hat{\Pi}_{D1}), \log(\hat{\Pi}_{D2})) = [\sum_{r=2}^{L} \sum_{s=2}^{L} \sum_{r'=1}^{r-1} \sum_{s'=1}^{s-1} Cov(\hat{\pi}_{r++1}\hat{\pi}_{r'++2}, \hat{\pi}_{+s+1}\hat{\pi}_{+s'+2})]/(\Pi_{D1}\Pi_{D2}).$$

We obtain the estimated asymptotic covariance $\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_2))$ by substituting parameter estimators for their corresponding parameters in (A.10). Using the same arguments, we can also obtain $\hat{Cov}(\log(\hat{G}_1), \log(\hat{G}_3))$ and $\hat{Cov}(\log(\hat{G}_2), \log(\hat{G}_3))$. These lead us to obtain, as for the ordinal data with two repeated measurements, the summary test procedure, the trivariate test procedure, the test procedure for treatment-by-period interaction, as well as interval estimators for the GOR in the ordinal data with three repeated measurements. For example, consider the trivariate test procedure for testing $H_0: G_1 = G_2 = G_3 = 1$. We will reject H_0 at the α -level if

$$\left(\log(\hat{G}_1), \log(\hat{G}_2), \log(\hat{G}_3)\right) \underline{\hat{\Sigma}}^{-1} \begin{pmatrix} \log(\hat{G}_1) \\ \log(\hat{G}_2) \\ \log(\hat{G}_3) \end{pmatrix} > \chi_{\alpha}^2(3), \tag{A.11}$$

where $\hat{\Sigma}$ is the estimated covariance matrix with diagonal elements equal to $\hat{Var}(\log(\hat{G}_1))$, $\hat{Var}(\log(\hat{G}_2))$ $\text{ and } \textit{Var}\Big(\log(\hat{G}_3)\Big), \text{ and off-diagonal elements equal to } \textit{Cov}\Big(\log(\hat{G}_1), \log(\hat{G}_2)\Big), \textit{Cov}\Big(\log(\hat{G}_1), \log(\hat{G}_3)\Big) \text{ and } \text{ord} \text{ and } \text{ord} \text{ ord} \text{ ord}$ $\hat{Cov}(\log(\hat{G}_2), \log(\hat{G}_3))$, respectively.

References

- 1. Agresti A. A survey of models for repeated ordered categorical response data. Stat Med 1989;8:1209-24.
- 2. Francom SF, Chuang-Stein C, Landis JR. A log-linear model for ordinal data to characterize differential change among treatments. Stat Med 1989;8:571-82.
- 3. Kenward MG, Jones B. The analysis of categorical data from cross-over trials using a latent variable model. Stat Med 1991;10:1607-19.
- 4. Kenward MG, Jones B. Alternative approaches to the analysis of binary and categorical repeated measurements. J Biopharm Stat 1992:2:137-70.
- 5. Koch GG, Landis JR, Freeman JL, Freeman JrDH, Lehnen RG. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 1977;33:133-58.
- 6. McCullagh P. A logistic model for paired comparisons with ordered categorical data. Biometrika 1977;64:449-53.
- 7. Parsons NR, Costa ML, Achten J, Stallard N. Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. Comput Stat Data Anal 2009;53:632-41.
- 8. Stram DO, Wei LJ, Ware JH. The analysis of repeated ordinal categorical outcomes with possible missing observations and time-dependent covariates. J Am Stat Assoc 1989;83:631-7.
- 9. Ware JH, Lipsitz S, Speizer FE. Issues in the analysis of repeated categorical outcomes. Stat Med 1988;7:95-107.
- 10. Bartolucci F, Forcina A, Dardanoni V. Positive quadrant dependence and marginal modeling in two-way tales with ordered margins. J Am Stat Assoc 2001;96:1497-505.
- 11. Bartolucci F, Forcina A. Extended RC association models allowing for order restrictions and marginal modeling. J Am Stat Assoc 2002;97:1192-9.
- 12. Agresti A, Coull BA. The analysis of contingency tables under inequality constraints. J Stat Plan Infernce 2002;107:45-73.
- 13. Agresti A. Generalized odds ratios for ordinal data. Biometrics 1980:36:59-67.
- 14. Agresti A. Categorical Data Analysis. New York: Wiley, 1990.
- 15. Lui K-J. Statistical Estimation of Epidemiology Risk. New York: Wiley, 2004.
- 16. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by mold arthroplasty. An endresult study using a new method of result evaluation. J Bone Joint Surg 1969;51-A:737-55.
- 17. Ezzet F, Whitehead J. A random effects model for ordinal responses from a crossover trial. Stat Med 1991;10:901-7.
- 18. SAS Institute, Inc. User's Guide, 2nd ed. Cary, NC: SAS Institute, 2009.
- 19. Goodman LA, Kruskal WH. Measure of association for cross classification. J Am Stat Assoc 1954;49:732-64.
- 20. Edwardes MD, Baltzan M. The generalization of the odds ratio, risk ratio and risk difference to r k tables. Stat Med 2000;19:1901-14.
- 21. Lui K-J. Notes on estimation of the general odds ratio and the general risk difference for paired-sample data. Biometric J 2002a:44:957-68.
- 22. Lui K-J, Chang K-C. Hypothesis testing and estimation in ordinal data under a simple crossover design. J Biopharm Stat 2012;22:1137-47.
- 23. Lui K-J, Chang K-C. Notes on interval estimation of the generalized odds ratio under stratified random sampling. J Biopharm Stat 2013a:23:513-25.
- 24. Lui K-J, Chang K-C. Notes on testing noninferiority in ordinal data under the parallel groups design. J Biopharm Stat 2013b;23:1294-307.
- 25. Lui K-J. Interval estimation of generalized odds ratio in data with repeated measurements. Stat Med 2002b;21:
- 26. Gart JJ. An exact test for comparing matched proportions in crossover designs. Biometrika 1969;56:75-80.
- 27. Senn S. Cross-over Trials in Clinical Research, 2nd ed. Chichester: Wiley, 2002.
- 28. Fleiss JL. On the asserted invariance of the odds ratio. Br J Preventive Soc Med 1970;24:45-6.
- 29. Mosteller F. Association and estimation in contingency tables. J Am Stat Assoc 1968;63:1-28.