Kung-Jong Lui[1]

Testing Equality of Treatments under an Incomplete Block Crossover Design with Ordinal Responses

[1] Department of Mathematics and Statistics, College of Sciences, San Diego State University, San Diego, CA 92182-7720, USA, E-mail: kjl@rohan.sdsu.edu

Abstract:
The generalized odds ratio (GOR) for paired sample is considered to measure the relative treatment effect on patient responses in ordinal data. Under a three-treatment two-period incomplete block crossover design, both asymptotic and exact procedures are developed for testing equality between treatments with ordinal responses. Monte Carlo simulation is employed to evaluate and compare the finite-sample performance of these test procedures. A discussion on advantages and disadvantages of the proposed test procedures based on the GOR versus those based on Wald's tests under the normal random effects proportional odds model is provided. The data taken as a part of a crossover trial studying the effects of low and high doses of an analgesic versus a placebo for the relief of pain in primary dysmenorrhea over the first two periods are applied to illustrate the use of these test procedures.

Keywords: generalized odds ratio, incomplete block, crossover trial, ordinal data, Mantel-Haenszel test
DOI: 10.1515/ijb-2016-0069

1 Introduction

When studying non-curable chronic diseases, including angina pectoris, epilepsy, hypertension, asthma, etc., we may often consider using a crossover design to reduce the number of patients needed for a parallel group design [1–14]. When there are more than two treatments under comparison, the trial duration for a crossover design can be much longer than that for a parallel group design if each patient is to receive every treatment in use of the former. The longer the duration of a trial, the higher is the patient risk of being lost to follow-up. Furthermore, a lengthy trial duration can cause the difficulty in recruiting patients into a trial and ensuring patients to closely follow a study protocol. To alleviate these concerns, we may consider using an incomplete block crossover design, in which each patient is to receive only a subset of treatments [1]. For example, consider the double-blind placebo controlled crossover trial comparing 12μg and 24μg of formoterol solution aerosol with a placebo [1]. For practical reasons, it was decided that each patient could receive only two of the three treatments: the placebo, 12μg and 24μg of formoterol solution. Although there were some publications [1, 12, 15–18] on the incomplete block crossover design, all these focused discussion on either continuous or binary data. The discussion on testing equality of treatments with ordinal responses under an incomplete block crossover design is limited [1, 7, 8].

Because ordinal responses are not on an interval scale, it is generally not appropriate to apply arithmetic operation to ordinal data [19]. In practice, we may commonly assign arbitrary scores to ordinal responses and do hypothesis testing with use of the t-test. Since the relative distances between consecutive categories in ordinal data are not truly comparable, converting ordinal responses into a universally agreeable score scale is difficult. Also, how to provide a meaningful and easily-understood summary measure based on these arbitrarily assigned scores to quantify the relative treatment effect can be challenging. On the other hand, if we dichotomize the ordinal responses into binary outcomes, the test procedures for binary data will probably lose efficiency.

In this paper, we propose use of the generalized odds ratio (GOR) for paired samples [20–22] to measure the relative treatment effect on patient responses in ordinal data. We focus our discussion on an incomplete block two-period crossover trial comparing three treatments with ordinal responses. We derive asymptotic test procedures based on the weighted-least-squares (WLS) and Mantel-Haenszel (MH) estimators [23] for testing equality between treatments. We further derive the exact test procedures for testing equality of treatments for

Kung-Jong Lui is the corresponding author.

small-sample cases. We employ Monte Carlo simulation to evaluate the finite-sample performance of these test procedures in a variety of situations. We use the data taken as a part of the crossover trial [24] comparing the low and high doses of an analgesic with a placebo for the relief of pain in primary dysmenorrhea over the first two periods to illustrate the use of these test procedures.

2 Notation, assumption and methods

Consider comparing two experimental treatments A and B with a placebo P under an incomplete block crossover design with two periods. We let X-Y denote the treatment-receipt sequence of receiving treatment X at period 1 and then crossover to receive treatment Y at period 2. Suppose that we randomly assign n_g patients to group $g(= 1, 2, 3, 4, 5, 6)$, where $g = 1$ denotes the group with P-A treatment-receipt sequence; $g = 2$ denotes the group with A-P treatment-receipt sequence; $g = 3$ denotes the group with P-B treatment-receipt sequence; $g = 4$ denotes the group with B-P treatment-receipt sequence; $g = 5$ denotes the group with A-B treatment-receipt sequence; and $g = 6$ denotes the group with B-A treatment-receipt sequence. For patient $i(= 1, 2, \ldots, n_g)$ assigned to group $g(= 1, 2, 3, 4, 5, 6)$, we let $Y_{iz}^{(g)}$ denote the ordinal outcome of the patient at period $z(= 1, 2)$, and take one of possible ordinal values C_j, where $C_1 < C_2 < C_3 < \cdots < C_L$. We let $X_{iz1}^{(g)}$ denote the indicator function of treatment-receipt for treatment A, and $X_{iz1}^{(g)} = 1$ for patient i assigned to group g at period z receiving treatment A, and = 0, otherwise. Similarly, we let $X_{iz2}^{(g)}$ denote the indicator function of treatment-receipt for treatment B, and $X_{iz2}^{(g)} = 1$ for the corresponding patient at period z receiving treatment B, and = 0, otherwise. We let $1_{iz}^{(g)}$ represent the indicator function of period, setting $1_{iz}^{(g)} = 1$ for period $z = 2$, and = 0, otherwise. We further let $\mu_i^{(g)}$ denote the random effect due to the ith subject in group g, and assume $\mu_i^{(g)}$'s to independently follow an unspecified probability density $f_g(\mu)$. We assume that one can apply an adequate wash-out period on the basis of our subjective knowledge to nullify the carry-over effect. As noted elsewhere [1, 4, 11, 12, 25–27], if we cannot ensure this assumption to hold, we may not wish to employ the crossover design. For patient $i(= 1, 2, \ldots, n_g)$ in group $g(= 1, 2, 3, 4, 5, 6)$, we assume that the joint conditional probability of $(Y_{i1}^{(g)}, Y_{i2}^{(g)})$ between periods 1 and 2, given the random effect $\mu_i^{(g)}$ fixed, satisfies

$$\begin{aligned}
&P(Y_{i1}^{(g)} < Y_{i2}^{(g)} | \mu_i^{(g)}) = 1/(1 + \exp(\mu_i^{(g)} + \eta_{AP} X_{i11}^{(g)} + \eta_{BP} X_{i12}^{(g)} + \gamma 1_{i1}^{(g)})) \\
&\times \exp(\mu_i^{(g)} + \eta_{AP} X_{i21}^{(g)} + \eta_{BP} X_{i22}^{(g)} + \gamma 1_{i2}^{(g)})/(1 + \exp(\mu_i^{(g)} + \eta_{AP} X_{i21}^{(g)} + \eta_{BP} X_{i22}^{(g)} + \gamma 1_{i2}^{(g)})), \\
&P(Y_{i1}^{(g)} > Y_{i2}^{(g)} | \mu_i^{(g)}) = 1/(1 + \exp(\mu_i^{(g)} + \eta_{AP} X_{i21}^{(g)} + \eta_{BP} X_{i22}^{(g)} + \gamma 1_{i2}^{(g)})) \\
&\times \exp(\mu_i^{(g)} + \eta_{AP} X_{i11}^{(g)} + \eta_{BP} X_{i12}^{(g)} + \gamma 1_{i1}^{(g)})/(1 + \exp(\mu_i^{(g)} + \eta_{AP} X_{i11}^{(g)} + \eta_{BP} X_{i12}^{(g)} + \gamma 1_{i1}^{(g)})) \\
&\text{and} \quad P(Y_{i1}^{(g)} = Y_{i2}^{(g)} | \mu_i^{(g)}) = 1 - P(Y_{i1}^{(g)} > Y_{i2}^{(g)} | \mu_i^{(g)}) - P(Y_{i1}^{(g)} < Y_{i2}^{(g)} | \mu_i^{(g)}),
\end{aligned} \tag{1}$$

where η_{AP} and η_{BP} denote the respective effect of treatments A and B relative to placebo P, as well as γ denotes the effect of period 2 versus period 1. Based on model (1), the GOR of responses [20–22] on a given patient i in group g when he/she has covariates $(X_{i21}^{(g)}, X_{i22}^{(g)}, 1_{i2}^{(g)})$ at period 2 versus when he/she has covariates $(X_{i11}^{(g)}, X_{i12}^{(g)}, 1_{i1}^{(g)})$ at period 1 is, by definition, equal to

$$\begin{aligned}
&P(Y_{i1}^{(g)} < Y_{i2}^{(g)} | \mu_i^{(g)})/P(Y_{i1}^{(g)} > Y_{i2}^{(g)} | \mu_i^{(g)}) = \\
&\exp(\eta_{AP}(X_{i21}^{(g)} - X_{i11}^{(g)}) + \eta_{BP}(X_{i22}^{(g)} - X_{i12}^{(g)}) + \gamma(1_{i2}^{(g)} - 1_{i1}^{(g)})).
\end{aligned} \tag{2}$$

When $\eta_{AP} = 0$, we can see from eq. (2) that the GOR of responses remains unchanged despite of receiving treatment A or placebo P. When $\eta_{AP} > 0$, taking treatment A tends to increase the patient response as compared with taking placebo P, given all the other covariates fixed. When $\eta_{AP} < 0$, taking treatment A tends to decrease the patient response as compared with taking placebo P. Similar interpretations of η_{AP} are applied to parameters η_{BP} and γ. We define the GOR of responses for treatment A versus placebo P and that for treatment B versus placebo P as $GOR_{AP} = \exp(\eta_{AP})$ and $GOR_{BP} = \exp(\eta_{BP})$, respectively. Also, we define the GOR of responses for treatment B versus treatment A as $GOR_{BA} = \exp(\eta_{BP} - \eta_{AP})$.

On the basis of model (1), for a randomly selected patient i from group g the probability that the patient response $Y_{i1}^{(g)}$ at period 1 is less than his/her response $Y_{i2}^{(g)}$ at period 2 is

$$P(Y_{i1}^{(g)} < Y_{i2}^{(g)}) = \int 1/(1+\exp(\mu+\eta_{AP}X_{i11}^{(g)}+\eta_{BP}X_{i12}^{(g)}+\gamma 1_{i1}^{(g)})) \times \exp(\mu+\eta_{AP}X_{i21}^{(g)}+\eta_{BP}X_{i22}^{(g)}+\gamma 1_{i2}^{(g)})/(1+\exp(\mu+\eta_{AP}X_{i21}^{(g)}+\eta_{BP}X_{i22}^{(g)}+\gamma 1_{i2}^{(g)}))f_g(\mu)d\mu. \quad (3)$$

Similarly, for a randomly selected patient i from group g the probability that the patient response $Y_{i1}^{(g)}$ at period 1 is larger than his/her response $Y_{i2}^{(g)}$ at period 2 is

$$P(Y_{i1}^{(g)} > Y_{i2}^{(g)}) = \int 1/(1+\exp(\mu+\eta_{AP}X_{i21}^{(g)}+\eta_{BP}X_{i22}^{(g)}+\gamma 1_{i2}^{(g)})) \times \exp(\mu+\eta_{AP}X_{i11}^{(g)}+\eta_{BP}X_{i12}^{(g)}+\gamma 1_{i1}^{(g)})/(1+\exp(\mu+\eta_{AP}X_{i11}^{(g)}+\eta_{BP}X_{i12}^{(g)}+\gamma 1_{i1}^{(g)}))f_g(\mu)d\mu. \quad (4)$$

Note that because we do not assume any parametric *p.d.f.* for $f_g(\mu)$ in the following discussion, our approach is semi-parametric.

For simplicity in notation, we define $\Pi_C^{(g)} = P(Y_{i1}^{(g)} < Y_{i2}^{(g)})$ and $\Pi_D^{(g)} = P(Y_{i1}^{(g)} > Y_{i2}^{(g)})$. From eqs (3) and (4), we can see that for a randomly selected patient i from group g the GOR of patient responses between periods 2 and 1 is

$$GOR^{(g)} = \Pi_C^{(g)}/\Pi_D^{(g)} = \exp(\eta_{AP}(X_{i21}^{(g)} - X_{i11}^{(g)}) + \eta_{BP}(X_{i22}^{(g)} - X_{i12}^{(g)}) + \gamma(1_{i2}^{(g)} - 1_{i1}^{(g)})). \quad (5)$$

We denote for a randomly selected patient i from group g the probability $P(Y_{i1}^{(g)} = C_r, Y_{i2}^{(g)} = C_s)$ by $\pi_{rs}^{(g)}$, where $r = 1, 2, \ldots, L$, and $s = 1, 2, \ldots, L$. Thus, we have

$$\Pi_C^{(g)} = P(Y_{i1}^{(g)} < Y_{i2}^{(g)}) = \sum_{r=1}^{L-1}\sum_{s=r+1}^{L}\pi_{rs}^{(g)}, \quad \text{and} \quad \Pi_D^{(g)} = P(Y_{i1}^{(g)} > Y_{i2}^{(g)}) = \sum_{r=2}^{L}\sum_{s=1}^{r-1}\pi_{rs}^{(g)}. \quad (6)$$

These represent the probability that a randomly selected patient i from group g has the response at period 2 higher than his/her response at period 1, and the probability that a randomly selected patient has the response at period 1 higher than his/her response at period 2, respectively. When $L = 2$, $GOR^{(g)}$ reduces to $\pi_{12}^{(g)}/\pi_{21}^{(g)}$, the OR of responses between periods 2 and 1 in binary data with matched-pairs. On the basis of model (5), we can express the GOR of responses for treatment A versus placebo as (**Appendix I**)

$$GOR_{AP} = \exp(\eta_{AP}) = (GOR^{(1)}/GOR^{(2)})^{1/2} = GOR^{(3)}/GOR^{(5)} = GOR^{(6)}/GOR^{(4)}. \quad (7)$$

Let $n_{rs}^{(g)}$ denote the number of patients in group g (= 1, 2, 3, 4, 5, 6) with the vector of patient responses $(Y_{i1}^{(g)} = C_r, Y_{i2}^{(g)} = C_s)$ among n_g patients. The random cell frequencies $\{n_{rs}^{(g)} \mid r = 1, 2, 3, \cdots, L, s = 1, 2, 3, \cdots, L\}$ then follow the multinomial distribution with parameters n_g and $\{\pi_{rs}^{(g)} \mid r = 1, 2, 3, \cdots, L, s = 1, 2, 3, \cdots, L\}$. Note that we can estimate $\pi_{rs}^{(g)}$ by the unbiased consistent estimator $\hat{\pi}_{rs}^{(g)} = n_{rs}^{(g)}/n_g$. We define $n_C^{(g)} = \sum_{r=1}^{L-1}\sum_{s=r+1}^{L} n_{rs}^{(g)}$ and $n_D^{(g)} = \sum_{r=2}^{L}\sum_{s=1}^{r-1} n_{rs}^{(g)}$. When substituting $\hat{\pi}_{rs}^{(g)}$ for $\pi_{rs}^{(g)}$ in $\Pi_C^{(g)}$ and $\Pi_D^{(g)}$, we obtain $\hat{\Pi}_C^{(g)} = n_C^{(g)}/n_g$ and $\hat{\Pi}_D^{(g)} = n_D^{(g)}/n_g$. These lead us to obtain the estimator $\widehat{GOR}^{(g)} = \hat{\Pi}_C^{(g)}/\hat{\Pi}_D^{(g)}$. Using the delta method [22, 28], we obtain the estimated asymptotic variance $\widehat{Var}(\log(\widehat{GOR}^{(g)})) = (\hat{\Pi}_C^{(g)} + \hat{\Pi}_D^{(g)})/(n_g\hat{\Pi}_C^{(g)}\hat{\Pi}_D^{(g)})$. When substituting $\widehat{GOR}^{(g)}$ for $GOR^{(g)}$ in eq. (7), we obtain the following three consistent estimators for $GOR_{AP}(= \exp(\eta_{AP}))$ as

$$\hat{GOR}_{AP} = (\hat{GOR}^{(1)}/\hat{GOR}^{(2)})^{1/2} = [(n_C^{(1)}n_D^{(2)})/(n_D^{(1)}n_C^{(2)})]^{1/2} = \hat{GOR}^{(3)}/\hat{GOR}^{(5)} = (n_C^{(3)}n_D^{(5)})/(n_D^{(3)}n_C^{(5)}) = \hat{GOR}^{(6)}/\hat{GOR}^{(4)} = (n_C^{(6)}n_D^{(4)})/(n_D^{(6)}n_C^{(4)}). \quad (8)$$

For convenience, we define three 2×2 tables consisting of cell frequencies $(f_{11k}, f_{12k}, f_{21k}, f_{22k})$(for $k = 1, 2, 3$) corresponding to eq. (8) as

$$(f_{111} = n_C^{(1)}, f_{121} = n_C^{(2)}, f_{211} = n_D^{(1)}, f_{221} = n_D^{(2)}),$$

$$(f_{112} = n_C^{(3)}, f_{122} = n_C^{(5)}, f_{212} = n_D^{(3)}, f_{222} = n_D^{(5)}),$$

and

$$(f_{113} = n_C^{(6)}, f_{123} = n_C^{(4)}, f_{213} = n_D^{(6)}, f_{223} = n_D^{(4)}). \tag{9}$$

When testing $H_0 : GOR_{AP} = 1$ versus $H_a : GOR_{AP} \neq 1$, we first consider use of the WLS summary test procedure [23] based on eq. (9). We will reject $H_0 : GOR_{AP} = 1$at the α-level if

$$(\sum_{k=1}^{3} W_k LGOR_{AP}^{(k)} / \sum_{k=1}^{3} W_k)^2 (\sum_{k=1}^{3} W_k) > Z_{\alpha/2}^2 \tag{10}$$

where

$$W_1 = 4/(1/f_{111} + 1/f_{121} + 1/f_{211} + 1/f_{221}),$$

$$W_k = 1/(1/f_{11k} + 1/f_{12k} + 1/f_{21k} + 1/f_{22k})$$

for $k = 2, 3$, $\widehat{LGOR}_{AP}^{(1)} = \log((f_{111}f_{221})/(f_{121}f_{211}))^{1/2}$, $\widehat{LGOR}_{AP}^{(k)} = \log((f_{11k}f_{22k})/(f_{12k}f_{21k}))$for $k = 2, 3$, and Z_α is the upper $100(\alpha)$th percentile of the standard normal distribution. Note that if f_{ijk}= 0 for some observed frequencies in a 2×2 table k, we cannot employ the test procedure (10). We may apply the commonly-used ad hoc arbitrary adjustment for sparse data by adding 0.50 to each observed frequency f_{ijk} in this particular table k.

When the observed marginal frequencies $n_C^{(g)}$ and $n_D^{(g)}$ are not large, the WLS test procedure may lose accuracy because the weights $W_k$in eq. (10) can be subject to a large variation. This may lead us to consider use of the MH summary test procedure [28, 29]. When comparing treatment A with placebo, we will reject the null hypothesis $H_0 : GOR_{AP} = 1$ at the α-level if the test statistic

$$\begin{aligned}&(\textstyle\sum_k f_{11k} - \sum_k f_{1+k}f_{+1k}/f_{++k})^2 /\\ &\{\textstyle\sum_k f_{1+k}f_{2+k}f_{+1k}f_{+2k}/[f_{++k}^2(f_{++k} - 1)]\} > Z_{\alpha/2}^2.\end{aligned} \tag{11}$$

When both the numbers of patients $n_C^{(g)}$ and $n_D^{(g)}$ are small, the asymptotic WLS and MH test procedures may lose accuracy. Thus, we may consider use of the following exact test procedure. Define $n_{dis}^{(g)} = n_C^{(g)} + n_D^{(g)}$. Given $n_{dis}^{(g)}$ fixed, we can show that $n_C^{(g)}$ follows the binomial distribution with parameters $n_{dis}^{(g)}$ and $\Pi_C^{(g)}/(\Pi_C^{(g)} + \Pi_D^{(g)})(= GOR^{(g)}/(1+GOR^{(g)}))$. Under$GOR_{AP} = 1$, the conditional probability distribution of f_{11k}, given $f_{+1k}, f_{+2k}, f_{1+k}$and f_{2+k} fixed, is given by the hypergeometric distribution:

$$P(f_{11k}|f_{+1k}, f_{+2k}, f_{1+k}, f_{2+k}) = \frac{\binom{f_{+1k}}{f_{11k}}\binom{f_{+2k}}{f_{1+k} - f_{11k}}}{\binom{f_{++k}}{f_{1+k}}} \tag{12}$$

where $a_k \leq f_{11k} \leq b_k$, $a_k = \max\{0, f_{1+k} - f_{+2k}\}$and $b_k = \min\{f_{+1k}, f_{1+k}\}$ for $k = 1, 2, 3$.

Thus, the joint conditional probability distribution off_{111}, f_{112}andf_{113} is simply [30]

$$P(f_{-11}|f_{-+1}, f_{-+2}, f_{-1+}, f_{-2+}) = \prod_{k=1}^{3} \frac{\binom{f_{+1k}}{f_{11k}}\binom{f_{+2k}}{f_{1+k} - f_{11k}}}{\binom{f_{++k}}{f_{1+k}}}, \tag{13}$$

where $\underline{f}_{-11} = (f_{111}, f_{112}, f_{113})'$, $\underline{f}_{-u+} = (f_{u+1}, f_{u+2}, f_{u+3})'$ for u =1, 2, and $\underline{f}_{-+v} = (f_{+v1}, f_{+v2}, f_{+v3})'$ for v =1, 2. Given an observed value $\underline{f}^o_{-11} = (f^o_{111}, f^o_{112}, f^o_{113})'$, if the following p-value, calculated as [23, 30]

$$\sum_{\underline{f}_{11} \in C} \prod_{k=1}^{3} \frac{\begin{pmatrix} f_{+1k} \\ f_{11k} \end{pmatrix} \begin{pmatrix} f_{+2k} \\ f_{1+k} - f_{11k} \end{pmatrix}}{\begin{pmatrix} f_{++k} \\ f_{1+k} \end{pmatrix}} \tag{14}$$

where $C = \left\{ \underline{f}_{-11} \middle| P(\underline{f}_{-11} \middle| \underline{f}_{-+1}, \underline{f}_{-+2}, \underline{f}_{-1+}, \underline{f}_{-2+}) \le P\left(\underline{f}^o_{-11} | \underline{f}_{-+1}, \underline{f}_{-+2}, \underline{f}_{-1+}, \underline{f}_{-2+} \right) \right\}$, is less than a small given α-level, we will reject $H_0 : GOR_{AP} = 1$. Note that the exact test (14) is actually a direct extension of Fisher's exact test to a series of 2×2 tables [23, 28, 30] with modifications to accommodate ordinal responses.

As shown in **Appendix I**, we can easily modify asymptotic test procedures (10) and (11) and the exact test procedure (14) to account for testing$H_0 : GOR_{BP} = 1$ (or $H_0 : GOR_{BA} = 1$) with replacing f_{ijk} by f^*_{ijk} (or f^{**}_{ijk}), where f^*_{ijk} and f^{**}_{ijk} are defined in (26) and (28) (**Appendix I**), respectively.

3 Monte Carlo simulation

To evaluate and compare the performance of the WLS, MH and exact procedures for testing equality between treatments, we employ Monte Carlo simulation. By use of the conditional arguments, we do not need to estimate the nuisance period effect γ in use of these test procedures. We arbitrarily set γ equal to 0.10 in the simulation. Furthermore, our approach is valid for any assumed distribution for the random effects $\mu_i^{(g)}$. We consider the cases in which random effects $\mu_i^{(g)}$are independent and identically distributed (*i.i.d.*) as the normal distribution with mean 0 and standard deviation $\sigma = 0.5, 1$; as well as $\mu_i^{(g)}$are *i.i.d* from a gamma distribution with shape parameter $\alpha = x00BD$; and scale parameter $\beta = 1, 2$. We cover the situations in which the relative effect of treatment A versus the placebo, $\eta_{AP} = 0.0, 0.50$; the relative effect of treatment B versus treatment A, $\eta_{BA} = 0.0, 1.0$ (such that the relative effect of treatment B versus the placebo, $\eta_{BP} = \eta_{AP} + \eta_{BA}$); and the number of patients n ($= n_1 = n_2 = ... = n_6$) per group $n = 10, 15, 25$. Note that when $n = 10$, these include the cases in which the expected number of patients with discordant responses $n^{(g)}_{dis} (= n^{(g)}_C + n^{(g)}_D)$ is as small as approximate 3.71–4.84 patients. Note also that all test procedures proposed here depend on only the marginal totals $n^{(g)}_C$ and $n^{(g)}_D$instead of individual cell frequency $n^{(g)}_{rs}$. Thus, there is no need to consider the number of ordinal levels L, and so is $n^{(g)}_{rs}$. For each configuration determined by a combination of the above parameter values, we write programs in SAS [31] and generate 10,000 simulated samples of n patients per group, each having the bivariate responses $(Y^{(g)}_{i1}, Y^{(g)}_{i2})'$ with probability $P(Y^{(g)}_{i1} < Y^{(g)}_{i2} | \mu^{(g)}_i)$ and $P(Y^{(g)}_{i1} > Y^{(g)}_{i2} | \mu^{(g)}_i)$ given by model (1), to calculate the simulated Type I error and power at the 0.05-level for a given test procedure. Recall that the power function of a test procedure with rejection region is, by definition, the probability that the sample points fall into the rejection region. The power function will give Type I error when H_0 is true, and will give power when H_0 is false. Therefore, the simulated Type I error for a given test procedure can be calculated as the proportion of 10,000 simulated samples for which we reject H_0 when H_0 is true. Similarly, the simulated power for a given test procedure can be calculated as the proportion of 10,000 simulated samples for which we reject H_0 when H_0 is false. For readers' information, the SAS program for our simulation can be accessible at http://edoras.sdsu.edu/ kjl/exactso.htm.

4 Results

We summarize in Table 1 the estimated Type I error (in **boldface**) and power of using the WLS, MH and exact procedures for testing $H_0 : \eta_{AP} = 0$ and testing$H_0 : \eta_{BP} = 0$ at the 0.05-level when $\mu_i^{(g)}$are *i.i.d.* as the normal distribution with mean 0 and standard deviation $\sigma = 0.5, 1$. For example, when $\eta_{AP} = 0$ and $\eta_{BP} = 1.0$, the entries corresponding to procedures for testing $H_0 : \eta_{AP} = 0$ are Type I errors, while those corresponding to procedures for testing $H_0 : \eta_{BP} = 0$ are powers (Table 1). We can see that both the MH and exact tests can perform well, while the WLS test can be conservative especially when n is small (say, 10). We note that the MH test can be consistently of more power than the WLS and exact tests in almost all the situations considered in

Table 1. For example, when $\sigma = 0.50$, $\eta_{AP} = 0.50$, $\eta_{BP} = 1.50$ and $n = 15$, the powers for testing $H_0 : \eta_{AP} = 0$ and $H_0 : \eta_{BP} = 0$ are 0.175 and 0.825 for the MH test, while these powers are 0.124 and 0.781 for the WLS test, and are 0.132 and 0.731 for the exact test. We also note that the power for all test procedures increases as the number of subjects n increases, but decreases as the variation σ of responses between patients increases. Since all the findings on the performance of the WLS, MH and exact procedures with respect to Type I error, as well as the relative order of powers between these test procedures hold when $\mu_i^{(g)}$ are *i.i.d.* from the gamma distribution, we do not present these results for brevity. These results are, however, to readers upon request.

Table 1: The estimated Type I error (in **boldface**) and power of using the WLS, MH and Exact tests for testing $H_0 : \eta_{AP} = 0$ and testing $H_0 : \eta_{BP} = 0$ at the 0.05-level in situations in which the random effects $\mu_i^{(g)}$ are i.i.d. as the normal distribution with mean 0 and standard deviation $\sigma = 0.5, 1$; the relative effect of treatment A versus the placebo, $\eta_{AP} = 0.0, 0.50$; and the relative effect of treatment B versus treatment A, $\eta_{BA} = 0.0, 1.0$ (such that the relative effect of treatment B versus the placebo, $\eta_{BP} = \eta_{AP} + \eta_{BA}$), and the number of patients n ($= n_1 = n_2 = ... = n_6$) per group $n = 10, 15, 25$.

				Testing $H_0 : \eta_{AP} = 0$			**Testing** $H_0 : \eta_{BP} = 0$		
σ	η_{AP}	η_{BP}	n	WLS	MH	Exact	WLS	MH	Exact
0.5	0.00	0.00	10	**0.020**	**0.050**	**0.046**	**0.022**	**0.051**	**0.048**
			15	**0.028**	**0.048**	**0.049**	**0.031**	**0.051**	**0.049**
			25	**0.037**	**0.048**	**0.049**	**0.034**	**0.046**	**0.050**
		1.00	10	**0.016**	**0.047**	**0.040**	0.259	0.369	0.253
			15	**0.029**	**0.051**	**0.051**	0.458	0.520	0.401
			25	**0.034**	**0.048**	**0.047**	0.750	0.749	0.636
	0.50	0.50	10	0.073	0.129	0.089	0.073	0.131	0.090
			15	0.131	0.174	0.122	0.134	0.178	0.125
			25	0.251	0.269	0.195	0.255	0.271	0.199
		1.50	10	0.058	0.125	0.093	0.528	0.649	0.501
			15	0.124	0.175	0.132	0.781	0.825	0.731
			25	0.226	0.250	0.195	0.970	0.973	0.938
1.0	0.00	0.00	10	**0.018**	**0.052**	**0.043**	**0.017**	**0.046**	**0.044**
			15	**0.025**		**0.048**	**0.030**	**0.053**	**0.053**
			25	**0.036**	**0.051**	**0.051**	**0.033**	**0.047**	**0.049**
		1.00	10	**0.013**	**0.050**	**0.037**	0.208	0.330	0.222
			15	**0.021**	**0.049**	**0.047**	0.401	0.477	0.361
			25	**0.035**	**0.052**	**0.047**	0.681	0.697	0.583
	0.50	0.50	10	0.054	0.117	0.079	0.059	0.121	0.083
			15	0.114	0.159	0.112	0.115	0.160	0.109
			25	0.219	0.247	0.169	0.213	0.240	0.169
		1.50	10	0.043	0.112	0.074	0.457	0.597	0.447
			15	0.093	0.150	0.114	0.732	0.789	0.672
			25	0.198	0.230	0.174	0.947	0.951	0.904

5 An example

Consider the data (Table 2) taken as a part of a crossover trial comparing an analgesic at low (L) and high (H) doses with a placebo (P) for the relief of pain in primary dysmenorrhea patients over the first two-periods [24]. Here, we refer the low and high doses as treatments A and B. There were 86 patients randomly assigned to the six groups: P-L ($g = 1$); L-P ($g = 2$); P-H ($g = 3$); H-P ($g = 4$); L-H ($g = 5$); and H-L ($g = 6$). At the end of each treatment period, each patient was assessed the extent of relief on the ordinal scale: none (coded as 1), moderate (coded as 2) and complete (coded as 3). When applying the WLS and MH procedures, as well as the exact procedure to test $H_0 : GOR_{AP} = 1$, we obtain p-values 0.0119, 0.0013 and 0.0002. Similarly, when applying these procedures with replacing f_{ijk} by f_{ijk}^* to test $H_0 : GOR_{BP} = 1$, we obtain all p-values as 0.0017, 0.0000 and 0.0000. Thus, there is evidence that taking either low or high dose of the analgesic can help the relief of pain in primary dysmenorrhea as compared with the placebo. Furthermore, when applying the corresponding test procedures with replacing f_{ijk} by f_{ijk}^{**} to test $H_0 : GOR_{BA} = 1$, we obtain p-values 0.7994, 0.9986, and 1.0000. There is no evidence that taking the high dose can help, as compared with the low dose, the relief of pain among these patients with primary dysmenorrhea.

Table 2: The frequency of patients for the relief of pain (1: none or minimal; 2: moderate; 3: complete) in primary dysmenorrhea at the first two periods versus the groups determined by the treatment-receipt sequence (P: placebo; L: low dose; H: high dose).

	Group of Treatment-Receipt Sequence					
g =	**1**	**2**	**3**	**4**	**5**	**6**
Responses	**P-L**	**L-P**	**P-H**	**H-P**	**L-H**	**H-L**
(1,1)	2	1	2	3	1	1
(1,2)	9	2	3	2	0	0
(1,3)	2	2	6	0	1	1
(2,1)	0	4	1	1	0	4
(2,2)	2	1	2	1	6	1
(2,3)	0	0	2	1	1	2
(3,1)	0	5	0	6	1	4
(3,2)	0	0	0	0	2	0
(3,3)	0	0	0	0	0	1
n_g=	15	15	16	14	12	14

Note that if we assume the proportional odds model using cumulative logit [28, 31] with normal random effects [19] due to patients, we can apply Proc GLIMMIX [31] to analyze the data in Table 2. When employing this SAS procedure (http://edoras.sdsu.edu/ kjl/exaordex1.htm), we obtain the parameter estimates (and their estimated standard error (SD)) of the relative effect for the low dose versus the placebo, the high dose versus the placebo, and the high dose versus the low dose are: –1.7656 (SD = 0.3898), –2.3404 (SD = 0.4011), and –0.5748 (SD = 0.3537). On the basis of these estimates (of which the signs are all < 0), we may conclude that both the low dose and high dose can significantly improve, as compared with the placebo, the relief of pain at the 5 % level; both p-values are < 0.0001. Furthermore, though the high dose can improve the outcome of patients as compared with the low dose, this improvement is not significant at the 5 %-level. All the above results are essentially similar to those reported previously for the complete block crossover design over a three-period trial [32].

6 Discussion

We do not recommend, as noted previously [1, 4, 11, 12, 25–27], use of the crossover design if we cannot ensure ourselves to nullify the carry-over effects with an adequate wash-out period. On the other hand, if there are carry-over effects due to earlier treatments, we note that the test procedures proposed here can still be valid for use under the simple carry-over model (**Appendix II**). Also, we note that although one may apply the estimator as given in **Appendix II** for the difference in carry-over effects to test whether there are differential carry-over effects, we do not recommend using this test to determine whether the assumption of no carry-over effects holds. This is because the concerns raised by Freeman [33] and Senn [34] for using the two-stage test procedure suggested by Grizzle [2].

When employing the test procedures developed here, we do not need to assume any parametric distribution for the random effects due to patients. Thus, our procedures is, as noted before, semi-parametric. Also, the number of patients for a crossover trial is often small. The exact test procedure (14) can be of use in practice. By contrast, one needs to assume the random effects due to patients independently follow the normal distribution in use of Proc GLIMMIX [35]. Also, note that the proportional odds model can be badly violated by many bivariate distributions [20, 36, 37]. Furthermore, note that Wald's test can be invalid for use if the number of patients in a trial is small and the data are sparse.

We note that when the number of subjects per group n is small (say, 10), use of the WLS procedure can be conservative, while the MH and exact test procedure can perform well (Table 1). We further note that the MH test procedure is generally of more power than the other two procedures in the situations considered here. Because use of the MH procedure does not involve any sophisticated numerical procedure and can be calculated by a hand calculator, we may recommend the MH test procedure for general use when n is not large. We may use of the exact test procedure if one has the concern of normal approximation for a small n. When the number of subjects n per group is large (say, 40), however, we want to note that the WLS procedure can be of more power than the MH and exact procedures on the basis of Monte Carlo simulation (not shown here).

If we wish to study the relative period effect, we may apply similar ideas as above to derive the corresponding procedures for testing $H_0 : \gamma = 0$. For example, we can easily see from eqs (16)–(21) (in **Appendix I)** that the GOR of responses for period 2 versus period 1 is

$$GOR_{21} = \exp(\gamma) = (GOR^{(1)}GOR^{(2)})^{1/2} = (GOR^{(3)}GOR^{(4)})^{1/2} = (GOR^{(5)}GOR^{(6)})^{1/2}. \quad (15)$$

Following similar arguments as for comparing the treatment effect, we can derive on the basis of (15) the WLS, MH and exact procedures for testing$H_0 : GOR_{21} = 1$. Finally, note that when the patient response is dichotomous, the GOR reduces to the common OR for paired sample data. Thus, the MH and exact test procedures include those for testing equality of treatments in binary data [18] under the random effects logistic regression model as special cases.

In summary, we have derived the WLS, MH and exact procedures for testing equality between treatments under an incomplete block crossover design with ordinal responses. We have evaluated and compare their performance in a variety of situations based on Monte Carlo simulation. We have noted that the proposed test procedures are valid for use in the presence of simple carry-over effects. We have compared the proposed test procedures with use of Wald's test procedures assuming the normal random effects proportional odds model. We have noted that the proposed test procedures include those for testing equality of treatments in binary data as special cases. The results, findings and discussions should have use for biostatisticians and clinicians when they employ a two-period crossover design to compare three treatments in ordinal data.

Acknowledgements

The author wishes to thank the reviewer for valuable and useful comments and suggestions to improve the contents and clarity of this article.

Appendix I

On the basis of the assumed model $GOR^{(g)}(5)$, we may obtain the GOR of responses between periods 2 and 1 in group g (= 1, 2, 3, 4, 5, 6) as follows:

$$GOR^{(1)} = \exp(\eta_{AP} + \gamma), \quad (16)$$

$$GOR^{(2)} = \exp(-\eta_{AP} + \gamma), \quad (17)$$

$$GOR^{(3)} = \exp(\eta_{BP} + \gamma), \quad (18)$$

$$GOR^{(4)} = \exp(-\eta_{BP} + \gamma), \quad (19)$$

$$GOR^{(5)} = \exp(\eta_{BP} - \eta_{AP} + \gamma), \quad (20)$$

$$GOR^{(6)} = \exp(-\eta_{BP} + \eta_{AP} + \gamma). \quad (21)$$

On the basis of eqs (16) and (17), we have

$$GOR_{AP} = \exp(\eta_{AP}) = (GOR^{(1)}/GOR^{(2)})^{1/2}, \quad (22)$$

which is free from the period effect. Similarly, we can see from eqs (18) and (20) that

$$GOR_{AP} = GOR^{(3)}/GOR^{(5)}. \quad (23)$$

Furthermore, we can see from eqs (21) and (19) that

$$GOR_{AP} = GOR^{(6)}/GOR^{(4)}. \tag{24}$$

Following similar arguments as for deriving eqs (22)–(24), we obtain the following three consistent estimators for the GOR of responses between treatment B and placebo P as

$$\begin{aligned} \hat{GOR}_{BP} &= (\hat{GOR}^{(3)}/\hat{GOR}^{(4)})^{1/2} \\ &= \hat{GOR}^{(1)}/\hat{GOR}^{(6)} \\ &= \hat{GOR}^{(5)}/\hat{GOR}^{(2)}. \end{aligned} \tag{25}$$

Again, for convenience in the following discussion we define three 2×2 tables consisting of frequencies $(f^*_{11k}, f^*_{12k}, f^*_{21k}, f^*_{22k})$(for $k = 1, 2, 3$) corresponding to equation (25) as

$$\begin{aligned} &(f^*_{111} = n^{(3)}_C, f^*_{121} = n^{(4)}_C, f^*_{211} = n^{(3)}_D, f^*_{221} = n^{(4)}_D), \\ &(f^*_{112} = n^{(1)}_C, f^*_{122} = n^{(6)}_C, f^*_{212} = n^{(1)}_D, f^*_{222} = n^{(6)}_D), \end{aligned}$$

and

$$(f^*_{113} = n^{(5)}_C, f^*_{123} = n^{(2)}_C, f^*_{213} = n^{(5)}_D, f^*_{223} = n^{(2)}_D). \tag{26}$$

Again, using the same arguments as above, we may further obtain the following three consistent estimators for $GOR_{BA}(= \exp(\eta_{BP} - \eta_{AP}))$ as

$$\begin{aligned} \hat{GOR}_{BA} &= (\hat{GOR}^{(5)}/\hat{GOR}^{(6)})^{1/2} \\ &= \hat{GOR}^{(3)}/\hat{GOR}^{(1)} \\ &= \hat{GOR}^{(2)}/\hat{GOR}^{(4)}. \end{aligned} \tag{27}$$

Also, we define the following three 2×2 tables consisting of frequencies $(f^{**}_{11k}, f^{**}_{12k}, f^{**}_{21k}, f^{**}_{22k})$(for $k = 1, 2, 3$) corresponding to equation (27) as

$$\begin{aligned} &(f^{**}_{111} = n^{(5)}_C, f^{**}_{121} = n^{(6)}_C, f^{**}_{211} = n^{(5)}_D, f^{**}_{221} = n^{(6)}_D), \\ &(f^{**}_{112} = n^{(3)}_C, f^{**}_{122} = n^{(1)}_C, f^{**}_{212} = n^{(3)}_D, f^{**}_{222} = n^{(1)}_D), \end{aligned}$$

and

$$(f^{**}_{113} = n^{(2)}_C, f^{**}_{123} = n^{(4)}_C, f^{**}_{213} = n^{(2)}_D, f^{**}_{223} = n^{(4)}_D). \tag{28}$$

Appendix II

Using the simple carryover model, we assume that

$$P(Y^{(g)}_{i1} < Y^{(g)}_{i2} | \mu^{(g)}_i) = 1/(1 + \exp(\mu^{(g)}_i + \eta_{AP} X^{(g)}_{i11} + \eta_{BP} X^{(g)}_{i12} + \gamma 1^{(g)}_{i1}))$$

$$\times [\exp(\mu^{(g)}_i + \eta_{AP} X^{(g)}_{i21} + \eta_{BP} X^{(g)}_{i22} + \gamma 1^{(g)}_{i2} + \rho_P 1_i(g = 1, 3) + \rho_A 1_i(g = 2, 5) + \rho_B 1_i(g = 4, 6))$$

$$/(1 + \exp(\mu^{(g)}_i + \eta_{AP} X^{(g)}_{i21} + \eta_{BP} X^{(g)}_{i22} + \gamma 1^{(g)}_{i2} + \rho_P 1_i(g = 1, 3) + \rho_A 1_i(g = 2, 5) + \rho_B 1_i(g = 4, 6))],$$

$$P(Y^{(g)}_{i1} > Y^{(g)}_{i2} | \mu^{(g)}_i) =$$

$$[1/(1+\exp(\mu_i^{(g)}+\eta_{AP}X_{i21}^{(g)}+\eta_{BP}X_{i22}^{(g)}+\gamma 1_{i2}^{(g)}+\rho_P 1_i(g=1,3)+\rho_A 1_i(g=2,5)+\rho_B 1_i(g=4,6)))]$$

$$\times\exp(\mu_i^{(g)}+\eta_{AP}X_{i11}^{(g)}+\eta_{BP}X_{i12}^{(g)}+\gamma 1_{i1}^{(g)})/(1+\exp(\mu_i^{(g)}+\eta_{AP}X_{i11}^{(g)}+\eta_{BP}X_{i12}^{(g)}+\gamma 1_{i1}^{(g)}))$$

and

$$P(Y_{i1}^{(g)}=Y_{i2}^{(g)}\Big|\mu_i^{(g)})=1-P(Y_{i1}^{(g)}>Y_{i2}^{(g)}|\mu_i^{(g)})-P(Y_{i1}^{(g)}<Y_{i2}^{(g)}\Big|\mu_i^{(g)}), \tag{29}$$

where $1_i(g=g_1,g_2)=1$ for $g=g_1,g_2$ at period z = 2, and = 0, otherwise; as well as ρ_P, ρ_A, and ρ_B represent the carry-over effect due to placebo, treatment A and treatment B, respectively.

On the basis of the assumed model (29), we obtain

$$P(Y_{i1}^{(g)}<Y_{i2}^{(g)}|\mu_i^{(g)})/P(Y_{i1}^{(g)}>Y_{i2}^{(g)}|\mu_i^{(g)})=$$

$$\exp(\eta_{AP}(X_{i21}^{(g)}-X_{i11}^{(g)})+\eta_{BP}(X_{i22}^{(g)}-X_{i12}^{(g)})+\gamma(1_{i2}^{(g)}-1_{i1}^{(g)})+\rho_P 1_i(g=1,3)+\rho_A 1_i(g=2,5)+\rho_B 1_i(g=4,6)). \tag{30}$$

On the basis of model (30), we may obtain the GOR of responses between periods 2 and 1 in group g (= 1, 2, 3, 4, 5, 6) as

$$GOR^{(1)}=\exp(\eta_{AP}+\gamma+\rho_P), \tag{31}$$

$$GOR^{(2)}=\exp(-\eta_{AP}+\gamma+\rho_A), \tag{32}$$

$$GOR^{(3)}=\exp(\eta_{BP}+\gamma+\rho_P), \tag{33}$$

$$GOR^{(4)}=\exp(-\eta_{BP}+\gamma+\rho_B), \tag{34}$$

$$GOR^{(5)}=\exp(\eta_{BP}-\eta_{AP}+\gamma+\rho_A), \tag{35}$$

$$GOR^{(6)}=\exp(-\eta_{BP}+\eta_{AP}+\gamma+\rho_B). \tag{36}$$

From eqs (31) and (32), we have

$$GOR^{(1)}/GOR^{(2)}=\exp(2\eta_{AP}+\rho_P-\rho_A). \tag{37}$$

Similarly, from eqs (33)–(36), we obtain

$$GOR^{(3)}/GOR^{(5)}=\exp(\eta_{AP}+\rho_P-\rho_A), \tag{38}$$

and

$$GOR^{(6)}/GOR^{(4)}=\exp(\eta_{AP}). \tag{39}$$

Under $H_0: GOR_{AP}=1$ (*i.e.*, there is no difference in effects between treatment A and placebo), we can reasonably assume that the carry over effect ρ_A, due to treatment A and the carry-over effectρ_P due to placebo are equal. Thus, all procedures (10, 11) and (14) can still preserve Type I error. We can use the same arguments as noted here to account for the reason why the corresponding procedures for testing $H_0: GOR_{BP}=1$ or $H_0: GOR_{BA}=1$ can preserve Type I error as well.

We define $\rho_{PA}=\rho_P-\rho_A$. To estimate ρ_{PA}, we consider the following linear combination of estimators:

$$\hat{\rho}_{PA}=l_1\log(\widehat{GOR}^{(1)}/\widehat{GOR}^{(2)})+l_2\log(\widehat{GOR}^{(3)}/\widehat{GOR}^{(5)})+l_3\log(\widehat{GOR}^{(6)}/\widehat{GOR}^{(4)}), \tag{40}$$

where l_k are constants. We let V_1, V_2 and V_3 denote the variances $Var(\log(\widehat{GOR}^{(1)}/\widehat{GOR}^{(2)}))$, $Var(\log(\widehat{GOR}^{(3)}/\widehat{GOR}^{(5)}))$and $Var(\log(\widehat{GOR}^{(6)}/\widehat{GOR}^{(4)}))$, respectively. On the basis of eqs (37)–(39), we want to find constants $l_k$to minimize the variance $Var(\hat{\rho}_{PA}) = l_1^2V_1 + l_2^2V_2 + l_3^2V_3$, while these constants $l_k$are subject to constraints $2l_1 + l_2 + l_3 = 0$ and $l_1 + l_2 = 1$ so that $\hat{\rho}_{PA}$ is a consistent estimator for$\rho_{PA}(= \rho_P - \rho_A)$. Using Lagrange multiplier, we obtain

$$l_1 = (V_2 - V_3)/(V_1 + V_2 + V_3)$$

,

$$l_2 = (V_1 + 2V_3)/(V_1 + V_2 + V_3), \quad \text{and}$$

$$l_3 = -(V_1 + 2V_2)/(V_1 + V_2 + V_3). \tag{41}$$

Note that the weights (41) are function of unknown variances V_k. We may simply substitute $\hat{V}_k$ for V_k in (41) to obtain the estimated optimal weights $\hat{l}_k$, where $\hat{V}_k = (1/f_{11k} + 1/f_{12k} + 1/f_{21k} + 1/f_{22k})$ for $k = 1, 2, 3$. We can then employ this resulting consistent estimator for ρ_{PA} together with its estimated variance $\widehat{Var}(\hat{\rho}_{PA}) = \hat{l}_1^2\hat{V}_1 + \hat{l}_2^2\hat{V}_2 + \hat{l}_3^2\hat{V}_3$ to test $H_0 : \rho_{PA} = 0$ if one should decide to do so. Similar discussions as above can be done for studying $\rho_{PB} = \rho_P - \rho_B$ and $\rho_{AB} = \rho_A - \rho_B$.

References

[1] Senn S. Cross-over trials in clinical research, 2nd ed Chichester: Wiley, 2002.
[2] Grizzle JE. The two-period change-over design and its use in clinical trials. Biometrics. 1965;21:467–480.
[3] Hill M, Armitage P. The two-period cross-over clinical trial. Br J Clin Pharmacol. 1979;8:7–20.
[4] Fleiss JL. The design and analysis of clinical experiments. New York: Wiley, 1986.
[5] Lui K-J, Chang K-C. Hypothesis testing and estimation in ordinal data under a simple crossover design. J Biopharm Stat. 2012;22:1137–1147.
[6] Jones B, Kenward MG. Modelling binary data from a three-period cross-over trial. Stat Med. 1987;6:555–564.
[7] Jones B, Kenward MG. Design and analysis of cross-over trials, 3rd ed London: Chapman and Hall, Taylor & Francis Group, 2014.
[8] Senn S. Cross-over trials in Statistics in Medicine: The first '25' years. Stat Med. 2006;25:3430–3442.
[9] Gart JJ. An exact test for comparing matched proportions in crossover designs. Biometrika. 1969;56:75–80.
[10] Ezzet F, Whitehead J. A random effects model for binary data from cross-over trials. Appl Stat. 1992;41:117–126.
[11] Schouten H, Kester A. A simple analysis of a simple crossover trial with a dichotomous outcome measure. Stat Med. 2010;29:193–198.
[12] Lui K-J. Crossover designs: Testing, estimation, and sample size. New York: Wiley, 2016.
[13] Lui K-J, Chang K-C. Test non-inferiority (and equivalence) based on the odds ratio under a simple crossover trial. Stat Med. 2011;30:1230–1242.
[14] Lui K-J, Chang K-C. Exact sample size determination in testing non-inferiority under a simple crossover design. Pharm Stat. 2012;11:129–134.
[15] Lui K-J. Notes on testing carry-over effects in continuous data under an incomplete block crossover design. Commun Stat Simul Comput 2016. DOI:10.1080/03610918.2016.1146764.
[16] Lui K-J, Chang K-C. Test and estimation in binary data analysis under an incomplete block crossover design. Comput Stat Data Anal. 2015;81:130–138.
[17] Lui K-J. Estimation of the treatment effect under an incomplete block crossover design in binary data –a conditional likelihood approach. Stat Methods Med Res 2015. DOI:10.1177/0962280215595434.
[18] Lui K-J, Chang K-C. Exact tests in binary data under an incomplete block crossover design. Stat Methods Med Res 2016. DOI:10.1177/0962280216638382.
[19] Ezzet F, Whitehead J. A random effects model for ordinal responses from a crossover trial. Stat Med. 1991;10:901–907.
[20] Agresti A. generalized odds ratios for ordinal data. Biometrics. 1980;36:59–67.
[21] Lui K-J. Notes on estimation of the general odds ratio and the general risk difference for paired-sample data. Biometrical J. 2002;44:957–968.
[22] Lui K-J. Statistical estimation of epidemiology risk. Wiley: New York, 2004.
[23] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions, 3rd ed. New York: Wiley, 2003.
[24] Kenward M, Jones B. The analysis of categorical data from cross-over trials using a latent variable model. Stat Med. 1991;10:1607–1619.
[25] Senn SJ. Is the simple carry-over model useful?. Stat Med. 1992;11:715–726.
[26] Fleiss JL. On multiperiod crossover studies. Biometrics (Correspondence). 1986;42:449–450.
[27] Fleiss JL. A critique of recent research on the two-treatment crossover design. Control Clin Trials. 1989;10:237–243.
[28] Agresti A. Categorical data analysis. New York: Wiley, 1990.
[29] Ne B, Ne D. Statistical methods in cancer research, volume 1. The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980.

[30] Gart JJ. Point and interval estimation of the common odds ratio in the combination of 2x2 tables with fixed margins. Biometrika. 1970;57:471–475.
[31] SAS Institute, Inc. User's Guide, 2nd ed Cary, NC: SAS Institute, 2009.
[32] Lui K-J, Chang K-C, Lin C-D. Testing equality and interval estimation of the generalized odds ratio in ordinal data under a three-period crossover design. Stat Methods Med Res 2015. DOI:10.1177/0962280215569623.
[33] Freeman P. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. Stat Med. 1989;8:1421–1432.
[34] Senn SJ. Cross-over trials, carry-over effects and the art of self-delusion. Stat Med. 1988;7:1099–1101.
[35] Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. J R Stat Soc Ser. 1985;148:82–117.
[36] Fleiss JL. On the asserted invariance of the odds ratio. Br J Prev Soc Med. 1970;24:45–46.
[37] Mosteller F. Association and estimation in contingency tables. J Am Stat Assoc. 1968;63:1–28.