HOMOGENEITY TEST FOR CORRELATED DATA IN OPHTHALMOLOGIC STUDIES

CHANG-XING MA AND GUOGEN SHAN

ABSTRACT. In ophthalmologic studies, measurements obtained from both eyes of an individual are often highly correlated. Ignoring the correlation could lead to incorrect inferences. An asymptotic method was proposed by Tang et al. (2008) for testing equality of proportions between two groups. In this article, we investigate three testing procedures for general $g \geq 2$ groups. Our simulation results show the score testing procedure usually produces satisfactory type I error control and ~~having~~ has more power. Examples from ophthalmologic studies are used to illustrate our proposed methods.

1. INTRODUCTION

In randomized clinical trials [1], patients are usually randomized into two or more treatment groups, and patients within each group receive the same treatment. Often a control group or a group with standard treatment is included for testing the efficiency of new treatments. After the randomization, all patients are followed up in exactly the same way as designed, and the only difference is the treatment assigned ~~for~~ to each group. A randomized clinical trial is a good choice to eliminate many of the biases and to avoid ethical problems that may arise ~~for~~ from comparing treatments [2, 3]. For example, in a double-blinded two-arm clinical trial for an ophthalmologic study, all patients are randomized into two treatment groups and the same treatment is applied to ~~two~~ both eyes of patients from the same group. Such clustered data with a cluster size of two, often arise ~~s~~ from ~~many~~ statistical and medical applications, such as ophthalmologic studies, orthopaedic studies, otolaryngological studies and twin studies.

We wish to test if the outcomes are identical among the two or more treatment groups. Obviously, the information ~~contributed~~ collected from two eyes of a single person

Key words and phrases. homogeneity test, asymptotic test, intraclass correlation, score test.

Corresponding author: Chang-Xing Ma (cxma@buffalo.edu).

tends to be highly correlated. Any statistical methods such as t tests, analyses of variance, or chi-square tests, that ignores the feature of dependence could lead to incorrect inferences (see, [4], [5], [6], [7], [8]).

In this article, we consider the case of a dichotomous outcome, such as the presence of a disease or some other binary trait. Several statistical tests have been proposed. Rosner [4] proposed a parametric model and a test statistic for testing homogeneity of proportions among g groups, however, the maximum likelihood estimates (MLEs) and likelihood-based tests were not given. Tang et al. (2006, 2008) considered this problem for two groups only and proposed several asymptotic testing procedures, including the score test. It is difficult to extend the testing procedures from 2 groups to g groups ($g > 2$) due to the complexity of deriving the information matrix and maximum likelihood estimates, which can be obtained only by numerical iterations. The score test statistic has been shown to be advantageous to other testing procedures in the comparison between two treatment groups in terms of type I error control and power [10]. We expect the score test, investigated for comparing multiple treatment groups in this article, to perform well as compared to other procedures.

I would rephrase, "The score test statistic has demonstrated better type I error control and power than other testing procedures, when comparing two treatment groups."

In this article, we present the methods for comparing proportions among any g groups, $g \geq 2$. The maximum likelihood estimate under Rosner's model and three different methods (Likelihood Ratio test, Wald-type test, Score test) are derived and investigated in Section 2. In Section 3, Monte Carlo simulation studies are conducted to compare the performance of various tests and comparisons are evaluated with respect to actual type I error rates and powers. Examples from otolaryngological studies are illustrated to demonstrate our methodologies in Section 4. Finally, we give some concluding remarks in Section 5.

2. Methods

Suppose we wish to compare g groups of individuals from an ophthalmologic study with m_i individuals in the ith group, $i = 1, \ldots, g$; $N = \sum m_i$ total subjects (Table 1). Let $Z_{ijk} = 1$ if the kth eye of jth individual in the ith group has a response at the end of the study, and 0 otherwise, $i = 1, \ldots, g$, $j = 0, \ldots, m_i, k = 1, 2$. Let m_{ji} denote the number of subjects who has exactly j responses in the ith group, and S_j be the number of subjects who has exactly j responses (e.g., affected

TABLE 1. Frequencies of the number of affected eyes for persons in g groups

	group				
number of affected eyes	1	2	$\cdots$	g	total
0	m_{01}	m_{02}	$\cdots$	m_{0g}	S_0
1	m_{11}	m_{12}	$\cdots$	m_{1g}	S_1
2	m_{21}	m_{22}	$\cdots$	m_{2g}	S_2
total	m_1	m_2	$\cdots$	m_g	N

eyes)

$$S_j = \sum_{i=1}^{g} m_{ji}, j = 0, 1, 2.$$

A parametric model proposed by Rosner [4] is given as

$$Pr(Z_{ijk} = 1) = \pi_i, Pr(Z_{ijk} = 1 | Z_{ij,3-k} = 1) = R\pi_i, \tag{2.1}$$

$i = 1, \ldots, g,\ j = 0, \ldots, m_i, k = 1, 2$ for some positive R. The constant R is a measure of dependence between two eyes of the same person. If $R = 1$, then the two eyes from the same patient are completely independent, while if $R\pi_i = 1$, then the eyes of each patient in the i-th group are completely dependent. From the conditional probability in Equation (2.1), it is easy to show that the correlation between two eyes is

$$\rho_i = corr(Z_{ij1}, Z_{ij2}) = \frac{\pi_i}{1 - \pi_i}(R - 1), i = 1, \ldots, g.$$

We wish to test whether the response rates of the g groups are identical. The hypotheses are given as

$$H_0 : \pi_1 = \cdots = \pi_g = \pi$$

against

$$H_1 : \text{some of the } \pi_i \text{ are unequal.}$$

Based on the observed data $\tilde{M} = (m_{01}, \cdots, m_{0g}, m_{11}, \cdots, m_{1g}, m_{21}, \cdots, m_{2g})$, the corresponding log-likelihood can be expressed as

$$l(\pi_1, \ldots, \pi_g; R) = \sum_{i=1}^{g} \left[m_{0i} \log\left(R\,{\pi_i}^2 - 2\,\pi_i + 1\right) + m_{1i} \log\left(2\pi_i(1 - R\pi_i)\right) + m_{2i} \log\left(R\,\pi_i^2\right) \right].$$

Differentiating $l(\pi_1, \ldots, \pi_g; R)$ with respect to parameters $\pi_1, \ldots, \pi_g$ and R yields

$$\frac{\partial l}{\partial \pi_i} = \frac{2\,m_{2i}}{\pi_i} + \frac{(2\,R\,\pi_i - 2)\;m_{0i}}{R\,{\pi_i}^2 - 2\,\pi_i + 1} + \frac{(4\,R\,\pi_i - 2)\;m_{1i}}{2\,\pi_i\;(R\,\pi_i - 1)}, i = 1, \ldots, g \tag{2.2}$$

$$\frac{\partial l}{\partial R} = \sum_{i=1}^{g} \left(\frac{m_{2i}}{R} + \frac{{\pi_i}^2\,m_{0i}}{R\,{\pi_i}^2 - 2\,\pi_i + 1} + \frac{\pi_i\,m_{1i}}{R\,\pi_i - 1} \right) \tag{2.3}$$

Under the null hypothesis $H_0 : \pi_1 = \cdots = \pi_g = \pi$, the maximum likelihood estimates of π and R satisfy

$$\frac{\partial l}{\partial R} = 0 \text{ and } \frac{\partial l}{\partial \pi} = \frac{2\,S_2}{\pi} + \frac{(2\,R\,\pi - 2)\;S_0}{R\,\pi^2 - 2\,\pi + 1} + \frac{(4\,R\,\pi - 2)\;S_1}{2\,\pi\;(R\,\pi - 1)} = 0,$$

A direct algebra calculation results in the MLEs of $\pi_i's$ and R

$$\hat{\pi}_{H_0} = \frac{S_1 + 2S_2}{2N}$$

and

$$\hat{R}_{H_0} = \frac{4NS_2}{(S_1 + 2S_2)^2}.$$

Denote $\hat{\pi}_i, i = 1, \ldots, g$ and R as the maximum likelihood estimate of $\pi_i, i = 1, \ldots, g$ and R, respectively. $\hat{\pi}_i, i = 1, \ldots, g$ and R are the solution of the following equations

$$\frac{\partial l}{\partial \pi_i} = 0, i = 1, \ldots, g, \qquad \frac{\partial l}{\partial R} = 0.$$

There is no closed form solution and it has to be solved iteratively. We can simplify the formula in Equation (2.2) as the following 3rd order polynomial (for $i = 1, \ldots, g$)

$$\pi_i^3 - \frac{4m_{0i} + 5m_{1i} + 6m_{2i}}{2Rm_i}\pi_i^2 + \frac{m_{0i} + (1+R)m_{1i} + (2+R)m_{2i}}{R^2 m_i}\pi_i - \frac{m_{1i} + 2m_{2i}}{2R^2 m_i} = 0$$

The $(t+1)$th update for π_i can directly be obtained by the real root of above equation, and R can be updated by Fisher scoring method

$$R^{(t+1)} = R^{(t)} - \left(\frac{\partial^2 l}{\partial R^2}(\pi_1^{(t)}, \ldots, \pi_g^{(t)}; R^{(t)}) \right)^{-1} \frac{\partial l}{\partial R}(\pi_1^{(t)}, \ldots, \pi_g^{(t)}; R^{(t)}).$$

See Appendix for the formula of $\frac{\partial^2 l}{\partial R^2}$.

2.1. **Likelihood ratio test (T_{LR}).** The likelihood ratio (LR) test is given by

$$T_{LR} = 2[l(\hat{\pi}_1, \ldots, \hat{\pi}_g; \hat{R}) - l(\hat{\pi}_{H_0}, \ldots, \hat{\pi}_{H_0}; \hat{R}_{H_0})].$$

Under the null hypothesis, T_{LR} is asymptotically distributed as a chi-square distribution with $g-1$ degrees of freedom.

2.2. **Wald-type test (T_W).** The null hypothesis $H_0 : \pi_1 = \cdots = \pi_g$ can be alternatively expressed as $C\boldsymbol{\beta}^T = 0$ where $\boldsymbol{\beta} = (\pi_1, \cdots, \pi_g, R)$ and

$$C = \begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & 0 \\ & & \ddots & \ddots & & \vdots \\ & & & 1 & -1 & 0 \end{bmatrix}.$$

Wald-type test statistic (T_W) for testing H_0 can be expressed as

$$T_W = (\boldsymbol{\beta}C^T)(CI^{-1}C^T)^{-1}(C\boldsymbol{\beta}^T)|\boldsymbol{\beta} = (\hat{\pi}_1, \ldots, \hat{\pi}_g, \hat{R}),$$

where I is Fisher information matrix for $\boldsymbol{\beta}$ (See Appendix) and T_W is asymptotically distributed as a chi-square distribution with $g-1$ degrees of freedom. T_W can be simplified as

$$T_W = \frac{\sum_{i,j=1}^{g} \hat{\pi}_i \hat{\pi}_j D_{ij}}{\sum_{k=1}^{g} (b_k^2 - h * a_k)},$$

where

$$\begin{aligned}
D_{ij} &= \begin{cases} (b_i^2 - ha_i)\sum_{k\neq i} a_k + a_i(\sum_{k\neq i} b_i)^2, & \text{if } i = j, \\ b_i a_j \sum_{k\neq j} b_k + b_j a_i \sum_{k\neq i} b_k - ha_i a_j - b_i b_j \sum_{k\neq i,j} a_k, & \text{if } i \neq j, \end{cases} \\
a_k &= \frac{-2m_k(2\hat{R}^2\hat{\pi}_k^2 - \hat{R}\hat{\pi}_k^2 - 2\hat{R}\hat{\pi}_k + 1)}{\hat{\pi}_k(\hat{R}^2\hat{\pi}_k^3 - 3\hat{R}\hat{\pi}_k^2 + \hat{R}\hat{\pi}_k + 2\hat{\pi}_k - 1)}, \\
b_k &= \frac{2m_k(1-\hat{R})\hat{\pi}_k^2}{\hat{R}^2\hat{\pi}_k^3 - 3\hat{R}\hat{\pi}_k^2 + \hat{R}\hat{\pi}_k + 2\hat{\pi}_k - 1}, \\
h &= \sum_{i=1}^{g} m_i \left[\frac{\hat{\pi}_i^4}{1 - 2\hat{\pi}_i + \hat{R}\hat{\pi}_i^2} + \frac{\hat{\pi}_i^2}{\hat{R}} + \frac{2\hat{\pi}_i^3}{1 - \hat{R}\hat{\pi}_i} \right].
\end{aligned}$$

Other multivariate tests of π's can be done similarly by choosing corresponding C matrix in above statistic. Further, Wald-type test statistic for testing $H_{0a} : \pi_i = \pi_j$ vs $H_{1a} : \pi_i \neq \pi_j, i \neq j$ can be given by

$$T_{Wa}(i,j) = (\boldsymbol{\beta}c^T)(cI^{-1}c^T)^{-1}(c\boldsymbol{\beta}^T)|\boldsymbol{\beta} = (\hat{\pi}_1, \ldots, \hat{\pi}_g, \hat{R}),$$

where $c = (0, \ldots, 1, \ldots, -1, \ldots, 0)$ with 1 in ith element and -1 in jth element. T_{Wa} is asymptotically distributed as a chi-square distribution with 1 degree of freedom. $T_{Wa}(i,j)$ can be simplified as

$$T_{Wa}(i,j) = \frac{a_i a_j (\sum_{k=1}^{g} (b_k^2/a_k) - h)(\hat{\pi}_i - \hat{\pi}_j)^2)}{(a_i + a_j)(\sum_{k \neq i,j}^{g} (b_k^2/a_k) - h) + (b_i + b_j)^2}.$$

2.3. **Score test (T_{SC}).** The score test statistic T_{SC} is given by

$$T_{SC}^2 = UI(\pi, R)^{-1} U^T | \pi_1 = \cdots = \pi_g = \hat{\pi}_{H_0}, R = \hat{R}_{H_0}$$

where

$$U = \left(\frac{\partial l}{\partial \pi_1}, \ldots, \frac{\partial l}{\partial \pi_g}, 0 \right)$$

and see Appendix for the formula of the inverse of the information matrix $I(\pi, R)^{-1}$.

It can be simplified as

$$T_{SC}^2 = \sum_{k=1}^{g} \frac{N\left(S_1^2 m_{0k} - S_0 S_1 (m_{1k} + 2m_{2k}) + 2S_0 S_2 m_{1k}\right)^2}{S_0 S_1 \left(S_1^3 + S_0 S_1^2 + 4 S_0 S_2^2\right) m_k} \tag{2.4}$$

after lengthy algebra calculations.

Remark 2.1. One limitation of score statistic is that it cannot be computed if $S_0 = 0$ or $S_1 = 0$. We dealt with this problem by adding $1/(2g)$ to m_{ij} for such situations.

Remark 2.2. Tang et al. (2008) derived a score test T_{SC} for $g = 2$ as

$$T_{SC} = \frac{N[S_0 S_1 (m_{11} + 2m_{21}) - m_{01} S_1^2 - 2m_{11} S_0 S_2]}{\sqrt{m_1 m_2 S_0 S_1 [S_1^2 (S_0 + S_1) + 4 S_0 S_2^2]}},$$

which is equivalent to (2.4).

3. Monte Carlo simulation studies

We now investigate the performance of proposed statistics and testing procedures discussed in the previous section. First, we investigate the behavior of the type I error rates of various procedures for g=2,3,4,5; sample size $m_1 = \cdots = m_g = 20$, 40 and 60; $\pi_1 = \cdots = \pi_g = \pi_0 = 0.5(0.1)0.8$; and $R = 1 + \rho(1 - \pi_0)/\pi_0$ where $\rho = 0.4(0.1)0.6$. In each configuration, 50,000 samples are generated based on null hypothesis, and empirical type I error rates are computed as *the number of rejections*/50000, and the results are presented in Table 2. Following Tang et al. (2008), a test is said to be liberal if the ratio of its actual type I error rate to the nominal type I error rate is greater than 1.2 (e.g., > 0.06 for $\alpha = 0.05$, in bold);

conservative if the ratio of its actual type I error rate to the nominal type I error rate is less than 0.8 (e.g., < 0.04); and robust otherwise.

TABLE 2. The type I error rates (percent) of various procedures under $H_0 : \pi_1 = \cdots = \pi_g = \pi_0$ at $\alpha = 0.05$ based on 50,000 replicates

m	π_0	ρ	$g=2$			$g=3$			$g=4$			$g=5$		
			T^2_{LR}	T^2_W	T^2_{SC}	T^2_{LR}	T^2_W	T^2_{SC}	T^2_{LR}	T^2_W	T^2_{SC}	T^2_{LR}	T^2_W	T^2_{SC}
20	0.5	0.4	**6.70**	**6.63**	5.39	**6.95**	**8.09**	5.06	**7.10**	**9.64**	4.99	**7.19**	**10.66**	5.05
		0.5	**6.70**	5.76	5.34	**7.11**	**7.66**	5.10	**7.34**	**9.48**	5.01	**7.34**	**10.81**	4.82
		0.6	**7.11**	4.86	5.35	**7.27**	**6.94**	4.80	**7.72**	**9.51**	4.79	**8.19**	**11.97**	4.92
	0.6	0.4	**6.63**	**6.51**	5.25	**6.80**	**8.18**	5.16	**7.05**	**9.42**	5.15	**6.87**	**10.44**	4.80
		0.5	**6.64**	5.57	5.32	**6.82**	**7.14**	4.95	**6.96**	**8.93**	4.82	**7.38**	**10.45**	4.92
		0.6	**7.18**	4.58	5.37	**7.34**	**6.95**	4.91	**7.64**	**9.29**	4.89	**7.80**	**11.54**	4.85
	0.7	0.4	**6.46**	**6.28**	4.76	**6.79**	**7.96**	5.05	**6.85**	**9.28**	4.92	**6.99**	**10.55**	4.77
		0.5	**6.90**	5.58	5.11	**6.89**	**7.23**	4.90	**7.30**	**9.06**	4.77	**7.64**	**10.98**	4.91
		0.6	**7.46**	4.43	5.02	**7.72**	**6.94**	4.80	**7.97**	**10.36**	4.73	**8.42**	**13.80**	4.92
	0.8	0.4	**6.76**	**6.51**	4.94	**7.26**	**8.35**	4.85	**7.49**	**10.37**	4.81	**7.71**	**12.15**	4.75
		0.5	**7.58**	5.45	4.99	**7.92**	**7.89**	4.78	**8.04**	**11.50**	4.72	**8.24**	**14.92**	4.65
		0.6	**7.81**	4.00	4.37	**8.31**	**6.94**	4.46	**8.15**	**11.87**	4.42	**8.28**	**17.56**	4.52
40	0.5	0.4	5.72	5.61	5.07	6.00	**6.42**	5.16	5.86	**7.01**	4.97	6.00	**7.66**	5.03
		0.5	5.71	5.11	5.15	5.84	5.89	4.98	5.90	**6.80**	4.91	**6.11**	**7.55**	4.92
		0.6	5.76	4.66	5.13	5.84	5.48	4.98	**6.12**	**6.62**	4.97	**6.20**	**7.45**	5.02
	0.6	0.4	5.58	5.48	4.98	5.72	**6.08**	5.05	5.83	**6.90**	4.98	5.87	**7.48**	4.99
		0.5	5.74	5.13	5.14	5.82	5.83	5.03	**6.04**	**6.73**	5.18	**6.16**	**7.57**	5.13
		0.6	5.84	4.73	5.19	5.79	5.37	5.00	5.98	**6.49**	4.94	**6.05**	**7.43**	4.97
	0.7	0.4	5.62	5.41	5.03	5.84	**6.11**	5.13	5.58	**6.42**	4.85	5.71	**7.30**	4.89
		0.5	5.59	4.94	4.96	5.88	5.79	5.21	5.86	**6.65**	5.05	5.69	**7.28**	4.89
		0.6	5.74	4.56	5.00	5.73	5.60	4.88	**6.01**	**6.76**	5.00	**6.08**	**7.59**	4.97
	0.8	0.4	5.76	5.33	5.30	5.85	**6.16**	4.99	5.98	**7.19**	5.10	6.00	**8.00**	5.04
		0.5	5.58	4.63	4.97	5.68	5.78	4.76	5.81	**6.87**	4.68	**6.00**	**7.79**	4.92
		0.6	5.85	4.44	4.91	5.91	5.71	4.67	**6.20**	**7.23**	4.77	**6.43**	**8.51**	4.71
60	0.5	0.4	5.39	5.20	4.94	5.67	**6.02**	5.06	5.60	**6.27**	4.95	5.62	**6.77**	4.91
		0.5	5.57	5.10	5.11	5.85	5.83	5.23	5.75	**6.28**	5.02	5.74	**6.59**	4.99
		0.6	5.58	4.67	5.06	5.60	5.28	5.02	5.84	**6.08**	5.15	5.51	**6.39**	4.79
	0.6	0.4	5.57	5.46	5.19	5.57	5.77	5.06	5.38	**6.09**	4.86	5.80	**6.75**	5.14
		0.5	5.36	4.90	4.97	5.42	5.39	5.01	5.55	5.85	4.94	5.57	**6.39**	4.91
		0.6	5.56	4.79	5.21	5.56	5.33	4.98	5.52	5.78	4.93	5.69	**6.62**	4.95
	0.7	0.4	5.46	5.24	5.09	5.51	5.62	5.10	5.60	**6.23**	5.08	5.44	**6.47**	4.85
		0.5	5.44	4.99	5.10	5.54	5.51	5.08	5.48	**6.02**	4.97	5.47	**6.55**	4.85
		0.6	5.42	4.70	5.04	5.32	5.14	4.84	5.50	5.82	4.91	5.47	**6.57**	4.82
	0.8	0.4	5.41	5.05	5.07	5.52	5.69	5.01	5.47	**6.19**	4.94	5.50	**6.76**	4.97
		0.5	5.42	4.79	5.04	5.67	5.58	5.19	5.61	**6.14**	4.98	5.53	**6.91**	4.80
		0.6	5.36	4.56	4.87	5.76	5.59	5.08	5.78	**6.54**	4.88	5.81	**7.33**	4.94

Generally, score tests T^2_{SC} produce satisfactory type I error controls for any configuration while LR tests and Wald tests are liberal, especially for small samples and larger numbers of groups (g). When $g > 2$, Wald tests are more liberal than LR tests and these tests get closer when sample size ~~goes~~ becomes larger.

TABLE 3. The power (percent) of various procedures at $\alpha = 0.05$ based on 50,000 replicates

m	R	$g=2$				$g=3$				$g=4$			
		T^2_{SC}	T^2_{LR}	T^2_W	T	T^2_{SC}	T^2_{LR}	T^2_W	T	T^2_{SC}	T^2_{LR}	T^2_W	T
20	1.0	11.8	12.7	13.5	11.7	12.7	14.5	17.2	13.0	13.1	15.2	19.4	13.2
	1.5	10.2	12.4	12.9	10.4	11.0	14.0	16.2	11.1	11.6	15.2	19.2	11.8
	2.0	10.4	13.7	12.2	9.4	11.6	16.5	16.4	10.1	12.7	18.3	20.4	10.1
40	1.0	24.6	25.3	26.0	24.8	30.4	31.3	32.8	30.6	34.7	35.8	37.9	34.9
	1.5	22.2	23.5	24.0	21.6	26.6	28.4	29.8	25.7	30.6	32.7	34.8	29.3
	2.0	23.6	25.7	24.7	19.1	29.9	32.6	32.2	22.0	35.9	39.0	39.5	24.8
100	1.0	38.0	38.5	39.1	38.2	48.2	48.8	49.8	48.3	56.0	56.8	58.1	56.1
	1.5	33.5	34.4	34.8	32.4	42.5	43.9	44.9	40.5	49.8	51.2	52.5	47.0
	2.0	36.5	38.1	37.4	28.7	47.8	50.1	49.9	35.1	57.4	59.7	59.9	40.1
H_1 :	π=	(0.25, 0.325)				(0.25, 0.30, 0,35)				(0.25, 0.2875, 0.325)			

Note: T is the test statistic in Rosner (1982).

LR tests and Wald-type tests are extremely liberal for a moderate sample size (i.e., $m = 20$), and their actual sizes inflate with the increase of the correlation coefficient (i.e., ρ). Therefore, the score test will be recommended.

Next, we evaluate the power performance of proposed methods. We consider the alternative hypotheses with $H_1 : \pi = (0.25, 0.325), (0.25, 0.30, 0, 35)$, and (0.25, 0.2875, 0.325) for g=2, 3, and 4, respectively. R is chosen as 1, 1.5, and 2.0 and sample size $m_1 = \cdots = m_g = 20$, 40 and 100. Ronser (1982)'s statistic T is also considered in the simulation studies and the results are presented in Table 3.

Based on the simulation results, LR and Wald tests are generally more powerful than score tests and Ronser (1982)'s T ~~is~~ generally ~~with~~ has less power. However, LR and Wald tests inflate power because their empirical levels are larger than the nominal level (see Table 2). For moderate or large sample sizes, the powers of the three proposed ~~three~~ methods are close. Overall, the score test is highly recommended as it has more power with satisfactory type I error control.

4. WORK EXAMPLES

We reanalyze the data presented by Rosner [4] to illustrate the newly proposed methods. The outpatient population of 218 persons aged 20-39 with retinitis pigmentosa (RP) were classified on the basis of a detailed family history into the genetic types of autosomal dominant RP (DOM), autosomal recessive RP (AR), sex-linked RP (SL), and isolate RP (ISO) for a study of differences between these

TABLE 4. Distribution of the number of affected eyes for persons in each genetic type

	genetic type			
number of affected eyes	DOM	AR	SL	ISO
0	15	7	3	67
1	6	5	2	24
2	7	9	14	57

TABLE 5. Statistic and p-value for comparing VA for different genetic types of RP

method	T^2_{LR}	T^2_W	T^2_{SC}	T
statistic	5.8862	6.2966	6.8475	11.36
p-value	0.1173	0.0980	0.0769	0.010

TABLE 6. Wald-type test results comparing VA for different genetic types of RP

Group	MLE	Standard	Comparison group			
i	$\hat{\pi}_i$	Error	DOM	AR	SL	ISO
DOM	0.3930	0.0041	–	-0.0868 (p =0.3116)	-0.1698 (p =0.0207)	-0.1001 (p =0.1363)
AR	0.4798	0.0039		–	-0.0830 (p =0.2135	-0.0132 (p =0.8284)
SL	0.5628	0.0022			–	0.0697 (p =0.0748)
ISO	0.4931	0.0011				–
	$\hat{R} = 1.6639$					

four groups on the Snellen visual acuity (VA). An eye was considered affected if VA was 20/50 or worse, and normal if VA was 20/40 or better. The sample used for this analysis consists of 216 persons out of the sample of 218 persons, each of whom had complete information for VA on both eyes (Table 4).

An overall significant difference between the proportions of affected eyes in the four groups is from 0.0769 to 0.1173 based on proposed methods and 0.010 on Rosner's statistic T (Table 5).

The maximum likelihood estimates and pairwise comparisons are shown in Table 6. It shows a significant difference between DOM and AR (p=0.0207).

Another example was a recent study from a cross-sectional, population-based sample in Iran to assess the prevalence of avoidable blindness (Rajavi et al., 2011).

TABLE 7. Prevalence of avoidable blindness of a sample population in Iran

Age Group	Blindness			Sample Prevalence	MLE
	None	Unilateral	Bilaterial		
50-54 yrs	964	23	2	0.014	0.014
55-59 yrs	541	17	8	0.029	0.030
60-64 yrs	469	18	4	0.026	0.027
65-69 yrs	257	16	5	0.047	0.048
70-74 yrs	242	32	3	0.069	0.067
75-79 yrs	127	30	9	0.145	0.134
80+ yrs	104	29	10	0.171	0.149

Nearly 3000 persons were examined and the blindness are assessed for seven age groups (Table 7). Test statistics $T^2_{LR} = 134.7$, $T^2_W = 89.1$, $T^2_{SC} = 161.1$, and $T = 202.0$ all show the significant age differences consistently (p-value<0.0001), and MLE $\hat{R} = 3.35$ shows positive correlation between eyes in a person.

5. CONCLUDING REMARKS

In this article, we investigated three procedures for testing the homogeneity for correlated data with cluster size two. We derived the maximum likelihood estimate algorithm by utilizing the root of third order polynomial equations. Fisher scoring method is usually criticized for converging slowly, especially when the number of parameters is large (e.g. g is large). However, the algorithm derived in this paper is very efficient because only R is updated by Fisher scoring iterations, and $\pi_i, i = 1, \ldots, g$ are the roots of third order polynomials, a closed form solution.

Simulation results show that the proposed approach (score test) has satisfactory type I error control with very high power, regardless of number of groups, sample sizes, and parameter configurations. The LR test and Wald test have inflated type I error and therefore are not recommended.

Although there are many convenient ways to solve the MLE iteratively or hypothesis tests with today's computation power (e.g., in R or other software), the explicit form of test statistic (e.g., score statistic) is still very useful not only for its simplicity, but also for future development of the test. For example, in small sample situation, an exact test may overcome the inflated type I error rate; however, the

exact test requires extensive calculations and it is nearly impossible using iterative versions of test statistics.

To overcome inflated type I error control in asymptotic tests, Tang et al. (2006) and Shan et al. (2013) considered exact tests for g=2. We consider the exact tests for $g > 2$ as interesting future work.

References

[1] R. Simon, R. E. Wittes, and S. S. Ellenberg. Randomized phase II clinical trials. *Cancer treatment reports*, 69(12):1375–1381, December 1985.

[2] G. E. Wilding, G. Shan, and A. D. Hutson. Exact two-stage designs for phase II activity trials with rank-based endpoints. *Contemporary Clinical Trials*, 2011.

[3] A. Donner. Cluster randomization trials in epidemiology: theory and application. *Journal of Statistical Planning and Inference*, 42(1-2):37–56, November 1994.

[4] B. Rosner. Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. *Biometrics*, 38(1):105–114, March 1982.

[5] G.E. Dallal. Paired Bernoulli trials. *Biometrics* 44, 253257, 1988.

[6] A. Donner. Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics* 45, 605–611.

[7] B. Rosner, R.C. Milton. Significance testing for correlated binary outcome data. *Biometrics* 44, 505–512, 1988.

[8] C.A. Bodian. Intraclass correlation for two-by-two tables under three sampling designs. *Biometrics* 50, 183–193, 1994.

[9] M.L. Tang, N.S. Tang, B. Rosner. Statistical inference for correlated data in ophthalmologic studies. *Statistics in Medicine* 25, 2771–2783, 2006.

[10] N.-S. Tang, M.-L. Tang, and S.-F. Qiu. Testing the equality of proportions for correlated otolaryngologic data. *Computational Statistics & Data Analysis*, 52(7):3719–3729, March 2008.

[11] Z. Rajavi, M. Katibeh, H. Ziaei, N. Fardesmaeilpour, M. Sehat, H. Ahmadieh,M.A. Javadi. Rapid assessment of avoidable blindness in Iran, *Ophthalmology* 118, 1812–18, 2011.

[12] G. Shan and C.X. Ma. Exact methods for testing the equality of proportions for clustered data, submitted for publication, 2013.

Appendix

5.1. **Information matrix.** Differentiating $\frac{\partial l}{\partial \pi_i}, i = 1, \ldots, g$ and $\frac{\partial l}{\partial R}$ with respect to $\pi_i, i = 1, \ldots, g$ and R respectively yields

$$\begin{aligned}
\frac{\partial^2 l}{\partial \pi_i^2} &= \frac{m_{0i}\left(-2\,R^2\,\pi_i^2+4\,R\,\pi_i+2\,R-4\right)}{\left(R\,\pi_i^2-2\,\pi_i+1\right)^2}-\frac{2\,m_{2i}}{\pi_i^2}-\frac{\left(2\,R^2\,\pi_i^2-2\,R\,\pi_i+1\right)\,m_{1i}}{\pi_i^2\,\left(R\,\pi_i-1\right)^2},\\
\frac{\partial^2 l}{\partial \pi_i \partial R} &= -\frac{m_{1i}}{\left(R\,\pi_i-1\right)^2}-\frac{2\,\left(\pi_i-1\right)\,\pi_i m_{0i}}{\left(R\,{\pi_i}^2-2\,\pi_i+1\right)^2},\\
& \qquad i = 1, \ldots, g\\
\frac{\partial^2 l}{\partial \pi_i \partial \pi_j} &= 0, i \neq j,\\
\frac{\partial^2 l}{\partial R^2} &= -\frac{S_2}{R^2}-\sum_{i=1}^{g}\frac{\pi_i^2\,m_{1i}}{\left(R\,\pi_i-1\right)^2}-\sum_{i=1}^{g}\frac{\pi_i^4\,m_{0i}}{\left(R\,\pi_i^2-2\,\pi_i+1\right)^2}.
\end{aligned}$$

Then we have

$$\begin{aligned}
I_{ii} &= E\left(-\frac{\partial^2 l}{\partial \pi_i^2}\right)=\frac{2\,m_i\,\left(2\,R^2\,{\pi_i}^2-R\,{\pi_i}^2-2\,R\,\pi_i+1\right)}{\pi_i\,\left(R\,{\pi_i}^2-2\,\pi_i+1\right)(1-R\,\pi_i)},\\
I_{i,g+1} &= E\left(-\frac{\partial^2 l}{\partial \pi_i \partial R}\right)=-\frac{2\,\left(1-R\right)\,{\pi_i}^2\,m_i}{\left(R\,{\pi_i}^2-2\,\pi_i+1\right)(1-R\,\pi_i)},\\
& \qquad i = 1, \ldots, g\\
I_{ij} &= E\left(-\frac{\partial^2 l}{\partial \pi_i \partial \pi_j}\right)=0, i \neq j,\\
I_{g+1,g+1} &= E\left(-\frac{\partial^2 l}{\partial R^2}\right)=\sum_{i=1}^{g}\frac{{\pi_i}^2\,m_i(R\pi_i-2\,\pi_i+1)}{R(R\,{\pi_i}^2-2\,\pi_i+1)(1-R\,\pi_i)}.
\end{aligned}$$

The $(g+1)\times(g+1)$ information matrix is denoted as $I(\pi_1, \ldots, \pi_g; R) = (I_{ij})$.

Under null hypothesis $H_0 : \pi_1 = \cdots = \pi_g = \pi$, it is straightforward but tedious to show that the inverse of information matrix can be expressed as

$$I^{-1}(\pi; R) = \frac{\pi^4\,R\,(R-1)^2}{N(2\,\pi^2\,R^2-\pi^2\,R-2\,\pi\,R+1)}\begin{bmatrix} c_1 & 1 & \cdots & 1 & d\\ 1 & c_2 & 1 & 1 & d\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & 1 & \cdots & c_g & d\\ d & d & d & d & h\end{bmatrix}$$

where

$$\begin{aligned}
c_i &= \frac{\left(R\,\pi^2-2\,\pi+1\right)\,\left(1-\pi\,R\right)N}{2\,\pi^3\,R\,(R-1)^2 m_i}+1, i = 1, \ldots, g,\\
d &= \frac{2\,\pi^2\,R^2-\pi^2\,R-2\,\pi\,R+1}{\pi^3\,\left(R-1\right)},
\end{aligned}$$

and

$$h = \frac{\left(2\,\pi^2\,R^2 - \pi^2\,R - 2\,\pi\,R + 1\right)^2}{\pi^6\,(R-1)^2}.$$

Department of Biostatistics, University at Buffalo, New York 14214, USA

E-mail address: cxma@buffalo.edu

Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154

E-mail address: guogen.shan@unlv.edu