Chiang Mai J. Sci. 2015; 42(4) : 1031-1037
http://epg.science.cmu.ac.th/ejournal/
Contributed Paper

Comparing Tests for Association in Two-by-Two Tables with Zero Cell Counts

Nurin Dureh, Chamnein Choonpradub* and Phattrawan Tongkumchum
Department of Mathematics and Computer Sciences, Faculty of Science and Technology, Prince of Songkla University Pattani Campus, Pattani 94000, Thailand.
*Author for correspondence; e-mail: chamnein.c@gmail.com

Received: 7 October 2013
Acepted: 24 January 2014

ABSTRACT

This study compared the tests for association in two by two tables with zero cell counts. Pearson's uncorrected chi-squared test, Pearson's chi-squared test with the continuity correction, Pearson's uncorrected chi-squared Monte Carlo simulation test, Fisher's exact test, the Conditional Binomial Exact Test (CBET), Barnard's exact test, Liebermeister's test, Lancaster's mid P-test and logistic regression with penalized maximum likelihood were considered. Criteria used 72 two by two tables with smallest counts and average p-value for Fisher's exact test and Pearson's uncorrected chi-squared test close to 0.05. CBET, Lancaster's mid-P test, and the penalized maximum likelihood test give similar p-values closest to 0.05, suggesting that these three methods can be recommended for testing association in two by two tables with zero cell counts.

Keywords: zero cell, two by two table, association

1. INTRODUCTION

Several methods have been recommended for analysis of association in two-by-two tables. The most common test is Pearson's chi-squared test, which is appropriate for sufficiently large sample sizes. It is inaccurate if any expected count is less than five [1, 2, 3]. In cases of small sample sizes, Fisher's exact test is recommended [1, 2, 3]. This method eliminates the nuisance parameter in the model under the null hypothesis by conditioning on its marginal totals [4] but is conservative. Another way to reduce the conservatism of Fisher's exact test is to consider an unconditional approach, such as Barnard's test, which eliminates the nuisance parameter by taking its supreme value over all possible values in the space of the null model [5, 6]. Several alternative tests also have been proposed [7]. These include Lancaster's mid-P test [8], an adjustment to the Fisher's exact test that tends to have increased power while maintaining a Type I error rate close to the nominal level [7, 9]. Liebermeister's test also can be used in place of Fisher's exact test, and is less conservative than Fisher's test and just as easy to calculate [3]. In addition, the "conditional binomial exact test" (CBET) is proposed as an alternative test for comparing binomial proportions estimated from samples of larger populations [10].

Logistic regression provides a more

general method because it provides a model that accommodates more complex determinants. However, when one of the four cells in the table is equal to zero, maximum likelihood estimates fail to converge [3, 5, 7, 10, 11, 12, 13]. A solution to this problem was proposed by Firth [14], giving finite parameter estimates based on penalized maximum likelihood [14, 15, 16, 17, 18]. This method is available in statistical software such as SAS, S-PLUS and R [17, 19]. However, these estimates are biased away from zero [16], so it is important to know how substantial these biases are.

Tables with a zero cell count thus lead to numerical problems [20], so it is important to identify the methods which provide the most accurate results for particular data structures. With this information, researchers can select the appropriate method for their studies. Thus the main objective of this study was to compare the results when using recommended tests for association in two-by-two tables with small cell counts.

2. MATERIALS AND METHODS

Tests for an association in two-by-two table

There are several methods for testing the association in two-by-two tables. If the two-by-two table contains counts as in Table 1, a brief summary of computing a p-value of these tests may be described as follows.

Table 1: The general counts of a two-by-two table.

	1	2	Total
1	a	b	m
2	c	d	n
Total	z	v	m+n

Pearson's uncorrected chi-squared test

The functional form of this test is

$$\chi^2_p = \frac{m+n(ad-bc)^2}{(a+b)(c+b)(a+c)(b+d)} \qquad (1)$$

In general, the *p*-value is defined as the probability of the test statistic *T* being equal to or more extreme than its value for the observed table (t_{obs}), therefore, the approximate p-value for Pearson's chi-squared test is [6]

$$\text{p-value} = P(\chi^2 \geq t_{obs}).$$

Pearson's chi-squared test with the continuity correction (Pearson's CC)

A continuity correction for the Pearson's chi-squared test was proposed by Yates (1984). The corresponding formula for Pearson's CC test is

$$\chi^2_{PCC} = \frac{n(abs(ad-bc)-n/2)^2}{(a+b)(c+d)(a+c)(b+d)}. \qquad (2)$$

Pearson's uncorrected chi-squared test with Monte Carlo Simulation

This test uses a reference set of 10,000 samples to compute the p-value for Pearson's uncorrected chi-squared test in (1).

Fisher's Exact Test

Fisher's Exact P-value is obtained by conditioning on the total number of observed successes [3]. If r is the observed value in a cell, which can be greater or equal to a, the formula is

$$P_F = \sum_{r \geq a} \binom{m}{r}\binom{n}{z-r} \Big/ \binom{m+n}{z} \qquad (3)$$

Alternative test statistics which can be used in place of Fisher's exact test with small counts are as follows.

Liebermeister's test

This test is the quasi-exact test for two binomials. It is based on adjusting the observed table and can be obtained as the formula

$$P_L = \sum_{r \geq a+1} \binom{m+1}{r}\binom{n+1}{z+1-r} \Big/ \binom{m+n+2}{z+1} \qquad (4)$$

Lancaster's mid-P test

From (2), we may write Fisher's Exact P-value as P_F or P_F (a). As Lancaster's mid-P test is Fisher's exact test adjusted, the formula for this p-value is [3].

$$P_M = \left[P_F(a) + P_F(a+1)\right]/2 \quad (5)$$

Barnard's exact test

Suppose t = {X : X is a 2 by 2 table as in Table 1}

Barnard's exact test is an unconditional test. Suppose T(X) is a "discrepancy measure" or test statistic that measures how discrepant any table X is relative to the type of table one would expect under the null hypothesis. It generates the exact distribution of T(X) by considering all the tables $X \in \tau$. If $p(\pi)$ is the exact p-value for any given π, Barnard suggested that we calculate $p(\pi)$ for all possible values of $\pi \in (0,1)$ and choose the value π which maximizes $p(\pi)$, thus, Barnard's exact p-value is defined as [2].

$$P_B = \sup\{p(\pi): \pi \in (0,1)\} \quad (6)$$

Conditional Binomial Exact Test (CBET)

This test is derived from the joint distribution of two binomial samples and conditioned by the estimate of the probability of success p based on the combined samples [10].

Data Simulation

Data comprising 72 two-by-two tables were created based on the condition that one cell is always equal to zero and the rest are small counts that make the averaged p-value from Pearson's chi-square test and Fisher's exact test close to 0.05. We selected these 72 tables because they cover all such tables that fail to satisfy the sample size requirement in Pearson's chi-squared test that all expected counts are at least 5. We selected these two methods because they are most commonly preferred when testing independence in categorical data. Pearson's chi-square test is the conventioned method for testing independence and Fisher's exact test is the prefered method when the sample sizes are too small. Therefore, using the averaged p-value from these two methods as a reference value is reasonable.

Table 1. Cell counts in 72 two-by-two tables where one cell contains zero and the averaged p-value is close to 0.05.

Table	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
b	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	18	27	37	46	55	64	74	83	92	6	8	10	13	15	17	20	22	25
Table	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
a	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4
b	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	4	5	6	8	9	10	11	12	13	3	4	5	6	8	8	10	11	12
Table	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
a	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6
b	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	2	3	4	5	6	7	7	8	9	2	3	4	4	5	6	6	7	8
Table	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
a	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	8
b	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	1	2	3	4	4	5	5	6	6	2	2	3	3	4	4	5	6	6

The nine methods used for testing association in the two-by-two tables were Pearson's uncorrected chi-squared test, Pearson's chi-squared test with the continuity correction (Pearson's CC), Pearson's uncorrected chi-squared test with Monte Carlo Simulation, Fisher's exact test, Liebermeister's test, Lancaster's mid-P test , Barnard's exact test, Conditional Binomial Exact Test (CBET) and logistic regression with penalized maximum likelihood estimates. The reference p-value used was p=0.05 based on the average p-value of Pearson's uncorrected chi-squared test and Fisher's exact test.

3. RESULTS

The p-values from the nine methods applied to tables 1-36 are shown in Figure 1 and Figure 2 shows p-values from tables 37 - 72. The solid line represents p-values equal to 0.05 and each connected line denotes the p-values for each test.

The p-values from each test not entirely consistent but almost all are between 0.01 and 0.2. A group including Pearson's chi-squared test with the continuity correction, Pearson's uncorrected chi-squared with the Monte Carlo simulation test, Fisher's exact test and Barnard's test give p-values higher than 0.05.

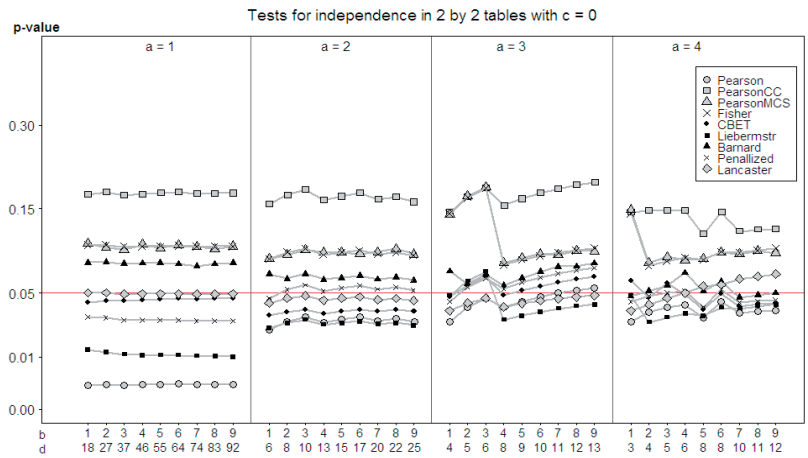

Figure 1. P-values from the recommended tests using data in two-by-two tables with c= 0 and a is 1, 2, 3 and 4.

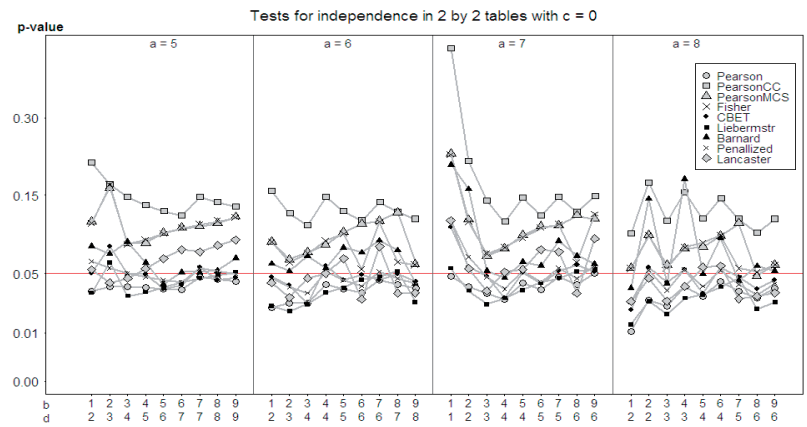

Figure 2. P-values from the recommended tests using data in two-by-two tables with c= 0 and a is 5, 6, 7 and 8.

Pearson's uncorrected chi-squared test and Liebermeister's test tend to give p-values lower than 0.05 and tend to increase when the sample size is increased. CBET, Lancaster mid-P test and the penalized maximum likelihood method gave p-values close to 0.05. There were no outliers distant from 0.05 for any of these three methods.

For comparison, Table 2 displays the average p-values of the 72 tables from those nine tests. It is clear that Lancaster's mid-P test, the Conditional Binomial Exact Test (CBET) and penalized maximum likelihood gave p-values closest to 0.05. Pearson's uncorrected chi-squared test and Liebermeister's test gave averaged p-values lower than 0.05. On the other hand, Fisher's exact test, Pearson's chi-squared test with the continuity correction, Pearson's uncorrected chi-squared test with Monte Carlo simulation and Barnard's exact test gave large biases. All tended to have a large p-values.

Table 2. Average p-values from nine recommended tests for testing the association in two-by-two tables with zero cell count.

Test	Averaged p-value
Pearson's uncorrected chi-squar	0.0320
Pearson's corrected chi-square	0.1528
Pearson's test with Monte Carlo simulation	0.0970
Fisher's exact test	0.0939
Conditional Binomial Exact Test (CBET)	0.0469
Barnard's exact test	0.0694
Liebermeister's test	0.0332
Penalized Maximum Likelihood	0.0478
Lancaster's mid-P test	0.0501

Example : Child Deaths from Perinatal Originating Conditions in Thai Provinces

The data shown in Table 3 are the numbers of child deaths from perinatal originating conditions in nine groups of provinces, based on the Thai 2005 Verbal Autopsy (VA) study [21, 22, 23, 24]. For the question, "Is the proportion of deaths from perinatal originating conditions in Chumporn province different from those in other provinces in Thailand ? ", the results are shown in Table 4.

Table 3. Number of child death from perinatal originating conditions by province.

Provinces	Cause of death		Total
	Other	Perinatal	
Other	84	59	143
Chumporn	6	0	6

Table 4. Average p-values from nine recommended tests for comparing proportion of child deaths from perinatal originating conditions.

Test	Averaged p-value
Pearson's uncorrected chi-square	0.0429
Pearson's corrected chi-square	0.1100
Pearson's test with Monte Carlo simulation	0.0838
Fisher's exact test	0.0815
Conditional Binomial Exact Test (CBET)	0.0483
Barnard's exact test	0.0541
Liebermeister's test	0.0423
Penalized Maximum Likelihood	0.0441
Lancaster's mid-P test	0.0588

The p-values from those nine recommended tests show that the tests including Pearson's uncorrected chi-square test, CBET, Liebermeister's test and Penalized Maximum Likelihood gave p-values 0.0429, 0.0483, 0.0423 and 0.0441, respectively, indicating that Chumporn province differed from other provinces.

4. Conclusion and discussion

This study compared the accuracy of nine separate tests of the association in two-by-two tables, where one cell contained a zero count, using a reference p-value equal to 0.05. When comparing the individual p-value with the reference p-value, most of the tests gave p-values in the range from 0.01 and 0.2. This study showed that the methods of Pearson's chi-squared test and Fisher's exact test were not appropriate for this condition of a zero count in a two-by-two tables, because of the high p-values resulting from their application.

Lancaster's mid-P test, Conditional Binomial Exact Test and a method using penalized maximum likelihood were identified as acceptable and clearly preferable in testing the association in two-by-two tables with zero counts. These three methods consistently produced results close to the reference (p=0.05), in average as shown in Table 2 and in range as shown in Figures 1 and 2.

For CBET, this confirms the finding by Rice (1988) that CBET can be used in place of Fisher's exact test when analyzing contingency tables that compare binomial proportions. In addition, this study can also recommend the use of Lancaster's mid-P test and penalized maximum likelihood. The three methods, Conditional Binomial Exact Test, Lancaster's mid-P test and penalized maximum likelihood can be recommended in cases of testing the association in two-by-two tables with zero cell counts. Since the main objective of this study was not to identify the best method but to compare the results when using recommended tests for association in two-by-two tables, this study can not conclude which method is best. The answer to this question are depend on many conditions, for example, the important of the data, software availability and simplicity of calculation.

Acknowledgements

This research received full financial support from the Thailand Research Fund through the Royal Golden Jubilee Ph.D.Program. We are grateful to Emeritus Professor Don McNeil for his guidance.

References

[1] Mehta C.R. and Patel NR., *Exact Inference for Categorical Data*, Harvard University and Cytel Software Corporation, 1997.

[2] Mehta C.R. and Senchaudhuri P, *Conditional versus Unconditional Exact Tests for Comparing Two Binomials*, Cytel Software Corporation, Cambridge, 2003.

[3] Seneta E. and Phipps M.C., *Biometrical J.,* 2001; **43**: 23-43. DOI:10.1002/1521-4036(200102)43:1<23::AID-BIMJ23>3.0.CO;2-8.

[4] Mehrotra D.V., Chan I.S. and Berger R.L., *Biometrics*, 2003; **59**: 441-450.

[5] Lin C.Y. and Yang M.C., *Commun. Stat. Simulat.,* 2009; **38**: 78-91. DOI:10.1080/03610910802417812.

[6] Lydersen S., Fagerland M.W. and Laake P., *Stat. Med.*, 2009; **28:** 1159-1175. DOI: 10.1002/sim.3531.

[7] Biddle D.A. and Morris S.B., *J. Appl. Psychol.*, 2011; **96**: 956-965. DOI: 10.1037/a0024223.

[8] Lancaster H.O., *J. Am. Stat. Assoc.,* 1961; **56**. DOI:10.1080/01621459.1961.10482105.

[9] Chen L.S. and Lin C.Y., *Commun. Stat. Theory,* 2009; **38**: 1635-1648. DOI:10.1080/03610920802513221.

[10] Rice W.R., *Biometrics*, 1988; **44**: 1-22.

[11] Aitkin M. and Chadwick T., Bayesian analysis of 2x2 contingency tables from comparative trials; Available at:http://www. mas.ncl.ac.uk/~nma9/Bayes2×2.pdf.

[12] Bester C.L. and Hansen C., Bias Reduction for Bayesian and Frequentist Estimators, 2005; Available at:http://faculty.chicagobooth.edu/christian.hansen/research/bh_brbayes.pdf.

[13] Sean R.E., *Nat. Biotechnol.*, 2004; **22**. DOI:10.1038/nbt0904-1177.

[14] Firth D., *Biometrika*, 1993; **80**: 27-38. DOI: 10.1093/biomet/80.1.27.

[15] Eyduran E., *J. Res. Med. Sci.*, 2008; **13**: 325-330.

[16] Heinze G. and Schemper M., *Stat. Med.*, 2002; **21**: 2409-2419. DOI: 10.1002/sim.1047.

[17] Heinze G. and Ploner M., *Comput. Meth. Prog. Bio.*, 2003; **71**: 181-187. DOI:10.1016/S0169-2607(02)00088-3.

[18] Heinze G., Avoiding infinite estimates in logistic regression- theory, solutions, examples, 2009; Available at:http://www.researchgate.net/publicztion/255594296_Avoiding_infinite_estimates_in_logistic_regression_theory_solutions_examples#full_tex+

[19] Heinze G. and Ploner M., A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems, 2004; Available at: http://www.meduniwien.ac.at/user/georg.heinze/techreps/tr2_2004.pdf.

[20] Brown M.B., *Comput. Stat. Data. An.*, 1983; **1**: 3-15. DOI:10.1016/0167-9473(83)90059-2.

[21] Pattaraarchachai J., Rao C., Polprasert W., Porapakkham Y., Pao-in W., Singwerathum N. and Lopez A.D., *Popul. Health Metrics,* 2010; **8**:12. DOI:10.1186/1478-7954-8-12.

[22] Polprasert W., Rao C., Adair T., Pattaraarchachai J., Porapakkham Y. and Lopez A.D., *Popul. Health Metrics,* 2010; **8**:13. DOI:10.1186/1478-7954-8-13.

[23] Porapakkham Y., Rao C., Pattaraarchachai J., Polprasert W., Vos T., Adair T. and Lopez A.D., *Popul. Health Metrics,* 2010; **8**:14. DOI:10.1186/1478-7954-8-14.

[24] Rao C., Porapakkham Y., Pattaraarchachai J., Polprasert W., Swampunyalert W. and Lopez A.D., *Popul. Health Metrics,* 2010; **8**:11. DOI:10.1186/1478-7954-8-11.