Far East Journal of Theoretical Statistics
Published Online: June 2014
Available online at http://pphmj.com/journals/fjts.htm
Volume 46, Number 2, 2014, Pages 137-146

SIMPLE IMPROVED CONFIDENCE INTERVALS FOR COMPARING BINARY MATCHED-PAIR DATA

Yohei Kawasaki and Emiko Haramo

Division of Biostatistics
Department of Clinical Study and Informatics
Center for Clinical Science
National Center for Global Health and Medicine
1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan

Multi-intelligence Analysis Group
System Development Department
Takumi Information Technology
2-40-13 Ikebukuro, Toshima-ku, Tokyo 171-0014, Japan

Abstract

Matched-pair data are employed in various clinical study designs. Agresti and Min [1] indicated that a Wald-type statistic confidence interval difference of marginal probability shows that the actual confidence level becomes lower than the nominal confidence level and proposed a confidence interval by adding k times the hypothetical trials to the Wald interval. In this study, we propose a new confidence interval based on Agresti's method concerning the ratio of marginal probability. Furthermore, we indicate the usability of the confidence interval, which we suggest in this paper through a simulation.

1. Introduction

Matched-pair data are used in various clinical study designs. For

Received: February 20, 2014; Accepted: March 10, 2014
2010 Mathematics Subject Classification: 62P10.
Keywords and phrases: matched-pair data, confidence interval, cross-over design, binary data.

example, repeated measure design, which measures one test subject repeatedly, uses a pretest and posttest to observe the efficacy or safety of a drug before and after its administration. In addition, these data are used in other designs, such as in matched-pair designs for case control studies or in cross-over designs used in drug development.

In these types of study designs, response variables that have binary data exhibit some interrelationship. We discuss a cross-over design example, in which a new drug and a standard drug are administered to the same test subject. For this case, the main outcome data are binary data of whether each test subject has improved, and four patterns result: both drugs cause an improvement, one improves and the other does not (and vice versa), or both drugs fail to cause an improvement. Thus, it is possible to get results for the same test subject each time, and the response variable is binary data with mutual relationships.

Nam and Blackwelder [7] indicated that analyzing binary data that has an interrelationship without considering the relationship was inefficient. Tango [9] proposed an analysis method for the difference of marginal probability. An inference method for the difference of marginal probability has often been discussed in recent papers. A Wald-type statistic confidence interval difference of marginal probability shows that the actual confidence level becomes lower than the nominal confidence level. A new confidence interval was proposed by adding k times the hypothetical trials to the Wald interval as an improved method. Lloyd [6] built a confidence interval that included a coverage probability that was not reduced below the nominal confidence level.

Lachenbruch and Lynch [5] indicated that there are cases when it is more suitable to use ratios rather than the difference between marginal probabilities in clinical studies. Nam and Blackwelder [7] proposed a confidence interval by using the maximum likelihood estimator with restrictions rather than a Wald-type test method. Bonett and Price [2] proposed a hybrid confidence interval that combined two Wilson-type confidence intervals proposed by Newcombe [8].

The remainder of the paper is structured as follows: Section 2 indicates the statistical model and notation method. Section 3 shows the construction method of the four confidence intervals and that of the new confidence interval that is extended to improve on the method of Agresti and Min [1]. In Section 4, we apply the confidence intervals proposed in Section 3 to actual clinical tests. Furthermore, we calculate the median of the actual confidence level and the confidence interval using a simulation, and we search for and verify the optimal confidence interval. Section 5 provides a discussion.

2. Materials and Methods

2.1. Model and notation

In this section, we indicate the probability model and notation method of the binary matched-pair data. X_{ij} is the probability variable of the reaction for the new drug *i* and standard drug *j*, where each *i*, *j* is 1 for success and 0 for failure. The population probability is p_{ij}, and the probability variables $X_{11}, X_{10}, X_{01}, X_{00}$ follow a multinomial distribution of parameters *n*, $p_{11}, p_{10}, p_{01}, p_{00}$.

$$f_X(x_{11}, x_{10}, x_{01}, x_{00}) = \frac{n!}{x_{11}!\,x_{10}!\,x_{01}!\,x_{00}!} p_{11}^{x_{11}} p_{10}^{x_{10}} p_{01}^{x_{01}} p_{00}^{x_{00}}.$$

x_{ij} is a non-negative integer and follows $\sum x_{ij} = n$. Under this condition, the marginal population probability (hereinafter marginal probability) is $p_1 = p_{11} + p_{10}$ (population probability of success with new drug) and $p_0 = p_{11} + p_{01}$ (population probability of success with standard drug). The ratio of the marginal probability of interest may be expressed as $\theta = p_1/p_0$ (Table 1).

Table 1. Random variable and probabilities in a 2×2 contingency table

	Standard				Standard		
New	Success	Failure	Total	New	Success	Failure	Total
Success	X_{11}	X_{10}	X_1	Success	p_{11}	p_{10}	p_1
Failure	X_{01}	X_{00}	$N - X_1$	Failure	p_{01}	p_{00}	$1 - p_1$
Total	X_0	$N - X_0$	N	Total	p_0	$1 - p_0$	1

2.2. Construction method

In this section, we indicate the construction method of the confidence interval of the parameter of interest θ. We use a Fieller-type statistic $F = \hat{p}_1 - \theta\hat{p}_0$ as an estimator of the parameter θ, where $\hat{p}_1 = (X_{11} + X_{10})/N$ and $\hat{p}_0 = (X_{11} + X_{01})/N$. The variance of this statistic and its estimated value is

$$E(F) = E(\hat{p}_1) - \theta E(\hat{p}_0) = 0,$$

$$Var(F) = Var(\hat{p}_1) - \theta^2 Var(\hat{p}_0) - 2\theta Cov(\hat{p}_1,\ \hat{p}_0) = \theta\,(p_1 + p_1)/n.$$

Therefore, the standardized statistic F becomes

$$Z = \frac{F - E(F)}{\sqrt{Var(F)}} = \frac{\hat{p}_1 - \theta\hat{p}_0}{\sqrt{\theta\,(p_{10} + p_{01})/n}}. \tag{1}$$

Furthermore, the Z statistic asymptotically follows standard normal distribution. Then it is expressed

$$P(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) \approx 1 - \alpha,$$

where $z_{1-\alpha}$ is the point $100(1-\alpha)\%$ in the standard normal distribution.

2.2.1. Wald interval

In Expression (1), $\hat{p}_{10}$ and $\hat{p}_{01}$ for the standard errors of the unknown parameter p_{10} and p_{01} may be estimated as

$$\frac{\hat{p}_1 - \theta\hat{p}_0}{\sqrt{\theta(\hat{p}_{10} + \hat{p}_{01})/n}} = \pm z_{\alpha/2},$$

where $\hat{p}_{10} = X_{10}/N$ and $\hat{p}_{01} = X_{01}/N$. Therefore, Desu and Raghavarao [3] indicated a $100(1-\alpha)\%$ confidence interval of the Wald-type statistic as

$$\exp[\ln(x_1/x_0) \pm z_{1-\alpha/2}\{(x_{10} + x_{01})/x_1x_0\}^{1/2}]. \tag{2}$$

2.2.2. Nam and Blackwelder interval

The Wald-type confidence interval for a small sample compared to the general parameter shows that the actual confidence level becomes much lower than the nominal confidence level. To improve this point, a score-type confidence interval is often constructed that is general. However, because the numerator in Expression (1) includes the unknown parameters p_{10}, p_{01}, it is not able to construct the confidence interval simply by solving θ. Hence, Nam and Blackwelder [7] used the maximum likelihood estimators $\tilde{p}_{10}$, $\tilde{p}_{01}$ to express

$$\frac{\hat{p}_1 - \theta\hat{p}_0}{\sqrt{\theta(\tilde{p}_{10} + \tilde{p}_{01})/n}} = \pm z_{\alpha/2},$$

where

$$\tilde{p}_{10} = \frac{-\hat{p}_1 + \theta^2(\hat{p}_0 + 2\hat{p}_{10}) + \sqrt{(\hat{p}_1 - \theta^2\hat{p}_0)^2 + 4\theta^2\hat{p}_{10}\hat{p}_{01}}}{2\theta(\theta+1)},$$

$$\tilde{p}_{01} = \theta\tilde{p}_{10} - (\theta - 1)(1 - \hat{p}_{00})$$

and $\hat{p}_{00} = X_{00}/N$. As shown here, a $100(1-\alpha)\%$ confidence interval of the score-type can be constructed.

2.2.3. Bonett interval

Newcombe [8] proposed a hybrid-type confidence interval which includes the combined confidence intervals of the differences between two individual sample proportions and the Wilson type. The Bonett-type confidence interval is practical, because it applies the method indicated previously to the estimation of the ratio of marginal proportions. Equation (1) may be expressed as

$$\left(\frac{\ln(\hat{p}_1') - kz_{1-\alpha/2}SE\{\ln(\hat{p}_1')\}}{\ln(\hat{p}_0') + kz_{1-\alpha/2}SE\{\ln(\hat{p}_0')\}}, \frac{\ln(\hat{p}_1') + kz_{1-\alpha/2}SE\{\ln(\hat{p}_1')\}}{\ln(\hat{p}_0') - kz_{1-\alpha/2}SE\{\ln(\hat{p}_0')\}}\right),$$

where

$$\hat{p}_1' = \frac{X_{11} + X_{10}}{X_{11} + X_{10} + X_{01}}, \quad \hat{p}_0' = \frac{X_{11} + X_{01}}{X_{11} + X_{10} + X_{01}},$$

$$k = \frac{SE\{\ln(\hat{p}_1') - \ln(\hat{p}_0')\}}{SE\{\ln(\hat{p}_1')\} + SE\{\ln(\hat{p}_0')\}}.$$

The proposed $100(1-\alpha)\%$ confidence interval for θ replaces the Wald interval estimates for $p_0' = E(\hat{p}_0')$ and $p_1' = E(\hat{p}_1')$ with Wilson interval estimates. The Wilson interval for p'_j is

$$\frac{(2X_j + kz_{1-\alpha/2}) \pm kz_{1-\alpha/2}\sqrt{(kz_{1-\alpha/2})^2 + 4X_j(1-\hat{p}'_j)}}{2(X_{00} + X_{10} + X_{01} + kz_{1-\alpha/2})}.$$

In order to use this confidence interval, it is necessary to estimate *k*. To estimate $SE\{\ln(\hat{p}'_j)\}$ and $SE\{\ln(\hat{p}_1') - \ln(\hat{p}_0')\}$, we use

$$\hat{S}E\{\ln(\hat{p}'_j)\} = \sqrt{\frac{1-\tilde{p}_j}{(X_{11} + X_{10} + X_{01} + 2)\tilde{p}_j}},$$

$$\hat{S}E\{\ln(\hat{p}_1') - \ln(\hat{p}_0')\} = \sqrt{\frac{X_{10} + X_{01} + 2}{(X_1 + 1)(X_0 + 1)}},$$

where $\tilde{p}_1 = \dfrac{X_{11} + X_{10} + 1}{X_{11} + X_{10} + X_{01} + 2}$ and $\tilde{p}_0 = \dfrac{X_{11} + X_{01} + 1}{X_{11} + X_{10} + X_{01} + 2}$.

2.2.4. Agresti interval

Agresti and Min [1] proposed a confidence interval by adding *k* times of a hypothetical trial to the Wald interval to generate a confidence interval of the difference of marginal population proportions. In this paper, we propose a new confidence interval by applying this improvement to the confidence interval of the ratio of the marginal population proportion. Using k times the hypothetical trial, the number of samples becomes $N' = N + k$. Furthermore, it is standard to set $X_{ij}^{'} = X_{ij} + k/4$ as each probability variance. Using these variances, the Agresti-type $100(1-\alpha)\%$ confidence interval is constructed by solving θ of

$$\frac{p_1' - \theta p_0'}{\sqrt{\theta(p_{10}' + p_{01}')/(n+k)}} = \pm z_{\alpha/2},$$

where

$$p_1' = (X_1 + (k/2))/(n+k), \quad p_0' = (X_0 + (k/2))/(n+k),$$

$$p_{10}' = (X_{10} + (k/4))/(n+k) \text{ and } p_{01}' = (X_{01} + (k/4))/(n+k).$$

3. Results

In this section, we indicate the confidence interval of the results of the actual clinical test. In addition, we compare and verify their performance under the two assumed conditions. Table 2 shows the results of a cross-over study conducted by Jones and Kenward [4], where a low dose and a high dose were administered to 86 women with dysmenorrhea. Table 3 indicates the confidence interval for each of the data in Table 2.

Table 2. Result of cross-over study

	High Dose		
Low Dose	Success	Failure	Total
Success	53	8	61
Failure	16	9	25
Total	69	17	86

Table 3. Each confidence interval for the data in Table 2

	95% Confidence interval			99% Confidence interval		
Method	Lower	Upper	Upper-lower	Lower	Upper	Upper-lower
Wald	0.762	1.025	0.263	0.728	1.074	0.347
Score	0.751	1.027	0.275	0.708	1.083	0.375
Bonett	0.761	1.028	0.267	0.724	1.080	0.356
Agresti + 1	0.763	1.026	0.264	0.728	1.076	0.348
Agresti + 2	0.862	1.193	0.330	0.819	1.256	0.436
Agresti + 4	0.862	1.193	0.330	0.819	1.255	0.436

4. Discussion

As shown in the results, the Agresti + 1 interval, which adds k times the hypothetical trial to the Wald interval, is narrow. Furthermore, the Agresti + 2 interval and the Agresti + 4 interval are wider and shift above other confidence intervals.

Table 4 indicates the coverage probability (CP) and median length (MEL) of the 95% confidence interval, when each parameter is set as $\theta = 0.9$, $p_1 = 0.63$, $p_0 = 0.7$, $p_{11} = 0.6$ for $n = 20, 50$. Table 5 shows examples of the case of low marginal probability.

According to these results, we found that the CP of the Agresti + 1 interval and the Agresti + 2 intervals are comparatively close to the nominal confidence level. Furthermore, we found that their MEL is narrow.

Table 4. $\theta = 0.9,\ p_1 = 0.63,\ p_0 = 0.7,\ p_{11} = 0.6$

	$N = 20$		$N = 50$	
Method	CP	MEL	CP	MEL
Wald	0.876	1.144	0.940	0.698
Score	0.987	1.233	0.947	0.722
Bonett	0.980	1.148	0.955	0.701
Agresti + 1	0.910	1.138	0.947	0.695
Agresti + 2	0.979	1.115	0.956	0.691
Agresti + 4	0.996	1.078	0.969	0.683

Table 5. $\theta = 1.2,\ p_1 = 0.144,\ p_0 = 0.12,\ p_{11} = 0.05$

	$N = 20$		$N = 50$	
Method	CP	MEL	CP	MEL
Wald	0.854	1.144	0.961	0.698
Score	0.903	1.233	0.952	0.722
Bonett	0.906	1.130	0.958	0.700
Agresti +1	0.992	1.138	0.967	0.695
Agresti +2	0.991	1.115	0.973	0.691
Agresti +4	0.996	1.078	0.977	0.683

5. Conclusion

In this paper, we proposed a construction method of a simple improvement interval that extends the improvement of Agresti. Using examples that apply to the results of the actual clinical test, the Agresti + 1 interval added k times the hypothetical trial to the Wald interval, and became narrow. Furthermore, according to simulation results of CP and MEL, we found that the CPs of Agresti + 2 is comparatively close to the nominal confidence level and that their MEL is narrow. As show here, we found that the new proposed confidence intervals are simple and have comparatively high performance.

References

[1] A. Agresti and Y. Min, Simple improved confidence intervals for comparing matched proportions, Statistics in Medicine 24 (2005), 729-740.

[2] D. G. Bonett and R. M. Price, Confidence intervals for a ratio of binomial proportions based on paired data, Statistics in Medicine 25 (2006), 3039-3047.

[3] M. M. Desu and D. Raghavarao, Nonparametric Statistical Methods for Complete and Censored Data, Chapman & Hall/CRC, Boca Raton, FL, 185, 2004.

[4] B. Jones and M. G. Kenward, Modelling binary data from a three-period cross-over trial, Statistics in Medicine 6 (1987), 555-564.

[5] P. A. Lachenbruch and C. J. Lynch, Assessing screening tests: extensions of McNemar's test, Statistics in Medicine 17 (1998), 2207-2217.

[6] C. J. Lloyd, A new exact and more powerful unconditional test of no treatment effect from binary matched pairs, Biometrics 64 (2008), 716-723.

[7] J. Nam and W. C. Blackwelder, Analysis of the ratio of marginal probabilities in a matched-pair setting, Statistics in Medicine 21 (2002), 689-699.

[8] R. G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, Statistics in Medicine 17 (1998), 857-872.

[9] T. Tango, Equivalence test and confidence interval for the difference in proportions for the paired-sample design, Statistics in Medicine 17 (1998), 891-908.