Statistics and Its Interface Volume 5 (2012) 443–449

A new non-inferiority test based on Bayesian estimation in matched-pairs design

Weiquan Chen and Hua Jin*

Non-inferiority of one treatment to another is a common issue in medical research. In this paper, a new test using an approximate p-value based on Bayesian estimation is proposed. Our test is based on only one point of the two-dimension nuisance parameter space for accuracy improvement and computational purposes. The sizes and powers of our test are considered. Simulation results suggest that our test can control the type I error rates with reasonable powers while the asymptotic normal test cannot for most cases. In comparison to Sidik's exact test, our test is much easier to implement.

AMS 2000 subject classifications: Primary 60K35; secondary 60K35.
Keywords and phrases: Matched-pairs sample, Test of non-inferiority, Bayesian estimation, Approximate p-value, Type I error rates of tests, Powers of tests.

1. INTRODUCTION

Tests for non-inferiority of one treatment to another have been a subject of interest in biostatistical research. Lu and Bean (1995) [1] first proposed test statistics and sample size formulae based on the McNemar test for comparison of non-inferiority of sensitivities. Nam (1997) [2] and Tango (1998) [3] introduced the asymptotic normal test. Hsueh et al (2001) [4] proposed two exact unconditional tests for non-inferiority based on the standard p-value [5]. Sidik (2003) [6] also proposed two exact unconditional tests of non-inferiority based on both standard [5] and confidence interval p-values [7] for a 2×2 matched-pairs sample. Lloyd (2008) [8] suggested a more powerful exact test, but the test is rather complex in computation. In this paper, we carry on the idea of Storer and Kim (1990) [9] and Seung-Ho (2000) [10] to estimate the unknown nuisance parameters to improve the accuracy [11], but we will use Bayesian estimation instead because, combining prior knowledge, Bayesian estimator may perform better when sample sizes are small.

Supposing there are n patients, we consider n pairs of matched samples, (X_{1j}, X_{0j}) for $j = 1, 2, \ldots, n$, where the X_{1j} and X_{0j} are binary responses (1 for a positive and 0 for a negative treatment result) of the new and standard treatments for the jth pair. The four resulting types of matched observations and probabilities can be displayed as follows:

New Treatment	Standard Treatment		Total
	1	0	
1	$x_{11}(p_{11})$	$x_{10}(p_{10})$	$x_{1+}(p_1)$
0	$x_{01}(p_{01})$	$x_{00}(p_{00})$	$x_{0+}(1-p_1)$
Total	$x_{+1}(p_0)$	$x_{+0}(1-p_0)$	$n(1)$

The random cell counts $\{x_{ij}, i; j = 0; 1\}$ follow a multinomial distribution with $\{p_{ij}, i; j = 0; 1\}$ as distribution parameters. Under this model structure, the positive (or response) rate of the new treatment is $p_1 = p_{11} + p_{10}$, and the positive rate of the standard is $p_0 = p_{11} + p_{01}$.

We say that the new treatment is non-inferior to the standard if $p_0 - p_1 \leq \delta_1$, and inferior to the standard if $p_0 - p_1 \geq \delta_0$, where $\delta_0 (> 0)$ and $\delta_1 (< \delta_0)$ are the specific threshold values of inferiority and non-inferiority. To test non-inferiority of the new treatment to the standard, we need to construct the statistical framework as follows:

$$(1) \qquad H_0 : p_0 - p_1 \geq \delta_0 \text{ versus } H_1 : p_0 - p_1 \leq \delta_1$$

The objective of our study is to present a new test for non-inferiority based on restricted Bayesian estimation using an approximate p-value, which is easy to implement, and less conservative and more powerful than the exact test of Sidik [6]. The remainder of the paper is structured as follows. Section 2 briefly reviews the asymptotic normal test of Nam [2] and Tango [3] and the exact unconditional test of Sidik [6], and then presents our new test of this study. In section 3, we carry out simulation studies to compare our test with the asymptotic normal test [2, 3] and the exact test of Sidik [6]. In Section 4, an example is presented for illustration. Conclusions of the results of this paper are presented in the last section.

2. TESTING STATISTICS

Nam [2] and Tango [3] introduced the asymptotic normal test statistic based on the restricted maximum likelihood estimate (RMLE) of p_{10} under the constraint $p_{01} - p_{10} = \delta_0$ is

$$Z(x_{01}, x_{10}) = \frac{x_{01} - x_{10} - n\delta_0}{\sqrt{n(\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2)}}$$

Here $\tilde{p}_{01} = \tilde{p}_{10} + \delta_0$, and $\tilde{p}_{10} = (-b + \sqrt{b^2 - 4ac})/(2a)$ is the RMLE of p_{10}, where $a = 2n$, $b = (2n + x_{10} - x_{01})\delta_0 - x_{01} - x_{10}$, and $c = -x_{10}\delta_0(1 - \delta_0)$.

*Corresponding author.

In a large sample, we can use Z to test the hypothesis (1), but the asymptotic test may not work well when sample sizes are small or moderately large. Sidik [6] defined an exact unconditional p-value of Z for testing (1) based on a standard p-value as follows:

$$\begin{aligned}P_Z(x_{01},x_{10}) &= \sup_{0\le p_{10}\le \frac{1-\delta_0}{2}} P_{p_{10}}(Z(X_{01},X_{10})\le Z(x_{01},x_{10}))\\ &= \sup_{0\le p_{10}\le \frac{1-\delta_0}{2}} \sum_{(u,v)\in R_Z} \frac{n!}{u!v!(n-t)!}(p_{10}+\delta_0)^u\\ &\quad \times p_{10}^v(1-2p_{10}-\delta_0)^{n-t}\end{aligned}$$

where $t=u+v$, $R_Z=\{(u,v): Z(u,v)\le Z(x_{01},x_{10})\}$.

However, this test PZ is complex to compute because it needs to search the supremum over the interval $[0,\frac{1-\delta_0}{2}]$ of a nuisance parameter p_{10} (It's so called in the sense that p-value of Z cannot be determined with its value unknown). Also, according to our simulation studies, this test PZ is somewhat conservative when the sample size is small or even moderately large. This should not be surprising. In fact, Sidik provides the computations of exact size and power based on the multinomial distribution and thus the sizes may not be exactly equal to (therefore less than) the specified alpha because of the discrete nature of data.

In this paper, we propose a new test using an approximate p-value based on Bayesian estimation. The procedure has mainly two steps:

(1) Get a Bayesian estimator of p_{10};

(2) Compute an approximate p-value of the test Z based on the Bayesian estimator of p_{10}.

First, we assume p_{10} to be uniformly distributed in the interval $[0,\frac{1-\delta_0}{2}]$ from Bayes' principle (in estimating a parameter, one should initially assume that each possible value has equal probability). So the prior probability density function of p_{10} is:

$$\pi(\theta)=\begin{cases}\dfrac{2}{1-\delta_0}, & 0\le\theta\le\dfrac{1-\delta_0}{2};\\ 0, & \text{otherwise.}\end{cases} \tag{2}$$

Therefore, the posterior probability density function of p_{10} is:

$$\begin{aligned}\pi(\theta|x) &= \Bigg(\frac{n!}{x_{01}!x_{10}!(n-x_{01}-x_{10})!}\theta^{x_{10}}(\theta+\delta_0)^{x_{01}}\\ &\quad\times(1-2\theta-\delta_0)^{n-x_{01}-x_{10}}\frac{2}{1-\delta_0}I\left(0\le\theta\le\frac{1-\delta_0}{2}\right)\Bigg)\\ &\quad\Big/\Bigg(\int_0^{\frac{1-\delta_0}{2}}\frac{n!}{x_{01}!x_{10}!(n-x_{01}-x_{10})!}\theta^{x_{10}}(\theta+\delta_0)^{x_{01}}\\ &\quad\times(1-2\theta-\delta_0)^{n-x_{01}-x_{10}}\frac{2}{1-\delta_0}d\theta\Bigg)\end{aligned} \tag{3}$$

where $x=(x_{01},x_{10})$. So we can get a Bayesian estimator of p_{10} as follows:

$$\tilde\theta_B=\int_0^{\frac{1-\delta_0}{2}}\theta\cdot\pi(\theta|x)d\theta \tag{4}$$

Now we define a new approximate p-value of Z for testing (1) using the restricted Bayesian estimation $\tilde\theta_B$ of p_{10} under the constraint $p_{01}-p_{10}=\delta_0$ as follows:

$$\begin{aligned}&P_{BZ}(x_{01},x_{10})\\ &\quad=\sum_{(i,j)\in R_Z}\frac{n!}{i!j!(n-i-j)!}(\tilde\theta_B+\delta_0)^i\tilde\theta_B^j\\ &\quad\times(1-2\tilde\theta_B-\delta_0)^{n-i-j}\end{aligned} \tag{5}$$

where

$$\tilde\theta_B=\frac{1-\delta_0}{2}\cdot\sum_{i=0}^{x_{01}}a_ib_i\Big/\sum_{i=0}^{x_{01}}b_i \tag{6}$$

Here $a_i=\frac{x_{10}+i+1}{n+i+2-x_{01}}$, $b_i=C_{x_{01}}^i(\frac{2\delta_0}{1-\delta_0})^{x_{01}-i}\cdot B(x_{10}+i+1, n-x_{01}-x_{10}+1)$.

We will reject the null hypothesis and conclude that the new treatment is non-inferior to the standard treatment if the proposed approximate p-value of Z is smaller than the chosen significance level α.

3. SIMULATION STUDIES

In the first simulation, to investigate the significance level of 5%, we repeated the experiment 10,000 times which provided a 95% confidence interval of type I error rate as

Table 1. Sizes of tests based on 10,000 simulations

		$\delta=0.05$			$\delta=0.10$		
n	p_{10}	Z	PZ	PBZ	Z	PZ	PBZ
15	0.05	0.0410	0.0105	0.0105	**<u>0.0573</u>**	0.0237	0.0237
	0.22	**<u>0.0580</u>**	0.0437	0.0437	**<u>0.0586</u>**	0.0450	0.0450
	0.26	**<u>0.0568</u>**	0.0428	0.0428	0.0542	0.0464	0.0464
	0.30	**<u>0.0563</u>**	0.0459	0.0459	0.0538	0.0495	0.0495
	0.45	0.0447	0.0447	0.0447	**<u>0.0771</u>**	0.0257	0.0257
20	0.09	**<u>0.0554</u>**	0.0401	0.0401	0.0435	0.0294	0.0435
	0.14	**<u>0.0571</u>**	0.0442	0.0442	0.0485	0.0410	0.0485
	0.28	**<u>0.0561</u>**	0.0448	0.0448	0.0463	0.0363	0.0451
	0.36	**<u>0.0580</u>**	0.0504	0.0504	0.0506	0.0399	0.0408
	0.40	**<u>0.0614</u>**	0.0465	0.0465	0.0490	0.0342	0.0417
30	0.20	**<u>0.0555</u>**	0.0448	0.0502	0.0455	0.0452	0.0452
	0.24	**<u>0.0589</u>**	0.0497	0.0507	0.0464	0.0458	0.0458
	0.37	**<u>0.0562</u>**	0.0470	0.0530	0.0491	0.0418	0.0446
	0.42	0.0509	0.0453	0.0509	**<u>0.0568</u>**	0.0417	0.0512
	0.45	0.0462	0.0389	0.0462	**<u>0.0714</u>**	0.0336	0.0336

Note that cases that the type I error of Z exceeded 5.43% appear in bold underlined.

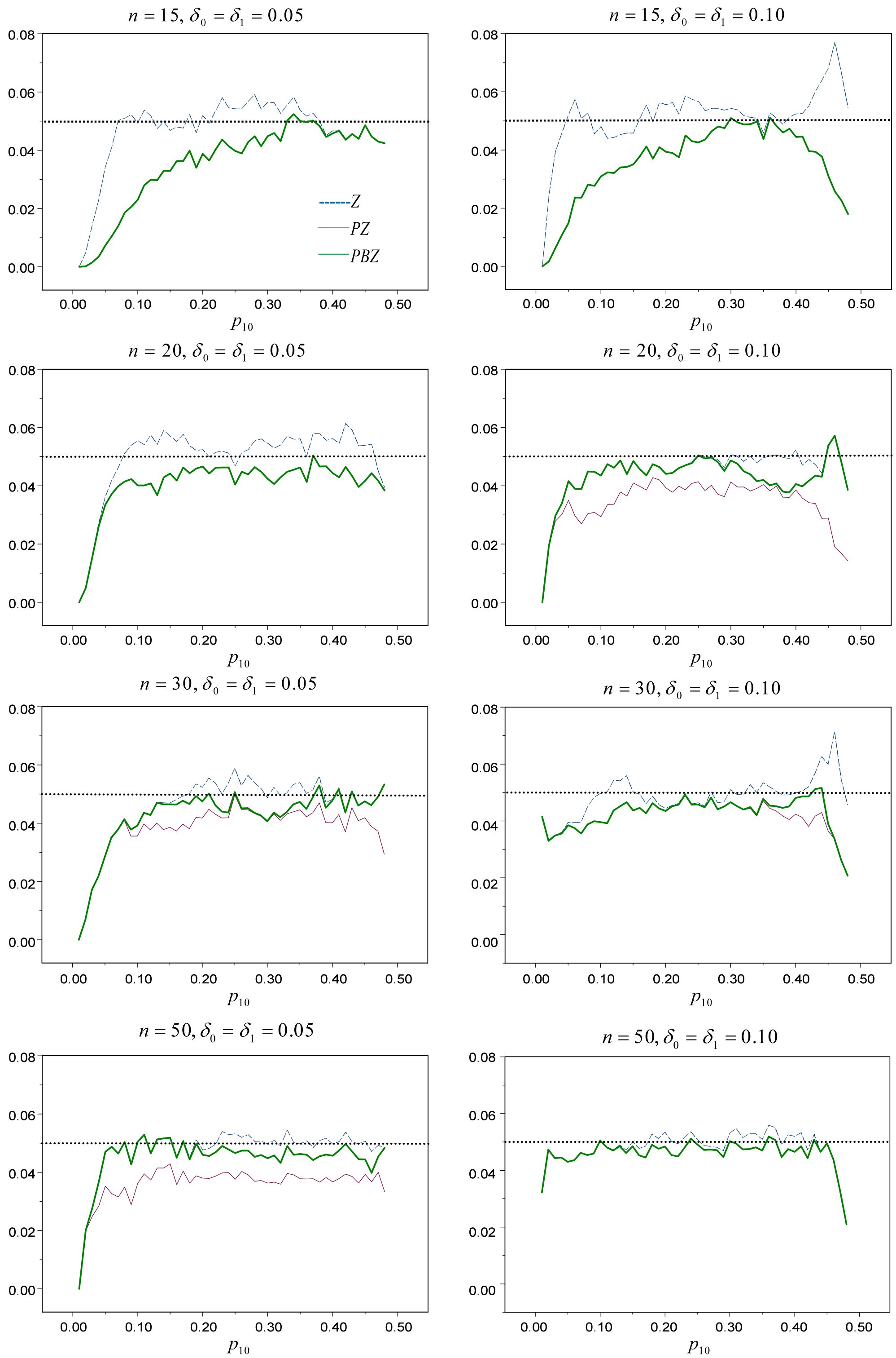

Figure 1. Size functions of tests based on 10,000 simulations.

(4.57%, 5.43%) for a 5% error rate. Figure 1 summarized the results of the comparisons of type I error rates of the asymptotic test Z [2, 3], Sidik's PZ [6] and our new test PBZ. As is shown in Figure 1 and Table 1, cases that type I error of Z was bigger than 5.43% appear. The type I error rates of our new test PBZ and Sidik's PZ test seemed the same when the sample size was as small as 15 or 20, but when the sample size was average or moderately large, such as $n = 30$ or 50, the type I error rate of our test PBZ was closer to 5% than that of Sidik's PZ test.

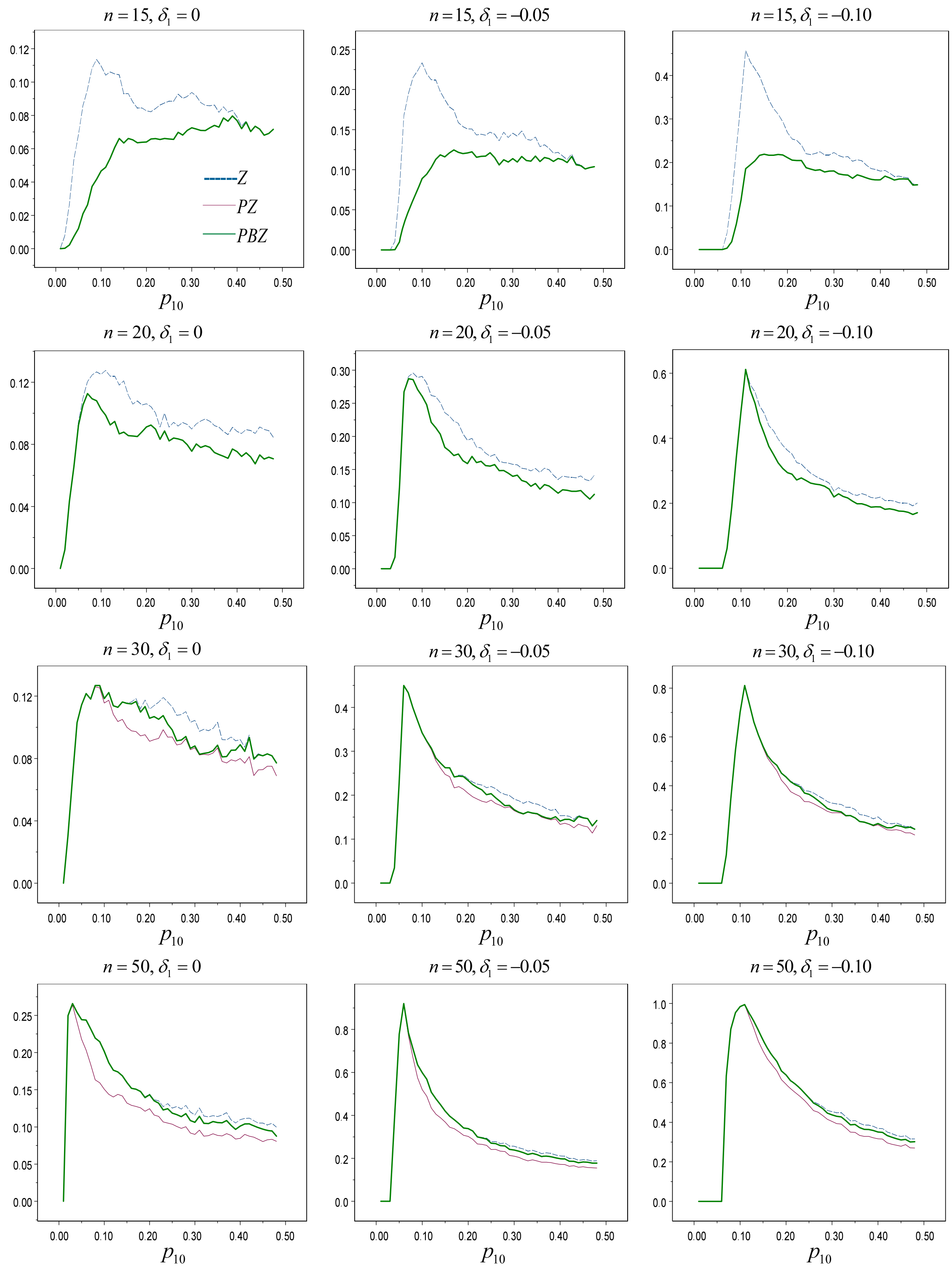

Figure 2. Power functions of the tests for $\delta_0 = 0.05$ based on 10,000 simulations.

In the second simulation experiment, we studied the powers of Z, PZ and PBZ. As is shown in Figure 2 and Table 2, the powers of PZ and PBZ seemed the same when the sample size was as small as 15 or 20, but when the sample size was average or moderately large, such as $n = 30$ or 50, our test PBZ was more powerful than Sidik's PZ test.

In the third simulation experiment, we evaluated the MSEs of $\tilde{\theta}_B$ and $\tilde{p}_{10}$. As is shown in Figure 3, the MSE (i.e. Mean Square Error) of $\tilde{\theta}_B$ seemed smaller than that of $\tilde{p}_{10}$ in mass.

Table 2. Statistical powers of tests for $\delta_0 = 0.05$ based on 10,000 simulations

		$\delta_1 = 0$			$\delta_1 = -0.05$			$\delta_1 = -0.10$		
n	p_{10}	Z	PZ	PBZ	Z	PZ	PBZ	Z	PZ	PBZ
15	0.10	0.1041	0.0488	0.0488	0.2197	0.0944	0.0944	0.4569	0.1858	0.1858
	0.15	0.0930	0.0661	0.0661	0.1784	0.1209	0.1209	0.3431	0.2166	0.2166
	0.25	0.0885	0.0656	0.0656	0.1443	0.1151	0.1151	0.2207	0.1820	0.1820
20	0.10	0.1276	0.0987	0.0987	0.2805	0.2479	0.2479	0.6117	0.6117	0.6117
	0.15	0.1117	0.0856	0.0856	0.2308	0.1782	0.1782	0.4387	0.3751	0.3751
	0.25	0.0941	0.0841	0.0841	0.1727	0.1575	0.1575	0.2837	0.2590	0.2590
30	0.10	0.1223	0.1174	0.1223	0.3236	0.3220	0.3236	0.8106	0.8106	0.8106
	0.15	0.1162	0.0977	0.1150	0.2627	0.2422	0.2623	0.5230	0.5136	0.5230
	0.25	0.1076	0.0887	0.0913	0.2147	0.1811	0.1940	0.3698	0.3265	0.3534
50	0.10	0.1863	0.1438	0.1863	0.5678	0.489	0.5678	0.9944	0.9944	0.9944
	0.15	0.152	0.129	0.152	0.3963	0.3445	0.3963	0.7736	0.7188	0.7736
	0.25	0.1274	0.1006	0.1165	0.2779	0.2417	0.2689	0.5023	0.4587	0.4969

4. AN EXAMPLE

Consider the same example data of comparing two diagnostic procedures MRI and CTAP for liver lesions used by Hsueh et al. [4] and Sidik [6]. Let $\delta_0 = 0.05$, then the asymptotic test statistic Z is -1.1291 with p-value 0.1294. Sidik [6] gave the p-value of PZ, $P_Z = 0.1441$. Now, we give the p-value of PBZ, $P_{BZ} = 0.1415$. If $\delta_0 = 0.10$ is used, then $Z = -1.8359$ with p-value 0.0332, $P_Z = 0.0444$ and $P_{BZ} = 0.0335$.

5. CONCLUSIONS

In this paper, we propose a new non-inferiority test using an approximate p-value based on a restricted Bayesian estimation for a 2×2 matched-pairs sample. Our new test is based on only one point of the two-dimension nuisance parameter space and Bayesian estimation in order to improve both statistical accuracy and computation time.

The sizes and powers of the existing asymptotic test Z, Sidik's PZ test and our test PBZ were discussed. According to our simulation studies, we found that the MSE of our restricted Bayesian estimation was smaller in mass. We also found that our test PBZ and Sidik's PZ test seemed the same when the sample size was small, such as $n = 15$ or 20. But Sidik's PZ test seemed conservative for data with average or even moderately large sample size, while our test PBZ performed better.

The asymptotic normal test Z [2, 3] can not control the type I error rates well when sample sizes are small or moderately large. Sidik's PZ test [6] and Chris J. Lloyd's exact test [8] are complex in computation, while our test PBZ is much easier to compute and it costs less time to implement. Because the PZ test searches the supremum over the boundary of H_0, its average time will be as about hundredfold as that of our approach.

APPENDIX A. APPENDIX SECTION

Now give the proof of (6).

According to (3), (4) and (5), we have

$$\tilde{\theta}_B = \left(\int_0^{\frac{1-\delta_0}{2}} \frac{n!}{x_{01}!x_{10}!(n-x_{01}-x_{10})!} \theta^{x_{10}+1}(\theta+\delta_0)^{x_{01}} \times (1-2\theta-\delta_0)^{n-x_{01}-x_{10}} \frac{2}{1-\delta_0} d\theta \right) \Big/ \left(\int_0^{\frac{1-\delta_0}{2}} \frac{n!}{x_{01}!x_{10}!(n-x_{01}-x_{10})!} \theta^{x_{10}}(\theta+\delta_0)^{x_{01}} \times (1-2\theta-\delta_0)^{n-x_{01}-x_{10}} \frac{2}{1-\delta_0} d\theta \right).$$

Then we let $t = \frac{2}{1-\delta_0}\theta$, and we know that $(t+\frac{2\delta_0}{1-\delta_0})^{x_{01}} = \sum_{i=0}^{x_{01}} C_{x_{01}}^i t^i (\frac{2\delta_0}{1-\delta_0})^{x_{01}-i}$, so we have

$$\tilde{\theta}_B = \frac{1-\delta_0}{2} \left(\int_0^1 \left[\sum_{i=0}^{x_{01}} C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0}\right)^{x_{01}-i} t^{x_{10}+1+i} \times (1-t)^{n-x_{01}-x_{10}} \right] dt \right) \Big/ \left(\int_0^1 \left[\sum_{i=0}^{x_{01}} C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0}\right)^{x_{01}-i} t^{x_{10}+i} \times (1-t)^{n-x_{01}-x_{10}} \right] dt \right)$$
$$= \frac{1-\delta_0}{2} \left(\sum_{i=0}^{x_{01}} \left[C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0}\right)^{x_{01}-i} \int_0^1 t^{x_{10}+1+i} \times (1-t)^{n-x_{01}-x_{10}} dt \right] \right) \Big/ \left(\sum_{i=0}^{x_{01}} \left[C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0}\right)^{x_{01}-i} \int_0^1 t^{x_{10}+i} \right. \right.$$

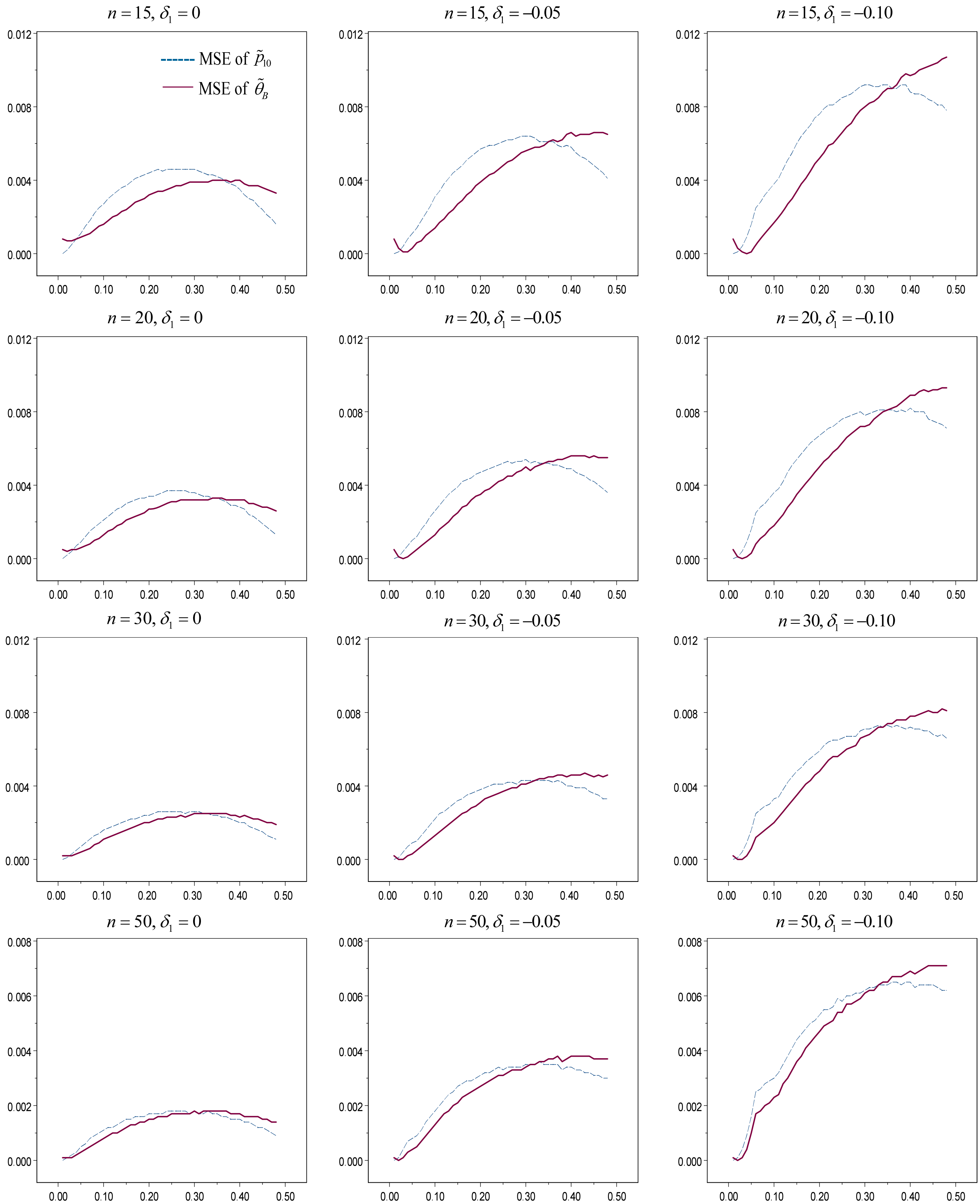

Figure 3. MSEs of $\tilde{\theta}_B$ and $\tilde{p}_{10}$ for $\delta_0 = 0.05$ based on 10,000 simulations.

$$\times (1-t)^{n-x_{01}-x_{10}} dt \Big] \Big).$$

And we know that $n - x_{01} - x_{10} + 1 > 0$, so

$$\begin{aligned}
\tilde{\theta}_B = \frac{1-\delta_0}{2} \Bigg(& \sum_{i=0}^{x_{01}} C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0} \right)^{x_{01}-i} \\
& \times B(x_{10}+i+2, n-x_{01}-x_{10}+1) \Bigg) \\
/ \Bigg(& \sum_{i=0}^{x_{01}} C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0} \right)^{x_{01}-i} \\
& \times B(x_{10}+i+1, n-x_{01}-x_{10}+1) \Bigg) \\
= \frac{1-\delta_0}{2} \Bigg(& \sum_{i=0}^{x_{01}} \frac{x_{10}+i+1}{n+i+2-x_{01}} \cdot C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0} \right)^{x_{01}-i} \\
& \times B(x_{10}+i+1, n-x_{01}-x_{10}+1) \Bigg) \\
/ \Bigg(& \sum_{i=0}^{x_{01}} C_{x_{01}}^i \left(\frac{2\delta_0}{1-\delta_0} \right)^{x_{01}-i} \\
& \times B(x_{10}+i+1, n-x_{01}-x_{10}+1) \Bigg).
\end{aligned}$$

Therefore, (6) is proved.

ACKNOWLEDGEMENTS

We would like to thank the two referees and editors for providing helpful comments that led to greater clarity of this study.

Received 8 January 2011

REFERENCES

[1] Lu, Y. and Bean, J. A. (1995). On the sample size for studies of bioequivalence based upon McNemar's test. *Statistics in Medicine* **14**(16) 1831–1839.

[2] Nam, J. (1997). Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics* **53** 1422–1430.

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

[4] Hsueh, H. M., Liu, J. P. and Chen, J. J. (2001). Unconditional exact tests for equivalence or noninferiority for paired binary endpoints. *Biometrics* **57** 478–483. MR1838117

[5] Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics*. Basic Ideas and Selected Topics. Holden-Day, California. MR0443141

[6] Sidik, K. (2003). Exact unconditional tests for testing non-inferiority in matched-pairs design. *Statistics in Medicine* **22** 265–278.

[7] Berger, R. L. and Boos D. D. (1994). P values maximized over a confidence set for the nuisance parameters. *Journal of the American Statistical Association* **89** 1012–1016. MR1294746

[8] Lloyd, C. J. and Moldovan, M. V. (2008). A more powerful exact test of non-inferiority from binary matched-pairs data. *Statistics in Medicine* **27** 3540–3549. MR2523970

[9] Storer, B. E. and Kim, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association* **85** 146–155.

[10] Kang, S. H. and Chen, J. J. (2000). An approximate unconditional test of non-inferiority between two proportions. *Statistics in Medicine* **19** 2089–2100.

[11] Skipka, G., Munk, A. and Freitag, G. (2004). Unconditional exact tests for the difference of binomial probabilities contrasted and compared. *Computational Statistics and Data Analysis* **47** 757–773. MR2101550

Weiquan Chen
School of Mathematical Sciences
South China Normal University
Guangzhou
China
E-mail address: morningcharming@126.com

Hua Jin
School of Mathematical Sciences
South China Normal University
Guangzhou
China
E-mail address: jinh1@163.com