Journal of Modern Applied Statistical Methods

Volume 12 | Issue 2 Article 15

11-1-2013

Comparison of Three Calculation Methods for a Bayesian Inference of P(π1 > π2)

Yohei Kawasaki
National Center for Global Health and Medicine, Tokyo, Japan, yk_sep10@yahoo.co.jp

Asanao Shimokawa
Tokyo University of Science, Tokyo, Japan

Etsuo Miyaoka
Tokyo University of Science, Tokyo, Japan

DOI: 10.22237/jmasm/1383279240

Journal of Modern Applied Statistical Methods
November 2013, Vol. 12, No. 2, 256-268.

ISSN 1538 − 9472

Comparison of Three Calculation Methods for a Bayesian Inference of $P(\pi_1 > \pi_2)$

Yohei Kawasaki
National Center for Global Health and Medicine
Tokyo, Japan

Asanao Shimokawa
Tokyo University of Science
Tokyo, Japan

Etsuo Miyaoka
Tokyo University of Science
Tokyo, Japan

In Bayesian inference, some researchers have examined the difference of binominal proportions using $\theta = \mathrm{P}\left(\pi_1 > \pi_2 - \Delta_0 | X_1, X_2\right)$, where X_i denote binomial random variable with parameter π_i. An approximate method and the MCMC method are compared with an exact method for θ, and results of actual clinical trials using θ are presented.

Keywords: Binominal proportions; Bayesian inference; MCMC method; hypergeometric series.

Introduction

Statistical inference concerning the difference between two independent binominal proportions is often discussed from the frequency rather than the Bayesian viewpoint. Some researchers have examined significant differences in binominal proportions using the index, $\theta = \mathrm{P}\left(\pi_1 > \pi_2 - \Delta_0 | X_1, X_2\right)$, which indicates the difference in the posterior density for two independent binomial proportions that are assumed to be random variables.

Originally, this index can be shown in the framework of frequency theory to be, $\mathrm{P}(Y_1 > Y_2)$, where Y_1 and Y_2 are random variables. The inference for $\mathrm{P}(Y_1 > Y_2)$ can be observed in various fields. In engineering, it is used in the `stress strength model' to evaluate the reliability of an industrial component (see for instance Kotz, et al. (2003)). In clinical research, it is used as an index for the comparison of two groups given different treatments. In addition, this probability corresponds to the area under the receiver operating characteristic (ROC) curve.

Dr. Kawasaki is a Senior Biostatistician at the National Center for Global Health and Medicine. Email at yk_sep10@yahoo.co.jp. Asanao Shimokawa is a graduate student. Dr. Miyaoka is a professor in the mathematics department.

In medicine, it is used as an index for evaluating the validity of a diagnostic method. Indeed, innumerable studies have been conducted for $\mathrm{P}(Y_1 > Y_2)$ in the framework of frequency theory (See for instance Sen (1960, 1967)). As for research papers on this index, Shirahata (1993), Zhou (2008) and Kawasaki and Miyaoka (2010) have published actively in recent years.

Conversely, there have been a number of studies to apply a construction of $\mathrm{P}(Y_1 > Y_2)$ to the Bayesian framework. Basu (1996) concisely showed the use of the Bayesian approach with respect to hypothesis testing. Berry (1995) using superior binomial proportions, presented a detailed comparison between two binomial proportions assumed to be random variables and presented some interesting examples. Zaslavsky (2009, 2010) applied θ to a one-side hypothesis based on a one-sample situation. Kawasaki and Miyaoka (2012) showed an exact expression for θ, and applied θ to a one-side hypothesis based on a two-sample situation.

There are some pending issues with the above-mentioned method. An approximate method and exact method of θ were adopted only while using a conjugate prior. The drawback of the approximate method is that it occasionally leads to a rough result in a small sample. The drawback of the exact method is that it is slightly complicated. In addition, the exact method requires extensive computing time with a large sample size. Hence, a Markov Chain Monte Carlo (MCMC) method is proposed for θ as a solution to these problems.

Methodology

Let X_1 and X_2 denote binomial random variables for n_1 and n_2 trials with parameters π_1 and π_2, respectively. The conjugate prior density for π_i is a beta distribution with parameters α_i and β_i, where $\alpha_i > 0$, $\beta_i > 0$, and $i = 1, 2$. The proposed posterior density for π_i is

$$g_i\left(\pi_i | X_i\right) = \frac{1}{\mathrm{B}\left(a_i, b_i\right)} \pi_i^{a_i - 1} \left(1 - \pi_i\right)^{b_i - 1} \dagger, \tag{1}$$

where $a_i = \alpha_i + x_i$, $b_i = n_i - x_i + \beta_i$, and $\mathrm{B}\left(a, b\right)$ is the proposed beta function. Let $\pi_{i,\dagger post}$ denote the binomial proportion following the posterior density.

Approximate method for θ

θ can be calculated via an approximation using the standard normal table. Assume that a_i and b_i of the posterior density are large. It is necessary to determine a Z-test statistic. The expected difference in the posterior density and the variance in this difference can be expressed as:

$$\mathrm{E}\left(\pi_{1,post} - \pi_{2,post}\right) = \mu_{1,post} - \mu_{2,post}\dagger, \tag{2}$$

$$\mathrm{V}\left(\pi_{1,post} - \pi_{2,post}\right) = \frac{\mu_{1,post}\left(1-\mu_{1,post}\right)}{a_1+b_1+1} + \frac{\mu_{2,post}\left(1-\mu_{2,post}\right)}{a_2+b_2+1}\dagger, \tag{3}$$

where $\mu_{i,post} = a_i / \left(a_i + b_i\right)$ denotes the posterior mean of π_i. The Z_g-test statistic,

$$Z_g = \frac{\left(\pi_{1,post} - \pi_{2,post}\right) - \mathrm{E}\left(\pi_{1,post} - \pi_{2,post}\right)}{\sqrt{\mathrm{V}\left(\pi_{1,post} - \pi_{2,post}\right)}}\dagger, \tag{4}$$

is approximately distributed as the standard normal distribution. Therefore, the approximate probability of θ is given by

$$\begin{aligned}\theta &= \mathrm{P}\left(\pi_1 > \pi_2 | X_1, X_2\right) \\ &\approx 1 - \Phi\left(\frac{-\left(\mu_{1,post} - \mu_{2,post}\right)}{\sqrt{\frac{\mu_{1,post}\left(1-\mu_{1,post}\right)}{a_1+b_1+1} + \frac{\mu_{2,post}\left(1-\mu_{2,post}\right)}{a_2+b_2+1}}}\right)\end{aligned} \tag{5}$$

where $\Phi\left(\cdot\right)$ is the cumulative distribution function (CDF) of the standard normal distribution. Thus, the approximate probability can easily be calculated.

Exact method for θ

Kawasaki and Miyaoka (2012) derived the exact expression for θ using the posterior density. The exact expression for θ is

$$\begin{aligned}\theta &= \mathrm{P}\left(\pi_1 > \pi_2 | X_1, X_2\right) \\ &= \frac{\mathrm{B}\left(a_1 + a_2, b_1\right)}{a_2 \mathrm{B}\left(a_1, b_1\right) \mathrm{B}\left(a_2, b_2\right)} \dagger_3 \mathrm{F}_2\left(a_2, 1-b_2, a_1+a_2; 1+a_2, \dagger a_1 + a_2 + b_1; 1\right),\end{aligned} \tag{6}$$

where

$$\dagger_3 \mathrm{F}_2\left(k_1, k_2, k_3; l_1, l_2; 1\right) = \sum_{t=0}^{\infty} \frac{\left(k_1\right)_t \left(k_2\right)_t \left(k_3\right)_t}{\left(l_1\right)_t \left(l_2\right)_t} \frac{1}{t!}, \dagger\dagger k_1 + k_2 + k_3 < l_1 + l_2 \tag{7}$$

is the hypergeometric series, and $\left(k\right)_t$ is the Pochhammer symbol.

MCMC method for θ

A computational procedure for θ using the MCMC method is now introduced. The MCMC method is a means of sampling from a posterior density. A random-walk Metropolis-Hasting algorithm was used as the MCMC Method. Given that the samples come from two independent populations, the posterior joint distribution of π_1 and π_2 is a product of its marginal distributions. For this reason, one can obtain samples from the posterior distribution of $\pi_1 - \pi_2$ by simulating k values from the posterior distribution of π_1 and π_2 using MCMC procedure of SAS, e.g., $\pi_{1,post}^1, \pi_{1,post}^2, \cdots, \pi_{1,post}^k$ and $\pi_{2,post}^1, \pi_{2,post}^2, \cdots, \pi_{2,post}^k \dagger$, respectively. Then, by computing $\pi_{1,post}^1 - \pi_{2,post}^1, \dagger \pi_{1,post}^2 - \pi_{2,post}^2 \dagger, \cdots, \pi_{1,post}^k - \pi_{2,post}^k \dagger$, the simulated values from the posterior distribution of $\pi_{1,post} - \pi_{2,post}$ are obtained. The posterior samples obtained by the MCMC method after the burn-in period are $\delta_1, \delta_2, \cdots, \delta_k$. Let $\Delta_1, \Delta_2, \cdots, \Delta_k$ be independent identically distributed random variables with distribution function *F*. The posterior samples is the observed value of $\Delta_1, \Delta_2, \cdots, \Delta_k$. Note the fact that $\theta = \mathrm{P}(\pi_{1,post} > \pi_{2,post})$ equals $\theta = \mathrm{P}(\pi_{1,post} - \pi_{2,post} > 0)$. Thus, θ can be expressed as,

$$\begin{aligned}\theta &= \mathrm{P}\left(\pi_1 > \pi_2 | X_1, X_2\right) \\ &= \mathrm{P}\left(\pi_1 - \pi_2 > 0 | X_1, X_2\right) \approx 1 - \hat{\mathrm{F}}_k\left(0\right)\end{aligned} \tag{8}$$

where

$$\hat{\mathrm{F}}_{\mathrm{k}}(\mathrm{s}) = \frac{1}{k}\sum_{i=1}^{k} \mathrm{I}(\Delta_{\mathrm{i}} \le \mathrm{s}) \tag{9}$$

and

$$\mathrm{I}(\Delta_i \le \mathrm{s}) = \begin{cases} 1 \dagger if \dagger \Delta_i \le s \\ 0 \dagger if \dagger \Delta_i > s \end{cases} \tag{10}$$

is the empirical distribution function.

Results

Comparison of three methods

Now the probabilities of the three methods for θ are compared. The difference between the sample proportions (horizontal axis) were plotted against the difference between the probabilities of the MCMC and exact methods (vertical axis), as shown in Figures 1, 3, and 5. Similarly, the difference between the sample proportions (horizontal axis) were plotted against the difference between the probabilities of the approximate and exact methods (vertical axis), as shown in Figures 2, 4, and 6. In Figures 1, and 2 consider small sample sizes, i.e., $\mathrm{n}_1 = \mathrm{n}_2 = 5$, 10, 15, and 20. Conversely, in Figures 3 and 4 consider large sample sizes, i.e., $\mathrm{n}_1 = \mathrm{n}_2 =$ 60, 70, 80, and 90. Figures 5 and 6 consider groups of different sample sizes, that is, $\mathrm{n}_1 = 15$, $\mathrm{n}_2 = 5$; $\mathrm{n}_1 = 15$, $\mathrm{n}_2 = 10$; $\mathrm{n}_1 = 15$, $\mathrm{n}_2 = 20$; and $\mathrm{n}_1 = 15$, $\mathrm{n}_2 = 20$. The following were confirmed from the results.

First, the relationship between the difference in the probabilities and the difference in the sample proportions is described. In Figure 1(d) and Figure 3(d), the probability of the MCMC method is more or less equal to that of the exact method when the difference between the sample proportions is 0.8. On the other hand, the difference between the probabilities of the MCMC and exact methods is around 0.01 when the difference between the sample proportions is 0.05. Overall, when the difference between the sample proportions is large, the probabilities of the MCMC and exact methods are roughly equal. In contrast, when the difference between the sample proportions is small, the probability of the MCMC method is

different from that of the exact method. This general pattern is similar for the difference in the probabilities of the approximation and exact methods.

Next, the relationship between the sample size and the difference in the probabilities is described. In Figure 2(a), the difference between the probabilities of the approximate and exact methods is around 0.013 when the difference between the sample proportions is 0.2. For a slightly larger sample size (Figure 2(d)), the difference between the probabilities of the approximate and exact methods is around 0.006 for the same difference between the sample proportions. In addition, there is virtually no difference between the probabilities of the approximate and exact methods when the sample size is further increased, as shown in Figure 4(d). Thus, the sample size influences the accuracy of the probability of the approximate method. It also shows the difference in the probabilities of the MCMC and exact methods. In Figure 1(a), the difference between the probabilities of the MCMC and exact methods is around 0.006 when the difference between the sample proportions is 0.2. For a slightly larger sample size (Figure 2(d)), the difference between the probabilities of the MCMC and exact methods is around 0.005 for the same difference between the sample proportions. Thus, the accuracy of the probability of the MCMC method always remains high even when the sample sizes are small.

Finally, the difference between the probabilities when groups of different sample sizes are considered is investigated. In Figure 2(d), the difference between the probabilities of the approximate and exact methods is around 0.006 when the difference between the sample proportions is 0.2. On the other hand, in Figure 6(d), the difference between the probabilities of the approximate and exact methods is around 0.012 for the same difference between the sample proportions. In both the cases, the total sample size ($n_1 + n_2$) is the same. However, the difference between the probabilities of the approximate and exact methods is slightly greater in the case of groups with different sample sizes. It is also shown the case of the MCMC method. In Figure 1(d), the difference between the probabilities of the MCMC and exact methods is around 0.005 when the difference between the sample proportions is 0.2. On the other hand, in Figure 5(d), the difference between the probability of the MCMC and exact methods is around 0.005 for the same difference between the sample proportions. Therefore, the difference between the probabilities of the MCMC and exact methods is the same regardless of whether the sample sizes are equal or different.

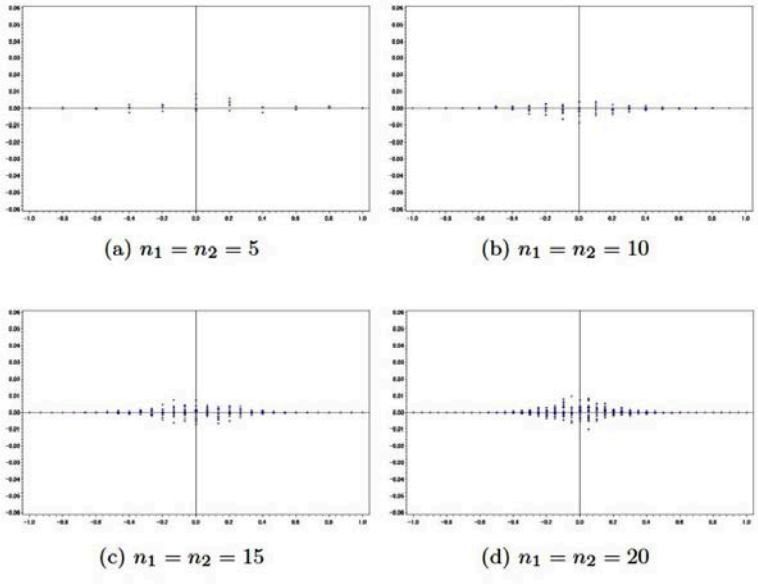

Figure 1: Comparison of the Exact and MCMC Method when sample sizes are small. (vertical axis：Differences of θ in Exact and MCMC method. Prior distribution is Beta(1,1). horizontal axis : Differences of two sample proportions.

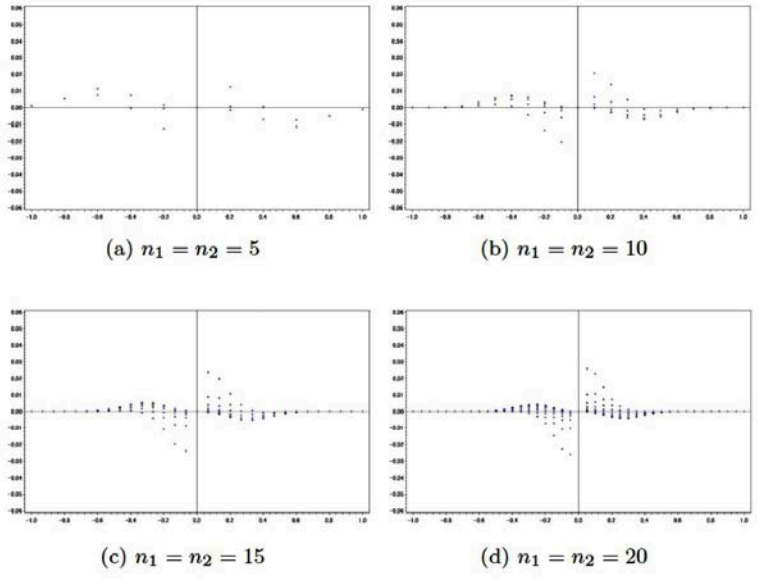

Figure 2: Comparison of the Exact and Approximate method when sample sizes are small. (vertical axis：Differences of θ in Exact and Approximation method. horizontal axis : Differences of two sample proportions.

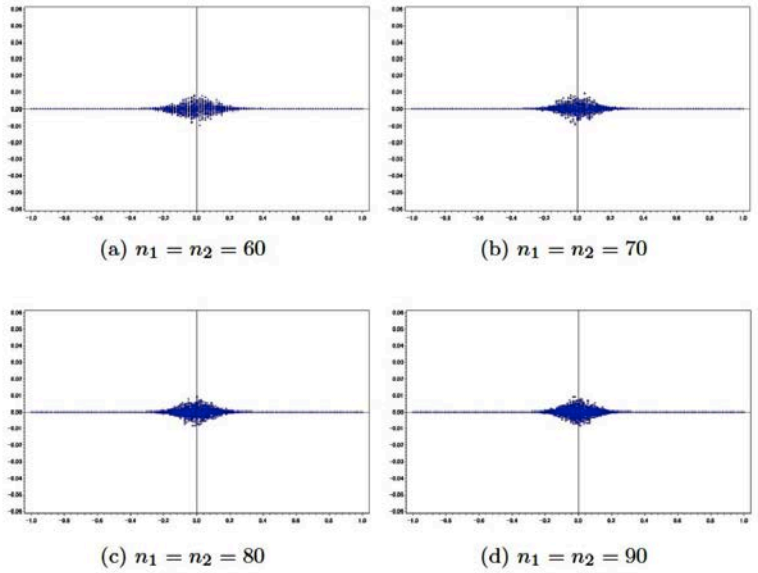

Figure 3: Comparison of the Exact and MCMC Method when sample sizes are large. (vertical axis：Differences of θ in Exact and MCMC method. Prior distribution is Beta(1,1). horizontal axis : Differences of two sample proportions.

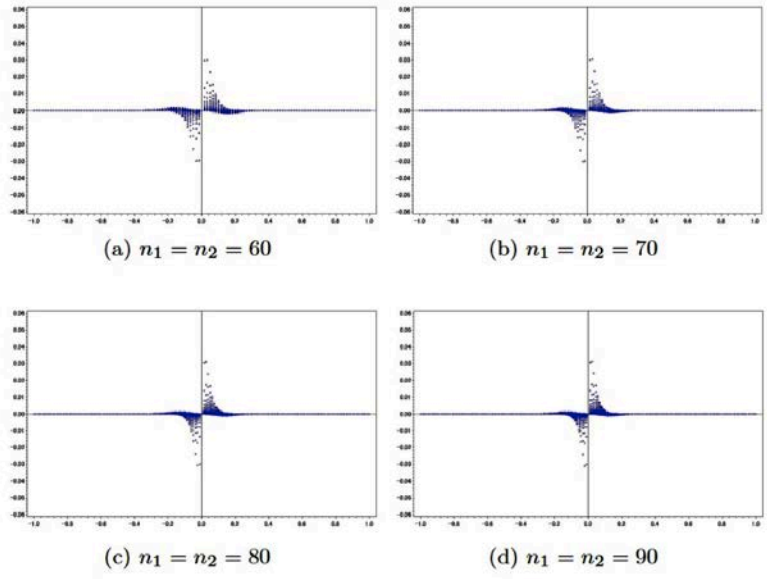

Figure 4: Comparison of the Exact and Approximate method when sample sizes are large. (vertical axis：Differences of θ in Exact and Approximation method. horizontal axis : Differences of two sample proportions.

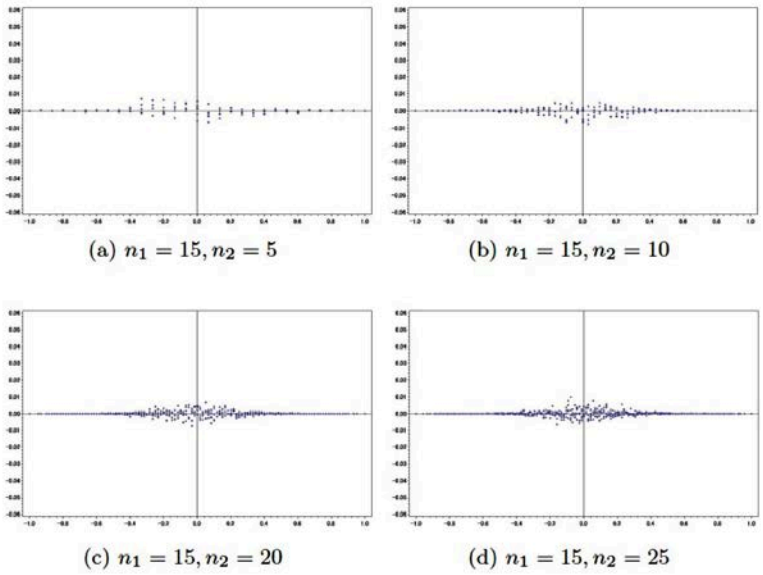

Figure 5: Comparison of the Exact and MCMC Method when sample sizes are unbalanced. (vertical axis : Differences of θ in Exact and MCMC method. Prior distribution is Beta(1,1). horizontal axis : Differences of two sample proportions.

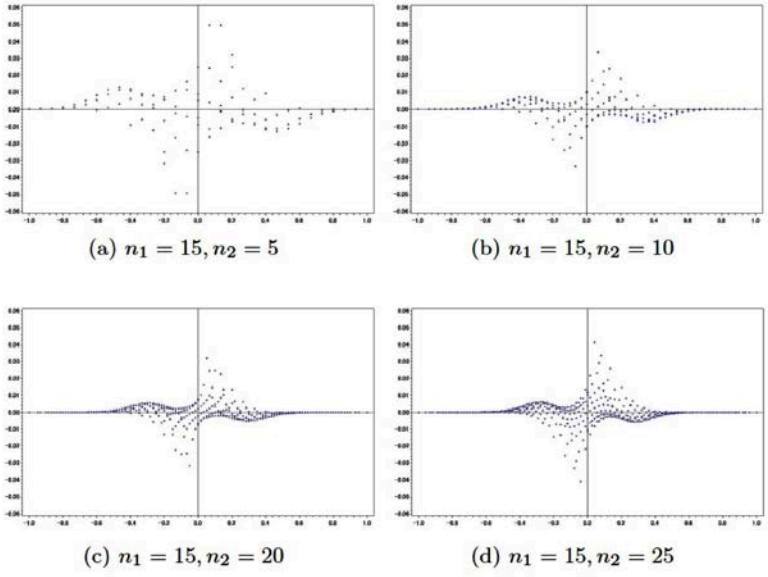

Figure 6: Comparison of the Exact and Approximate method when sample sizes are unbalanced. (vertical axis : Differences of θ in Exact and Approximation method. horizontal axis : Differences of two sample proportions.

Example

Next the utility of θ is illustrated by applying it to the results of clinical trials. A non-informative prior was assumed. Table 1 lists the results of a double-blind, randomized, 41-center study that compares the efficacy of TJN-318 cream with that of Bifonazole (BFZ) cream in the treatment of patients suffering from cutaneous mycosis (TJN-318 Solution Study Group (1992)). The main purpose of this clinical trial was to show that TJN-318 cream is more effective than BFZ cream in the treatment of cutaneous mycosis. The primary end point of this clinical trial is a binary variable. In other words, the patient either recovers or does not recover. In short, the alternative hypothesis is $\pi_1 > \pi_2$. In general, the frequentist approach can be adopted to verify the purpose of the clinical trial via the calculation of a *p*-value. The *p*-value was calculated using the Z-test statistic for the purpose of reference,

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}\left(1-\hat{\pi}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{11}$$

where $\hat{\pi}_i = x_i / n_i$ and $\hat{\pi} = \left(x_1 + x_2\right) / \left(n_1 + n_2\right)$. The values of θ are listed in the rightmost column of Table 1. Consequently, a non-informative prior was adopted, that is, $\alpha_i = \beta_i = 1$ and $i = 1, \dagger 2$. Clearly, θ increases when the *p*-value is low, and $\theta \approx 1$ when the null hypothesis is rejected. Moreover, $\theta \approx 1 - \dagger p$-value.

Next, the results of a double-blind, randomized, phase-3 clinical trial that compares the efficacies of follitropin alpha (hereafter, the study drug) and human menopausal gonadotropin (hereafter, the control drug) in the treatment of patients suffering from no-ovulation-cycle syndrome (from the assessment report of PMDA (2009)) was employed. Table 2 lists the resulting ovulation rate, that is, the primary end point. The Z-test affords a *p*-value of 0.764, which suggests no significant differences. Using the non-informative prior, the approximate probability of θ is obtained as 0.238, whereas the exact probability and the MCMC probability is obtained as 0.237.

Table 1: The result of primary end point in clinical trial for TJN-318 cream vs Bifonazole cream.

		Outcome			θ		
Disease Name	**Drug Name**	**Cure**	**Non-Cure**	***p*-value**	**Approximate**	**Exact**	**MCMC**
Tinea Pedis	TJN-318	110	27	0.264	0.734	0.735	0.735
	BFZ	96	31				
Tine Corporis	TJN-318	70	13	0.417	0.581	0.582	0.581
	BFZ	69	14				
Candidal Intertigo	TJN-318	39	4	0.472	0.531	0.530	0.531
	BFZ	37	4				
Candidal Interdigital	TJN-318	25	2	0.021	0.978	0.977	0.977
	BFZ	23	9				
Ptyriasis Versicolor	TJN-318	59	2	0.236	0.749	0.756	0.757
	BFZ	46	3				

Table 2: The result of primary end point in clinical trial for follitropin alpha vs human menopausal gonadotropin.

	Outcome					θ		
Drug Name	**Cure**	**Non-cure**	**Total**	**Ovulation Ratio**	***p*-value**	**Approximate**	**Exact**	**MCMC**
Study	102	27	129	79.1%	0.764	0.238	0.237	0.237
Control	109	23	132	82.6%				

Conclusion

Three methods for the index $\theta = \mathrm{P}(\pi_1 > \pi_2 \mid X_1, \dagger X_2)$ were presented to determine the probability that the binomial proportion for a study drug is superior to that for a control drug. In particular, a new procedure was described based on the MCMC method. The probabilities of these three methods were compared to test the relative effectiveness of each.

The expression for the exact method was presented, which includes a hypergeometric series. It is speculated that this series causes the decrease in calculation efficiency when the sample size is very large. In addition, hypergeometric series are not built into SAS, which is a statistical software

program frequently used in pharmaceutical development. Therefore, if SAS is used, a calculation program for hypergeometric series must be developed.

It is easy to calculate the probability for using the approximation method. This is an advantage when the approximate probability is used. Conversely, when the difference in the sample proportions is small and the sample sizes are unbalanced, the accuracy the approximation method is poor. That is, the accuracy of the probability of the approximation method depends on the sample size.

This study showed that the accuracy of the MCMC method was greater than that of the approximation method. Moreover, the probability of the MCMC method can be easily calculated using SAS. In addition, it is possible to use the non-conjugate prior for the prior distribution in the MCMC method. The authors consider this as one of the advantages of the MCMC method

References

Basu, S. (1996). Bayesian hypotheses testing using posterior density ratios. *Statistics and Probability Letters, 30,* 79-86.

Berry, D. S. (1995). *Statistics: a Bayesian perspective*. Duxbury Press, New York.

Kawasaki, Y. & Miyaoka, E. (2010). On confidence intervals for $P(X < Y)$. *Journal of the Japanese Society of Computational Statistics, 23,* 1-12.

Kawasaki, Y. & Miyaoka, E. (2012). A Bayesian inference of $P(\pi_1 > \pi_2)$ for two proportions. *Journal of Biopharmaceutical Statistics, 22,* 425-437.

Kotz, S., Lumelskii, Y. & Pensky, M. (2003). *The stress-strength model and its generalizations.* World Scientific Publishing, New Jersey.

Pharmaceuticals & Medical Devices Agency. (2009). *Follitropin alpha of assessment report.* PMDA, Tokyo, (in Japanese).

Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statistical Association Bulletin, 10,* 1-18.

Sen, P. K. (1967). A note on asymptotically distribution-free confidence bounds for $P(X < Y)$, based on two independent samples. *Sankhya Series, A, 29,* 95-102.

Shirahata, S. (1993). Estimate of variance of Wilcoxon-Mann-Whitney statistic. *Journal of the Japanese Society of Computational Statistics, 6,* 1-10.

TJN-318 Solution Study Group. (1992). Double-blind study on TJN-318 cream and Bifonazole cream to the patient with the cutaneous mycosis. *The Nishinihon Journal of Dermatology, 54*, 977—992, (in Japanese).

Zaslavsky, B. G. (2009). Bayes models of clinical trials with dichotomous outcomes and sample size determination. *Statistics in Biopharmaceutical Research, 1,* 149-158.

Zaslavsky, B. G. (2010). Bayesian versus frequentist hypotheses testing in clinical trials with dichotomous and countable outcomes. *Journal of Biopharmaceutical Statistics, 20,* 985-997.

Zhou, W. (2008). Statistical inference for $\mathrm{P}(X < Y)$. *Statistics in Medicine, 27*, 257-279.