American Journal of Biostatistics 1 (1): 23-31, 2010
ISSN 1948-9889


# Comparisons of Test Statistics for Noninferiority Test for the Difference between Two Independent Binominal Proportions

[1,3]Youhei Kawasaki, [2]Fanghong Zhang and [3]Etsuo Miyaoka
[1]Biostatistics Group, Development Division, Department of Data Science,
Mitsubishi Tanabe Pharma Corporation, 2-6 Nihonbashi-Honcho 2-Chome,
Chuo-Ku, Tokyo 103-8405, Japan
[2]Development and Medical Affairs Division, Department of Biomedical Data Sciences,
GlaxoSmithKline K.K., 6-15 Sendagaya 4-Chome, Shibuya-Ku, Tokyo 151-8566, Japan
[3]Department of Mathematics, Tokyo University of Science,
26 Wakamiya-Cho, Shinjuku-Ku, Tokyo 162-0827, Japan

**Abstract: Problem statement:** Noninferiority tests are frequently used in clinical trials to demonstrate that the response for study drugs is not much worse than the response for reference drugs. Several test statistics exist. However, a detailed comparison of those test statistics is not researched. Moreover, a little complex calculation might be necessary in some of those test statistics. **Approach:** In this study, we investigated the performance of the existing test statistics and propose new test statistics. Further, we compare them with existing test methods by means of simulation and devise a suitable technique of using of these test statistics. **Results:** We found that for the proposed test statistics, the actual type I error was close to the nominal level. Further, when the sample size is moderate it is found that, the new test statistics have a little higher power than other test statistics. **Conclusion:** One of the biggest advantages of our method is that it does not require complicated calculations.

**Key words:** On-sided testing, non-inferiority trials, binominal proportions

## INTRODUCTION

A noninferiority test, whose main purpose is to indicate whether the response for study drugs shows clinically not much worse than the response for reference drugs, is often conducted in clinical trials. A noninferiority test is especially, employed to derive the difference between two binomial proportions if the response is an independent binominal. The ICH-E9 guidelines and the European medicines agency guidelines showed the framework of noninferior setting comparisons between treatment groups.

Research pertaining to noninferiority tests for deriving the differences between proportions has been conducted since a long time. However, few theses consider the behavior of test statistics in detail. Moreover, research in this field has been initiated only recently.

Dunnett and Gent (1977) selected an example of noninferiority test from a clinical trial. In their research, an estimator weighted by a noninferiority margin was used for the unknown parameter with test statistics.

Farrington and Manning (1990) proposed three methods for estimating for an unknown parameter in standard error measurement and they recommended using a restricted maximum likelihood estimator, which is a restricted value of the null hypothesis, proposed by Miettinen and Nurminen (1985). The statistical analysis software-power analysis and sample size-can calculate power in eight ways. Almendra-Arao (2009) showed that non-inferiority test sizes are calculated for the difference between two independent proportions based on Z-statistic with pooled variance, for several continuity corrections and the behavior of these test sizes is analyzed. Hirotsu *et al*. (1997) provides confidence intervals that correct skewness and discusses the design issue of the required sample size for the noninferiority test. Dann and Koch (2008) proposed a method of evaluating the noninferiority test on the basis of some confidence intervals. They also showed the relationship between the confidence intervals and the noninferiority test for the difference between two independent binominal proportions.

**Corresponding Author:** Youhei Kawasaki, Biostatistics Group, Development Division, Department of Data Science, Mitsubishi Tanabe Pharma Corporation, 2-6 Nihonbashi-Honcho 2-Chome, Chuo-Ku, Tokyo 103-8405, Japan Tel: +81-3-3241-4198

Zhang *et al*. (2006) proposed a new test statistic for the noninferiority test for ordered categorical data and they expanded their test statistic to the difference of proportions. In this study, we propose a new test statistic, distinct from the method proposed by Zhang *et al*. (2006).

We present a method of deriving an estimator, focusing on the noninferiority test for the difference between two independent binominal proportions and we detect and verify a well-performing estimator in this study.

## MATERIALS AND METHODS

Suppose that $X_1$ and $X_2$ are two independent random variables with a binomial distribution. The first random variable is size $n_1$ and it has a binomial proportion $\pi_1$, denoted as $X_2$~B ($n_1$, $\pi_2$). The second random variable is size $n_2$ and it has a binomial proportion $\pi_2$, denoted as $X_2$~B ($n_2$, $\pi_2$). In this study, we assume that a large binominal proportion is preferred consistently. Here, the hypothesis of the noninferiority test for deriving the difference between proportions is:

$$\begin{aligned} &H_0 : \pi_1 - \pi_2 = -\Delta_0 \\ &H_1 : \pi_1 - \pi_2 > -\Delta_0 \end{aligned} \quad (1)$$

where, the noninferiority margin is $\Delta_0 > 0$. We assume that $\delta = \pi_2 - \pi_2$. The difference between sample proportion, $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_2$, is the estimator for $\delta$, where $\hat{\pi}_1 = X_1 / n_1$ and $\hat{\pi}_2 = X_2 / m_2$. Therefore, the expected value under the null hypothesis is:

$$E(\hat{\delta}) = \pi_1 - \pi_2 = -\Delta_0$$

The variance is:

$$V(\hat{\delta}) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \quad (2)$$

Therefore, the statistic of standardized $\hat{\delta}$ is given by:

$$Z_{CE} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\left(\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}\right)\frac{1-\Delta_0^2}{1-(\hat{\pi}_2 - \hat{\pi}_1)^2}}}$$

This Z-test statistic asymptotically has a standard normal distribution. However, several test statistics have been proposed since the unknown parameter involved in Z-test statistics.

**Pooled variance:** The variance of the estimator under the null hypothesis in a significance test is:

$$V(\hat{\delta}) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\pi(1-\pi)$$

where, the unknown parameter is $\pi = \pi = \pi_2$. This variance is generally known as pooled variance. By replacing the unknown parameter in this variance with the estimator $\hat{\pi}$, the $Z_p$ test statistic is given by:

$$Z_p = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\pi}(1-\hat{\pi})}}$$

where, the estimator for $\pi$ is $\hat{\pi} = \frac{X_1 + X_2}{n_1 + n_2}$.

**Alternative hypothesis variance:** The variance of the Z-test statistic is identical to the one used under the alternative hypothesis. Each maximum likelihood estimator not related to the hypothesis is used for the unknown parameter in the variance. The $Z_W$ test statistic is shown as:

$$Z_W = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}}$$

This is known as the Wald test statistic. Many researchers have indicated in many study that the performance of the Wald statistic suffers when the sample size is small. Further, Munzel and Hsuschke (2003) showed the framework of the noninferiority test for ordered categorical data. When the number of categories is assumed to be two, it is regarded as a problem with regard to the difference between proportions. Hence, this test statistic is derived by extending the method proposed by Munzel and Hsuschke (2003) to the noninferiority test for deriving the difference between proportions.

**Null hypothesis variance 1:** The variance of the noninferiority test under the null hypothesis is:

$$V(\hat{\delta}) = \frac{(\pi_2 - \Delta_0)(1 - \pi_2 + \Delta_0)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

Dunnett and Gent (1977) proposed the estimator:

$$\hat{\pi}_2' = \frac{X_1 + X_2 + n_1\Delta_0}{n_1 + n_2} \quad (3)$$

for the unknown parameter $\pi_2$. By using this estimator, the $Z_D$ test statistic is shown as:

$$Z_D = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\frac{(\hat{\pi}_2' - \Delta_0)(1 - \hat{\pi}_2' + \Delta_0)}{n_1} + \frac{\hat{\pi}_2'(1 - \hat{\pi}_2')}{n_2}}}$$

This is called the Dunnett-Gent test statistic. We suggest that the problem was that the estimator (3) exceeded the limit value 1.

**Null hypothesis variance 2:** Miettinen and Nurminen (1985) constructed a maximum likelihood estimator with a restriction for the binominal proportion $\pi_2$ under the null hypothesis. Farrington and Manning (1990) proposed a test statistic using this estimator. The log-likelihood function under the restricted null hypothesis $\pi_1-\pi_2 = -\Delta_0$ is:

$$\begin{aligned} l(\pi_2) \propto\ & x_1 \ln(\pi_2 - \Delta_0) + (n_1 - x_1)\ln(1 - \pi_2 + \Delta_0) \\ & + x_2 \ln(\pi_2) + (n_2 - x_2)\ln(1 - \pi_2) \end{aligned}$$

The solution $\pi_2$, which maximizes this function is given by solving the following cubic equation:

$$a\pi_2^3 + b\pi_2^2 + c\pi_2 + d = 0$$

Where:

$$a = n_1 + n_2$$
$$b = -\{n_1 + n_2 + x_1 + x_2 + \Delta_0(n_1 + 2n_2)\}$$
$$c = n_2\Delta_0^2 + \Delta_0(2x_2 + n_1 + n_2) + x_1 + x_2$$
$$d = -x_2\Delta_0(1 + \Delta_0)$$

Therefore, the maximum likelihood estimator is:

$$\tilde{\pi}_2 = 2u\cos(w) - b/3a$$

Where:

$$w = \{\pi + \mathrm{Cos}^{-1}(v/u^3)\}/3$$
$$v = (b/3a)^3 - bc/6a^2 + d/2a$$
$$u = \mathrm{sign}(v)\sqrt{(b/3a)^2 - c/3a}$$

Using this restricted maximum likelihood estimator, the $Z_F$ test statistic can be shown as:

$$Z_F = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\frac{(\tilde{\pi}_2 - \Delta_0)(1 - \tilde{\pi}_2 + \Delta_0)}{n_1} + \frac{\tilde{\pi}_2(1 - \tilde{\pi}_2)}{n_2}}}$$

**Null hypothesis variance 3:** Zhang *et al.* (2006) proposed a new test statistic for noninferiority test in ordered categorical data. They extended it to derive the difference between proportions and introduce the $Z_C$ test statistic as:

$$Z_C = \frac{-(\hat{p}_1 - p_{10})}{\sqrt{\left(\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{01}^2}{n_2}\right)\frac{p_{10}(1 - p_{10})}{\sigma_{00}^2}}}$$

Where:

$$\hat{p}_1 = \frac{1}{2} - \frac{1}{2}(\hat{\pi}_1 - \hat{\pi}_2)$$
$$p_{10} = \frac{1}{2} - \frac{1}{2}(\Delta_0)$$
$$\sigma_{00}^2 = \frac{1}{4}(1 - (\pi_1 - \pi_2)^2)$$
$$\sigma_{10}^2 = \frac{1}{4}\pi_1(1 - \pi_1)$$
$$\sigma_{01}^2 = \frac{1}{4}\pi_2(1 - \pi_2)$$

Using each maximum likelihood estimator for the unknown parameter in the $Z_C$ test statistic, the $Z_{CE}$ statistic is defined by:

$$Z_{CE} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\left(\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}\right)\frac{1 - \Delta_0^2}{1 - (\hat{\pi}_2 - \hat{\pi}_1)^2}}}$$

Kawasaki *et al.* (2008) applied this test statistic to the confidence interval for the difference between two independent binominal proportions. They showed that the new confidence interval showed a greater improvement in performance than the Wald interval.

**Null hypothesis variance 4:** In the test statistic used by Zhang *et al.* (2006), the estimator for the unknown parameter in variance is not unbiased. In this study, we use these unbiased estimators for the unknown parameter to propose a new test statistic that is defined as:

$$Z_{CU} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\left(\frac{\tilde{\sigma}_{10}^2}{n_1} + \frac{\tilde{\sigma}_{01}^2}{n_2}\right)\frac{1 - \Delta_0^2}{\tilde{\sigma}_{00}^2}}}$$

where, the unbiased estimators are:

$$\tilde{\sigma}_{00}^2 = \frac{(n_1n_2 - n_1 - n_2 + 2)(1-(\hat{\pi}_1 - \hat{\pi}_2)^2) + (n_2 - 2)\hat{\pi}_1(1-\hat{\pi}_1) + (n_1 - 2)\hat{\pi}_2(1-\hat{\pi}_2)}{4(n_1 - 1)(n_2 - 1)} \quad (4)$$

$$\tilde{\sigma}_{10}^2 = \frac{(1-(\hat{\pi}_1 - \hat{\pi}_2)^2) + (n_1n_2 - n_1 - 1)\hat{\pi}_1(1-\hat{\pi}_1) - \hat{\pi}_2(1-\hat{\pi}_2)}{4(n_1 - 1)(n_2 - 1)} \quad (5)$$

$$\tilde{\sigma}_{01}^2 = \frac{(1-(\hat{\pi}_1 - \hat{\pi}_2)^2) + (n_1n_2 - n_2 - 1)\hat{\pi}_2(1-\hat{\pi}_2) - \hat{\pi}_1(1-\hat{\pi}_1)}{4(n_1 - 1)(n_2 - 1)} \quad (6)$$

The derivation for these unbiased estimators is illustrated in the Discussion.

## RESULT

We show the validity and usability of each test statistic. In this research, with regard to the validity of the test, it is assumed that the type I error is close to the nominal level. Further, usability of the test is assumed to be high power.

In Table 1, we evaluate whether the actual type I error is at the nominal level of 2.5%. In Table 2, we show that the actual type I error is at the nominal level of 5%. The actual type I errors of each method are calculated by conducting a simulation 100,000 times under each condition. The following points are indicated in Table 1 and 2.

Table 1: Actual type I error (%) of test for the no inferiority hypothesis (1), nominal level is 2.5%

| Sample size | $\pi_1$ | $\Delta_0$ | Method (%) $Z_P$ | $Z_W$ | $Z_D$ | $Z_F$ | $Z_{CE}$ | $Z_{CU}$ |
|---|---|---|---|---|---|---|---|---|
| $n_1 = 10$ | 0.3 | 0.05 | 3.04 | 3.43 | 3.08 | 3.06 | 3.06 | 2.73 |
| $n_2 = 10$ | 0.3 | 0.10 | 2.85 | 4.22 | 2.83 | 2.60 | 2.88 | 2.85 |
| | 0.5 | 0.05 | 3.65 | 3.52 | 3.66 | 3.61 | 3.76 | 2.71 |
| | 0.5 | 0.10 | 2.38 | 3.49 | 2.46 | 2.42 | 2.50 | 2.37 |
| | 0.7 | 0.05 | 2.85 | 3.64 | 2.86 | 2.89 | 2.81 | 2.78 |
| | 0.7 | 0.10 | 4.04 | 4.79 | 4.15 | 2.53 | 4.13 | 3.72 |
| $n_1 = 30$ | 0.3 | 0.05 | 2.71 | 2.98 | 2.65 | 2.64 | 2.64 | 2.67 |
| $n_2 = 30$ | 0.3 | 0.10 | 2.71 | 2.55 | 2.59 | 2.65 | 2.61 | 2.68 |
| | 0.5 | 0.05 | 2.39 | 3.50 | 2.33 | 2.24 | 2.26 | 2.25 |
| | 0.5 | 0.10 | 2.66 | 2.54 | 2.61 | 2.52 | 2.56 | 2.57 |
| | 0.7 | 0.05 | 2.53 | 3.02 | 2.53 | 2.57 | 2.58 | 2.52 |
| | 0.7 | 0.10 | 2.96 | 3.00 | 2.86 | 2.53 | 2.99 | 2.91 |
| $n_1 = 50$ | 0.3 | 0.05 | 2.69 | 2.78 | 2.69 | 2.69 | 2.62 | 2.69 |
| $n_2 = 50$ | 0.3 | 0.10 | 2.58 | 2.58 | 2.53 | 2.60 | 2.66 | 2.54 |
| | 0.5 | 0.05 | 2.32 | 2.37 | 2.27 | 2.27 | 2.31 | 2.31 |
| | 0.5 | 0.10 | 2.68 | 2.75 | 2.78 | 2.85 | 2.76 | 2.79 |
| | 0.7 | 0.05 | 2.52 | 2.61 | 2.55 | 2.66 | 2.65 | 2.54 |
| | 0.7 | 0.10 | 2.89 | 2.87 | 2.85 | 2.54 | 2.89 | 2.89 |

Table 2: Actual type I error (%) of test for the noninferiority hypothesis (1), nominal level is 5.0%

| Sample size | $\pi_1$ | $\Delta_0$ | Method (%) $Z_P$ | $Z_W$ | $Z_D$ | $Z_F$ | $Z_{CE}$ | $Z_{CU}$ |
|---|---|---|---|---|---|---|---|---|
| $n_1 = 10$ | 0.3 | 0.05 | 6.61 | 7.76 | 6.63 | 4.80 | 6.63 | 4.86 |
| $n_2 = 10$ | 0.3 | 0.10 | 6.26 | 6.12 | 6.23 | 5.10 | 6.28 | 5.13 |
| | 0.5 | 0.05 | 5.45 | 8.89 | 5.46 | 4.00 | 5.52 | 3.92 |
| | 0.5 | 0.10 | 6.03 | 5.96 | 5.92 | 5.87 | 6.03 | 5.71 |
| | 0.7 | 0.05 | 6.80 | 7.37 | 6.82 | 5.38 | 6.93 | 5.31 |
| | 0.7 | 0.10 | 7.30 | 7.38 | 7.45 | 4.16 | 7.51 | 5.37 |
| $n_1 = 30$ | 0.3 | 0.05 | 5.14 | 5.27 | 5.16 | 5.10 | 5.12 | 5.25 |
| $n_2 = 30$ | 0.3 | 0.10 | 5.00 | 4.95 | 5.07 | 5.04 | 5.08 | 5.05 |
| | 0.5 | 0.05 | 5.94 | 5.94 | 6.05 | 5.94 | 6.16 | 6.10 |
| | 0.5 | 0.10 | 4.59 | 4.64 | 4.57 | 4.68 | 4.55 | 4.64 |
| | 0.7 | 0.05 | 5.18 | 5.32 | 5.19 | 5.12 | 5.23 | 5.21 |
| | 0.7 | 0.10 | 5.31 | 5.29 | 5.29 | 5.19 | 5.35 | 5.40 |
| $n_1 = 50$ | 0.3 | 0.05 | 5.05 | 4.98 | 4.97 | 5.14 | 5.06 | 5.00 |
| $n_2 = 50$ | 0.3 | 0.10 | 5.51 | 5.46 | 5.47 | 5.33 | 5.54 | 5.19 |
| | 0.5 | 0.05 | 5.49 | 5.48 | 5.47 | 5.47 | 5.41 | 5.43 |
| | 0.5 | 0.10 | 4.61 | 4.70 | 4.71 | 4.67 | 4.74 | 4.48 |
| | 0.7 | 0.05 | 5.27 | 5.22 | 5.29 | 5.37 | 5.31 | 5.31 |
| | 0.7 | 0.10 | 5.08 | 5.50 | 5.54 | 4.95 | 5.52 | 5.08 |

Table 3: Actual power (%) of test for the no inferiority hypothesis (1), nominal level is 2.5%

| Sample size | $\pi_1$ | $\pi_2$ | $\Delta_0$ | Method (%) $Z_P$ | $Z_W$ | $Z_D$ | $Z_F$ | $Z_{CE}$ | $Z_{CU}$ |
|---|---|---|---|---|---|---|---|---|---|
| $n_1 = 30$ | 0.3 | 0.20 | 0.05 | 26.50 | 29.16 | 26.60 | 26.73 | 26.58 | 26.82 |
| $n_2 = 30$ | 0.5 | 0.40 | 0.05 | 20.45 | 25.67 | 20.67 | 20.66 | 20.29 | 20.48 |
| | 0.7 | 0.60 | 0.05 | 23.71 | 25.67 | 23.92 | 24.05 | 24.16 | 23.86 |
| | 0.8 | 0.70 | 0.05 | 26.52 | 29.04 | 26.72 | 26.52 | 26.71 | 26.75 |
| | 0.3 | 0.30 | 0.10 | 14.25 | 14.41 | 14.29 | 13.90 | 14.11 | 14.07 |
| | 0.5 | 0.50 | 0.10 | 12.25 | 12.27 | 12.18 | 12.26 | 12.41 | 12.27 |
| | 0.7 | 0.70 | 0.10 | 14.20 | 14.15 | 14.31 | 13.73 | 14.27 | 14.27 |
| | 0.8 | 0.80 | 0.10 | 17.32 | 17.41 | 17.45 | 14.96 | 17.28 | 17.32 |
| | 0.3 | 0.35 | 0.20 | 23.43 | 23.64 | 25.83 | 23.43 | 25.77 | 25.74 |
| | 0.5 | 0.55 | 0.20 | 22.12 | 21.98 | 22.18 | 22.09 | 22.01 | 21.91 |
| | 0.7 | 0.75 | 0.20 | 26.11 | 26.12 | 28.72 | 25.77 | 28.54 | 27.55 |
| $n_1 = 50$ | 0.3 | 0.20 | 0.05 | 41.54 | 41.49 | 41.55 | 41.71 | 41.66 | 41.74 |
| $n_2 = 50$ | 0.5 | 0.40 | 0.05 | 31.30 | 31.91 | 31.23 | 31.23 | 31.45 | 31.21 |
| | 0.7 | 0.60 | 0.05 | 35.55 | 36.73 | 35.83 | 35.92 | 35.71 | 35.91 |
| | 0.8 | 0.70 | 0.05 | 41.60 | 41.73 | 41.61 | 41.48 | 41.36 | 41.45 |
| | 0.3 | 0.30 | 0.10 | 20.10 | 20.11 | 19.70 | 19.90 | 20.05 | 20.19 |
| | 0.5 | 0.50 | 0.10 | 18.47 | 18.36 | 18.65 | 18.33 | 18.55 | 18.33 |
| | 0.7 | 0.70 | 0.10 | 20.15 | 19.97 | 20.20 | 19.78 | 19.94 | 20.20 |
| | 0.8 | 0.80 | 0.10 | 25.42 | 25.40 | 25.20 | 23.27 | 25.58 | 25.28 |
| | 0.3 | 0.35 | 0.20 | 35.89 | 36.00 | 37.45 | 37.43 | 37.52 | 37.35 |
| | 0.5 | 0.55 | 0.20 | 34.01 | 34.53 | 34.64 | 34.45 | 34.37 | 34.10 |
| | 0.7 | 0.75 | 0.20 | 39.36 | 40.03 | 41.59 | 40.94 | 41.33 | 40.74 |
| $n_1 = 100$ | 0.3 | 0.20 | 0.05 | 69.30 | 69.10 | 69.43 | 69.34 | 69.33 | 69.37 |
| $n_2 = 100$ | 0.5 | 0.40 | 0.05 | 58.47 | 58.47 | 58.33 | 58.49 | 58.39 | 58.53 |
| | 0.7 | 0.60 | 0.05 | 60.25 | 61.07 | 60.52 | 60.21 | 60.46 | 60.30 |
| | 0.8 | 0.70 | 0.05 | 69.20 | 69.32 | 69.28 | 69.36 | 69.46 | 69.46 |
| | 0.3 | 0.30 | 0.10 | 34.16 | 34.55 | 34.71 | 34.24 | 34.56 | 34.58 |
| | 0.5 | 0.50 | 0.10 | 31.35 | 30.80 | 31.22 | 31.27 | 31.03 | 30.85 |
| | 0.7 | 0.70 | 0.10 | 34.37 | 34.35 | 34.68 | 34.50 | 34.59 | 34.42 |
| | 0.8 | 0.80 | 0.10 | 43.31 | 43.48 | 43.17 | 41.76 | 43.31 | 43.32 |
| | 0.3 | 0.35 | 0.20 | 62.32 | 62.13 | 64.01 | 63.37 | 64.12 | 63.78 |
| | 0.5 | 0.55 | 0.20 | 58.14 | 58.36 | 58.34 | 58.44 | 58.51 | 58.37 |
| | 0.7 | 0.75 | 0.20 | 66.05 | 66.71 | 68.43 | 66.94 | 67.69 | 67.57 |

The actual type I error of $Z_W$ exceeded the nominal level with a small sample size and even when the sample size was moderate, it often exceeded the nominal level. The actual type I errors of $Z_{CE}$, $Z_D$ and $Z_P$ showed similar behaviors. Besides, the actual type I errors of these methods are close to the nominal level, except in cases where the small sample sizes are small. We found that the actual type I errors of $Z_F$ and $Z_{CU}$ came close to the nominal level even though the sample size was small. Further, when the population proportion was an extreme value, the actual type I error of only $Z_F$ was close to the nominal level. Therefore, we recommend the use of $Z_F$ test statistics in cases where the population proportion is assumed to be extreme. Thus, all of the above indicate that $Z_F$ and $Z_{CU}$ test statistics have high validity. In Table 3, we showed the actual power in the one-sided test at the nominal level of 2.5%. The actual power of each method is calculated by a simulation conducted 100,000 times under each condition, as we did for the type I error.

We deduced the following points from Table 3. It indicated that the $Z_W$ statistic and the $Z_D$ statistic have higher powers than the others, especially with a small sample size. However, we do not infer that it has usability since the validity of the $Z_W$ statistic is not assured. It shows that $Z_D$, $Z_{CE}$ and $Z_{CU}$ have higher powers when the sample size is moderate. In particular, in cases with moderate sample size, it was found that $Z_{CU}$ has a stable high power.

Further, we found that $Z_F$ and $Z_P$ had lower power; in particular, $Z_F$ had lower power even at large sample size. We also found that the characters of the power of each statistic were changed by the value of the noninferiority margin only in a few cases. From the above result, it was inferred that $Z_D$, $Z_{CE}$ and $Z_{CU}$ test statistics have high usability.

## DISCUSSION

The derivation for these unbiased estimators is illustrated in this section. Let us consider a nonparametric two-sample situation, where it is assumed that the variables $Y_{11}$, $Y_{12,\ldots,}$ $Y_{1n1}$~$Y_1$ and $Y_{21}$, $Y_{22},\ldots,Y_{2n2}$~$Y_2$ are mutually independent. For the

purpose of formulating a nonparametric test, a pivotal probability is advocated by some authors. The nonparametric test for noninferiority may be formulated as:

$$p_1 = P(Y_1 < Y_2) + \frac{1}{2}P(Y_1 = Y_2)$$

The nonparametric test for noninferiority may be formulated as:

$$\begin{aligned} &H_0 : p_1 = p_{10} = 1/2 - \delta_0 \\ &H_1 : p_1 < p_{10} = 1/2 - \delta_0 \end{aligned} \quad (7)$$

where, $\delta_0$ is the noninferiority margin and $\delta_0<0$. Let $\varphi$ be a function of two real variables:

$$\varphi(x,y) = \begin{cases} 1 & x < y \\ 1/2 & x = y \\ 0 & x > y \end{cases}$$

and let:

$$U = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U_{ij}$$

where, $U_{ij} = \varphi(Y_{1i}, Y_{2j})$. The unbiased estimator of $P_1$ becomes:

$$\hat{p}_1 = \frac{1}{n_1 n_2} U$$

Let $\sigma_{11}^2$ be the variance of $U_{ij}$ nd let $\sigma_{10}^2$ and $\sigma_{01}^2$ denote the covariance:

$$\begin{aligned} &\sigma_{11}^2 = V(U_{ij}), \\ &\sigma_{10}^2 = Cov(U_{ij}, U_{il}), j \neq l \\ &\sigma_{01}^2 = Cov(U_{ij}, U_{kj}), i \neq k. \end{aligned}$$

In addition、$\sigma_{10}^2$ and $\sigma_{01}^2$ are represented as:

$$\begin{aligned} &\sigma_{10}^2 = E(U_{ij}U_{il}) - E(U_{ij})E(U_{il}) = p_2 - p_1^2, j \neq l, \\ &\sigma_{01}^2 = E(U_{ij}U_{kj}) - E(U_{ij})E(U_{kj}) = p_3 - p_1^2, i \neq k. \end{aligned}$$

The variance of $\hat{p}_1$ is given by:

$$V(\hat{p}_1) = \frac{1}{n_1 n_2}[\sigma_{11}^2 + (n-1)\sigma_{10}^2 + (m-1)\sigma_{01}^2]$$

The test statistic:

$$T = \sqrt{N}(\hat{p}_1 - p_{10})$$

is asymptotically normally distributed with expectation 0 and variance:

$$\sigma_N^2 = N\left(\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{01}^2}{n_2}\right)$$

where, $N = n_1+n_2$ let $Z = T/\sqrt{\sigma_N^2}$ ; then:

$$Z = \frac{\hat{p}_1 - p_{10}}{\sqrt{\frac{\sigma_N^2}{N}}} = \frac{\hat{p}_1 - p_{10}}{\sqrt{\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{01}^2}{n_2}}}$$

is an asymptotically normal distribution. However, we cannot use this test statistic. We should replace the unknown parameter in the Z-test statistics by estimators.

Munzel and Hsuschke (2003) proposed that the test statistics for hypothesis (7) is:

$$Z_M = \frac{-(\hat{p}_1 - p_{10})}{\sqrt{\frac{\hat{\sigma}_{10}^2}{n_1} + \frac{\hat{\sigma}_{01}^2}{n_2}}}$$

where, the estimator is:

$$\hat{\sigma}_{10}^2 = \frac{1}{n_1 n_2^2}\sum_{i=1}^{n_1}\left[U_{i\cdot} - \bar{U}_{1\cdot}\right]^2 \quad (8)$$

where, $\bar{U}_{1\cdot} = \sum_{i=1}^{n_1} U_{i\cdot}/n_1$ . Similarly, let us denote $\bar{U}_{2\cdot} = \sum_{j=1}^{n_2} U_{j}/n_2$ , $U_{\cdot j} = \sum_{i=1}^{n_1} U_{ij}$ and:

$$\hat{\sigma}_{01}^2 = \frac{1}{n_1^2 n_2}\sum_{j=1}^{n_2}\left[U_{\cdot j} - \bar{U}_{2\cdot}\right]^2 \quad (9)$$

Moreover the empirical estimators of $P^2$ and $P^3$ are:

$$\hat{p}_2 = \frac{1}{n_1 n_2^2}\sum_{i=1}^{n_1} U_{i\cdot}^2, \hat{p}_3 = \frac{1}{n_1^2 n_2}\sum_{j=1}^{n_2} U_{\cdot j}^2$$

Zhang *et al.* (2006) pointed out that one problem with this is that it used the variance under an alternative hypothesis. They proposed the test statistic:

$$Z_C = \frac{-\left(\hat{p}_1 - p_{10}\right)}{\sqrt{\left(\frac{\sigma_{10}^2}{m} + \frac{\sigma_{01}^2}{n}\right)\frac{p_{10}(1-p_{10})}{\sigma_{00}^2}}}$$

in which $\sigma_{00}^2 = p_1(1-p_1)$ and used a variance under a null hypothesis. This test statistic $Z_C$ follows the asymptotic standard normal distribution. However, we cannot use it as it is. They used expressions (8) and (9) and proposed the $Z_{CE}$ test statistic:

$$Z_{CE} = \frac{-(\hat{p}_1 - p_{10})}{\sqrt{\left(\frac{\hat{\sigma}_{10}^2}{m} + \frac{\hat{\sigma}_{01}^2}{n}\right)\frac{p_{10}(1-p_{10})}{\hat{\sigma}_{00}^2}}}$$

where, $\hat{\sigma}_{00}^2 = \hat{p}_1(1-\hat{p}_1)$. Zhang *et al.* (2006) call the $Z_{CE}$ test statistic an empirical test statistic. We derive unbiased estimators for the unknown parameter with $Z_C$ test statistics. The unbiased estimators of $P_2$ and $P_3$ are given by:

$$\tilde{p}_2 = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{i}^{n_1} \sum_{j}^{n_2} \sum_{j \neq l}^{n_2} U_{ij} U_{il}$$

$$\tilde{p}_3 = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{i}^{n_1} \sum_{j}^{n_2} \sum_{i \neq k}^{n_1} U_{ij} U_{kj}.$$

We show that the unbiased estimator of $\sigma_{00}^2$ is:

$$\tilde{\sigma}_{00}^2 = \frac{n_1 n_2 (\hat{p}_1 - \hat{p}_1^2) - (n_2 - 1)(\hat{p}_1 - \tilde{p}_2) - (n_1 - 1)(\hat{p}_1 - \tilde{p}_3)}{(n_1 - 1)(n_2 - 1)} \tag{10}$$

Moreover, the unbiased estimators of $\sigma_{10}^2$ and $\sigma_{01}^2$ can be:

$$\tilde{\sigma}_{10}^2 = \frac{n_1 n_2 (\hat{p}_1 - \hat{p}_1^2) - n_1 (n_2 - 1)(\hat{p}_1 - \tilde{p}_2) - (n_1 - 1)(\hat{p}_1 - \tilde{p}_3)}{(n_1 - 1)(n_2 - 1)} \tag{11}$$

$$\tilde{\sigma}_{01}^2 = \frac{n_1 n_2 (\hat{p}_1 - \hat{p}_1^2) - (n_2 - 1)(\hat{p}_1 - \tilde{p}_2) - n_2 (n_1 - 1)(\hat{p}_1 - \tilde{p}_3)}{(n_1 - 1)(n_2 - 1)} \tag{12}$$

Because $\hat{p}_1$, $\tilde{p}_2$ and $\tilde{p}_3$ are unbiased and consistent, $\tilde{\sigma}_{10}^2$ and $\tilde{\sigma}_{01}^2$ are unbiased and consistent. Therefore, the $Z_{CU}$ test statistic is proposed as:

$$Z_{CU} = \frac{-\left(\hat{p}_1 - p_{10}\right)}{\sqrt{\left(\frac{\tilde{\sigma}_{10}^2}{m} + \frac{\tilde{\sigma}_{01}^2}{n}\right)\frac{p_{10}(1-p_{10})}{\tilde{\sigma}_{00}^2}}}$$

We let $Y_1$ and $Y_2$ be two independent Bernoulli random variables with $\pi_1$ and $\pi_2$ respectively. Through simple calculation, we obtain:

$$p_1 = \frac{1}{2} - \frac{1}{2}(\pi_1 - \pi_2)$$

Therefore the estimator of $P_1$ is given by:

$$\hat{p}_1 = \frac{1}{2} - \frac{1}{2}(\hat{\pi}_1 - \hat{\pi}_2)$$

The hypothesis for noninferiority, expression (7), can be represented as:

$$H_0 : \pi_1 - \pi_2 = -\Delta_0$$
$$H_1 : \pi_1 - \pi_2 > -\Delta_0,$$

where, $\Delta_0 = -2\delta_0$. The important components of the asymptotical variance $\hat{p}_1$ are:

$$\sigma_{10}^2 = \frac{1}{4}\pi_1(1-\pi_1), \sigma_{01}^2 = \frac{1}{4}\pi_2(1-\pi_2), \sigma_{00}^2 = \frac{1}{4}[1-(\pi_1-\pi_2)^2]$$

Therefore, the $Z_{CE}$ test statistic is:

$$Z_{CE} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta_0}{\sqrt{\left(\frac{\hat{\sigma}_{10}^2}{m} + \frac{\hat{\sigma}_{01}^2}{n}\right)\frac{p_{10}(1-p_{10})}{\hat{\sigma}_{00}^2}}}$$

Where:

$$\hat{\sigma}_{10}^2 = \frac{1}{4}\pi_1(1-\pi_1)$$
$$\hat{\sigma}_{01}^2 = \frac{1}{4}\pi_2(1-\pi_2)$$
$$\hat{\sigma}_{00}^2 = \frac{1}{4}[1-(\pi_1-\pi_2)^2]$$

We can obtain other expressions of the relationship between the empirical estimator and unbiased estimator for $p_2$ and $P_3$ as:

$$\tilde{p}_2 = \hat{p}_2 - \frac{1}{n_2 - 1}(\hat{\sigma}_{11}^2 - \hat{\sigma}_{01}^2) \quad (13)$$

$$\tilde{p}_3 = \hat{p}_3 - \frac{1}{n_1 - 1}(\hat{\sigma}_{11}^2 - \hat{\sigma}_{10}^2) \quad (14)$$

Substituting (13) and (14) into (10-12) and noticing that $\hat{\sigma}_{11}^2 = \hat{\sigma}_{10}^2 + \hat{\sigma}_{01}^2$ for the Bernoulli variable, the unbiased estimator of $\sigma_{00}^2$, $\sigma_{10}^2$ and $\sigma_{01}^2$ can be derived as expressions (4-6) respectively.

## CONCLUSION

In this study, we investigated the validity and usability of test statistics in the noninferiority test for the difference between two independent binominal proportions.

It was deduced that the power of the $Z_P$ test statistic is generally low. We suppose that this is a result of the use of the variance with an assumed null hypothesis for a significance test.

We found that the $Z_W$ test statistic showed higher power than the $Z_P$ test statistic. However, it also showed that the actual level frequently exceeded the nominal level. Therefore, the $Z_W$ test statistic does not fulfill the validity of testing. Hence, using this method only because its power is high might lead to a wrong conclusion.

The power of the $Z_D$ test statistic performed better. However, it is best if this test statistic is used judiciously since the estimator of a nuisance parameter used in this test statistic may exceed the limit value.

We have deduced that the $Z_F$ test statistic is the method that passes the validity in the noninferiority test. Especially, we also found that this is also the only method in which the type I error comes close to the nominal level when the population proportion is an extreme value. However, we also found that the power of this method is comparatively low. Moreover, the method of calculating this test statistic is a little complicated since this method uses a restricted maximum likelihood estimator.

In conclusion, we prove that the proposed $Z_{CE}$ and $Z_{CU}$ test statistics are methods that show that their type I errors are comparatively closer to the nominal level and also that they have reasonably higher powers; This is particularly true in the case of the $Z_{CU}$ test statistic, which uses an unbiased estimator that shows a stable positive behavior in the hypothesis test. In addition, one of the biggest advantages of our method is that it does not require complicated calculations.

## REFERENCES

Almendra-Arao, F., 2009. Behavior of the asymptotic pooled Z-statistic. JP J. Biostat., 3: 247-256. http://pphmj.com/abstract/4362.htm

Dann, R.S. and G.G. Koch, 2008. Methods for one-sided testing of the difference between proportions and sample size considerations related to noninferiority clinical trials. Pharm. Stat., 7: 130-141. DOI: 10.1002/pst.287

Dunnett, C.W. and M. Gent, 1977. Significance testing to establish equivalence between treatments with special reference to data in the form of 2×2 tables. Biometrics, 33: 593-602. http://www.jstor.org/pss/2529457

Farrington, C.P. and G. Manning, 1990. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of nonzero risk difference or nonentity relative risk. Stat. Med., 9: 1447-1454. DOI: 10.1002/sim.4780091208

Hirotsu, C., W. Hashimoto, K. Nishihar and E. Adachi, 1997. Calculation of the confidence interval with skewness correction for the difference of two binominal probabilities. JP. J. Applied Stat., 26: 83-97. DOI: 10.5023/jappstat.26.83

Kawasaki, Y., S. Midorikawa and E. Miyaoka, 2008. Comparisons of confidence intervals for the difference between two independent binomial proportions. Proceedings of the International Association for Statistical Computing, pp: 829-838.

Miettinen, O. and M. Nurminen, 1985. Comparative analysis of two rates. Stat. Med., 4: 213-226. DOI: 10.1002/sim.4780040211

Munzel, U. and D. Hsuschke, 2003. A nonparametric test for proving no inferiority in clinical trials with ordered categorical data. Pharma. Stat., 2: 31-37. DOI: 10.1002/pst.17

Zhang, F., H. Fuping and E. Miyaoka, 2006. No inferior nonparametric test at order categorical data. Proceedings of the Organized Sessions Japanese Joint Statistical Meeting, pp: 268. https://www.conferenceissjp.com/upload/upload_date.php?id=00061