Journal of Biopharmaceutical Statistics

ISSN: 1054-3406 (Print) 1520-5711 (Online) Journal homepage: http://www.tandfonline.com/loi/lbps20

# Confidence interval construction for the difference between two correlated proportions with missing observations

Nian-Sheng Tang, Hui-Qiong Li, Man-Lai Tang & Jie Li

# Confidence interval construction for the difference between two correlated proportions with missing observations

Nian-Sheng Tang[a]*, Hui-Qiong Li[a], Man-Lai Tang[b] and Jie Li[a]

[a]*Department of Statistics, Yunnan University, Kunming 650091, P. R. of China*

[b]*Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong*

## Abstract

Under the assumption of missing at random, eight confidence intervals (CIs) for the difference between two correlated proportions in the presence of incomplete paired binary data are constructed on the basis of the likelihood ratio statistic, the score statistic, the Wald-type statistic, the hybrid method incorporated with the Wilson score and Agresti-Coull (AC) intervals, and the Bootstrap-resampling method. Extensive simulation studies are conducted to evaluate the performance of the presented CIs in terms of coverage probability and expected interval width. Our empirical results evidence that the Wilson-Score-based hybrid CI and the Wald-type CI together with the constrained maximum likelihood estimates perform well for small to moderate sample sizes in the sense that (i) their empirical coverage probabilities are quite close to the pre-specified confidence level, (ii) their expected interval widths are shorter and (iii) their ratios of the mesial non-coverage to non-coverage probabilities lie in interval [0.4, 0.6]. An example from a neurological study is used to illustrate the proposed methodologies.

*Key words:* Bootstrap confidence interval; Correlated proportion difference; Squaring-and-adding confidence interval; Missing data; Paired binary data.

*Corresponding author: Dr. Nian-Sheng Tang, Department of Statistics, Yunnan University, Kunming 650091, P. R. of China. E-mail: nstang@ynu.edu.cn





## 1 Introduction

Incomplete matched-pair data are often encountered in paired-comparison studies of two treatments or two different conditions of the same treatment. For example, in a neurological study of meningitis patients (Choi and Stablein, 1982), 33 young meningitis patients at the St. Louis Children's Hospital were given neurological tests at the time of admission and at the end of a standard treatment on neurological complication. In that study, 25 patients received neurological tests at the beginning and at the end of the standard treatment, 6 patients received neurological tests only at the beginning but not at the end of the standard treatment, and 2 patients received neurological tests only at the end but not at the beginning of the standard treatment. Thus, the resultant data included two parts: the complete and paired observations, and the unpaired observations. The data are presented in Table 1 in which 1 and 0 represent the absence and presence of neurological complication, respectively.

In the aforementioned neurological clinical trial, one would like to test the equality of the incidence rates of neurological complication before and after the standard treatment. To this end, one could construct a $100(1-\alpha)\%$ confidence interval (CI) for the difference between two correlated proportions in the presence of incomplete paired binary data. If the resultant CI entirely lies in the interval $(-\delta_0, \delta_0)$ with $\delta_0$ (> 0) being some pre-specified clinically acceptable threshold, one cannot reject the equality of two proportions at the significance level α. Hence, motivated by the aforementioned neurological data, we consider CI construction for the difference between two correlated proportions in the presence of incomplete paired binary data.

The problem of testing the equality and CI construction of the difference between two correlated proportions in the presence of incomplete paired binary data has received considerable attention in past years. For example, ones can consult Choi and Stablein (1982), Ekbohm (1982), Campbell (1984), Bhoj and Snijders (1986), Thomson (1995) for the large sample method, and Pradhan, Menon and Das (2013) for the corrected profile likelihood method. When sample size is small, Tang and Tang (2004) proposed the exact unconditional test procedure for testing equality of two correlated proportions with incomplete correlated data. Tang, Ling and Tian (2009) developed the exact unconditional and approximate unconditional CIs for proportion difference in the presence of incomplete paired binary data. Lin *et al.* (2009) presented a Bayesian method to test equality of two correlated proportions with incomplete correlated data. However, all the aforementioned methods were developed under the assumption of missing completely at random (MCAR), i.e., the probability of missing is independent of treatment and outcome (Choi and Stablein, 1982).

Statistical inference on incomplete paired binary data under the assumption of missing at random (MAR) has received limited attention (see, Choi and Stablein, 1988; Little and Robin, 2002; Chang, 2009). For example, Choi and Stablein (1988) discussed the problem of testing the equality of two correlated proportions under the assumption of MAR, and pointed out that those tests that utilize all the data are generally more efficient than those discarding part of the data. Chang (2009) proposed an EM algorithm to evaluate the maximum likelihood estimates (MLEs) of two correlated proportions under the assumption of MAR and concluded that their proposed estimators are more efficient than those conventional estimators in terms of asymptotic relative efficiency. However, to our knowledge, little work has been done on CI construction for the difference between two correlated proportions under the assumption of MAR.

Inspired by Shao and Tu (1995), Newcombe (1998b) and Zou and Donner (2008), we develop eight CIs for proportion difference in the presence of incomplete paired binary data under the assumption of MAR (Chang, 2009; Choi and Stablein, 1988) based on the likelihood ratio test, score test, Wald-type test, hybrid method and Bootstrap-resampling method. The derived hybrid CI possesses a closed-form expression, which largely reduces the computational burden, and the presented Bootstrap-resampling CIs have not been considered in the literature related to missing observations. These CIs can be used for analysis of incomplete paired binary data as well as of complete paired binary data.

The rest of this paper is organized as follows. Section 2 first reviews the missing mechanism given in Chang (2009). Five different methods are then presented to construct CIs for correlated proportion difference under the MAR assumption in Section 2. Simulation studies are conducted to evaluate the performance of the proposed CIs in terms of coverage probability, expected interval width, and mesial and distal non-coverage probabilities in Section 3. An example from the aforementioned neurological study is used to illustrate the proposed methodologies in Section 4. A brief discussion is given in Section 5.

## 2 Model and confidence interval estimators

### 2.1 Model

Consider a crossover design in which two treatments (e.g., treatment A and treatment B) are sequentially performed on the same subject. We may assume that *X* and *Y* are outcomes of two treatments sequentially applied to the same subject. Denote all the possible values of *X* and *Y* by 0 and 1. We consider the situation where a portion of observations is complete and paired, and the remainder is incomplete and unpaired. Let $n_{ij}$ be the number of subjects who sequentially underwent both treatments (i.e., with both *X* and *Y* being observed) with outcome $X = i$ and $Y = j$ for $i, j = 0,1$, $n_x$ be the number of subjects who alone underwent treatment A (i.e., with only *X* being observed) and $n_y$ be the number of subjects who alone underwent treatment B (i.e., with only *Y* being observed), $n_{x0}$ be the number of subjects who alone underwent treatment A with outcome *X* = 0, $n_{y0}$ be the number of subjects who alone underwent treatment B with outcome *Y* = 0, and *z* be the number of subjects who did not undergo either treatment A or treatment B. Thus, $n_{x1} = n_x - n_{x0}$ represents the number of subjects who alone underwent treatment A with outcome *X* = 1, and $n_{y1} = n_y - n_{y0}$ represents the number of subjects who alone underwent treatment B with outcome *Y* = 1. Denote $n = n_{00} + n_{01} + n_{10} + n_{11}$, $n_{0+} = n_{00} + n_{01}$, $n_{1+} = n_{10} + n_{11}$, $n_{+0} = n_{00} + n_{10}$, $n_{+1} = n_{01} + n_{11}$. Since subjects who did not undergo either treatment A or treatment B do not provide any information for estimating response rates, these subjects will be excluded from the final data analysis (e.g., Chang, 2009). Hence, similar to Chang

(2009), we assume that $z = 0$, which indicates that $N = n + n_x + n_y$. The data can be summarized in Table 2.

Let $P_{EE}$ be the probability that a subject sequentially underwent both treatments A and B, $P_{EI}$ be the probability that a subject alone underwent treatment A, $P_{IE}$ be the probability that a subject alone underwent treatment B, and $P_{II}$ be the probability that a subject did not undergo either of treatments A and B. Similar to Chang (2009), we assume $P_{II} = 0$. Let $p_{ij}$ be the conditional probability of the subject having experimental outcome $X = i$ and $Y = j\,(i, j = 0,1)$ given a subject sequentially underwent both treatments A and B. Hence, we have $P_{EE} + P_{EI} + P_{IE} = 1$ and $p_{00} + p_{01} + p_{10} + p_{11} = 1$. Denote $p_{0+} = p_{00} + p_{01}$ and $p_{+0} = p_{00} + p_{10}$, which are the response rates for $X = 0$ and $Y = 0$, respectively. Following Choi and Stablein (1988) and Chang (2009), we assume that the missing mechanism is MAR, i.e., the probability of missing is independent of the outcome but dependent of the treatment (Little and Robin, 2002). In other words, the assumption of MAR is equivalent to that the probability of a missing observation differs for different treatments but is constant for the same treatment irrespective of the outcome (Choi and Stablein, 1988). Mathematically, $P$ (outcome (i.e., $X$ or $Y$) is missing | outcome, treatment (i.e., $A$ or $B$)) = $P$ (outcome (i.e., $X$ or $Y$) is missing | treatment (i.e., $A$ or $B$)). Hence, it follows from the multiplication rule of probability that the cell probability corresponding to $n_{ij}$ is $\pi_{ij} = P$ (sequentially underwent both treatments and outcome $X = i$ and $Y = j$) = $P_{EE}P$ (outcome $X = i$, $Y = j$ | sequentially underwent both treatments) = $P_{EE}p_{ij}$ for $i, j = 0, 1$. Similarly, the probabilities corresponding to $n_{x0}$ and $n_{y0}$ are $P_{EI}\,p_{0+}$ and $P_{IE}\,p_{+0}$, respectively. Therefore, the observed data $\boldsymbol{D}=\left\{n_{00},n_{01},n_{10},n_{11},n_{x0},n_{x1},n_{y0},n_{y1}\right\}$ can be assumed to come from the following multinomial distribution:

$$
\begin{aligned}
\mathrm{P}\left(\boldsymbol{D}\mid N,\boldsymbol{\theta},P_{EE},P_{EI}\right) &= c\cdot\left(P_{EE}\,p_{00}\right)^{n_{00}}\left(P_{EE}\,p_{01}\right)^{n_{01}}\left(P_{EE}\,p_{10}\right)^{n_{10}}\left(P_{EE}\,p_{11}\right)^{n_{11}} \\
&\quad\times\left(P_{EI}\,p_{0+}\right)^{n_{x0}}\left(P_{EI}\,p_{1+}\right)^{n_{x1}}\left(P_{IE}\,p_{+0}\right)^{n_{y0}}\left(P_{IE}\,p_{+1}\right)^{n_{y1}} \qquad (2.1)\\
&= c\cdot p_{00}^{n_{00}}\,p_{01}^{n_{01}}\,p_{10}^{n_{10}}\,p_{11}^{n_{11}}\,p_{0+}^{n_{x0}}\,p_{1+}^{n_{x1}}\,p_{+0}^{n_{y0}}\,p_{+1}^{n_{y1}}\,P_{EE}^{n}\,P_{EI}^{n_x}\,P_{IE}^{n_y},
\end{aligned}
$$

where $c=N!/\left\{n_{00}!n_{01}!n_{10}!n_{11}!n_{x0}!n_{x1}!n_{y0}!n_{y1}!\right\}$, $n_{x1}=n_x-n_{x0}$, $n_{y1}=n_y-n_{y0}$, $p_{1+}=1-p_{0+}$, $p_{+1}=1-p_{+0}$ and $\boldsymbol{\theta}=\left(p_{00},p_{01},p_{10}\right)$.

## 2.2 Confidence interval estimators

In this paper, our main purpose is to construct CI for the correlated proportion difference $\Delta=p_{0+}-p_{+0}=p_{01}-p_{10}$.

(i) *Confidence interval based on the likelihood ratio test*

Let $\hat{p}_{ij}$ be the maximum likelihood estimator (MLE) of $p_{ij}$ for $i$, $j$ = 0, 1. It follows from Chang (2009) that $\hat{p}_{00}$, $\hat{p}_{01}$, $\hat{p}_{10}$ and $\hat{p}_{11}$ can be obtained via the EM algorithm, which is presented in Appendix A. It can be shown from Equation (2.1) that the MLEs of $P_{EE}$, $P_{EI}$ and $P_{IE}$ are given by $\hat{P}_{EE}=n/N$, $\hat{P}_{EI}=n_x/N$ and $\hat{P}_{IE}=n_y/N$, respectively.

Let $\tilde{p}_{ij}$ be the constrained MLE of $p_{ij}$ under $H_0: p_{0+} - p_{+0} = \Delta$ for $i$, $j$ = 0, 1. Thus, it follows from Equation (2.1) that $\tilde{p}_{00}$ and $\tilde{p}_{10}$ satisfy the following equations:

$$\begin{cases} \frac{n_{00}}{\tilde{p}_{00}} - \frac{n_{11}}{\tilde{p}_{11}} + \frac{n_{x0}}{\tilde{p}_{0+}} - \frac{n_{x1}}{1-\tilde{p}_{0+}} + \frac{n_{y0}}{\tilde{p}_{+0}} - \frac{n_{y1}}{1-\tilde{p}_{+0}} = 0, \\ \frac{n_{01}}{\tilde{p}_{01}} + \frac{n_{10}}{\tilde{p}_{10}} - \frac{2n_{11}}{\tilde{p}_{11}} + \frac{n_{x0}}{\tilde{p}_{0+}} - \frac{n_{x1}}{1-\tilde{p}_{0+}} + \frac{n_{y0}}{\tilde{p}_{+0}} - \frac{n_{y1}}{1-\tilde{p}_{+0}} = 0, \end{cases} \quad (2.2)$$

where $\tilde{p}_{01} = \tilde{p}_{10} + \Delta$, $\tilde{p}_{11} = 1 - \tilde{p}_{00} - 2\tilde{p}_{10} - \Delta$, $\tilde{p}_{0+} = \tilde{p}_{+0} + \Delta$, and $\tilde{p}_{+0} = \tilde{p}_{00} + \tilde{p}_{10}$. The likelihood ratio statistic (Choi and Stablein, 1982) for testing $H_0: p_{0+} - p_{+0} = \Delta$ is then given by

$$T_l(\Delta) = 2\left\{ l\left(\hat{p}_{00}, \hat{p}_{01}, \hat{p}_{10}, \hat{p}_{11}\right) - l\left(\tilde{p}_{00}, \tilde{p}_{01}, \tilde{p}_{10}, \tilde{p}_{11}\right)\right\},$$

which is asymptotically distributed as the chi-squared distribution with one degree of freedom under $H_0: p_{0+} - p_{+0} = \Delta$, where

$l\left(p_{00}, p_{01}, p_{10}, p_{11}\right) = n_{00}\log p_{00} + n_{01}\log p_{01} + n_{10}\log p_{10} + n_{11}\log p_{11} + n_{x0}\log p_{0+} + n_{x1}\log\left(1-p_{0+}\right) + n_{y0}\log p_{+0} + n$

. Therefore, the approximate 100(1 − $\alpha$)% likelihood-ratio-test-based CI for Δ is given by $\left[\Delta_L, \Delta_U\right]$, where $-1 \le \Delta_L \le \Delta_U \le 1$ are the smaller and larger roots of Δ to the following equation

$$T_l\left(\Delta\right) = \chi^2_{1,\alpha},$$

where $\chi^2_{1,\alpha}$ is the upper $\alpha$-percentile of the central $\chi^2$ distribution with one degree of freedom. There are no closed-form expressions for $\Delta_L$ and $\Delta_U$. Hence, the bisection searching algorithm can be used to obtain $\Delta_L$ and $\Delta_U$.

(ii) *Confidence interval based on the score test*

After some routine computation, the score statistic for testing the null hypothesis $H_0 : p_{0+} - p_{+0} = \Delta$ can be shown to be

$$T_s(\Delta) = \left( \frac{n_{01}}{\tilde{p}_{01}} - \frac{n_{11}}{\tilde{p}_{11}} + \frac{n_{x0}}{\tilde{p}_{0+}} - \frac{n_{x1}}{\tilde{p}_{1+}} \right) \sqrt{\frac{\tilde{\mathbb{A}}_2 + \left(\tilde{a} + \tilde{b}\right)\tilde{\mathbb{A}}_1}{\tilde{\mathbb{B}}_1 + \tilde{\mathbb{A}}_1 \tilde{a}\tilde{b} + \tilde{\mathbb{B}}_2 \tilde{a} + \tilde{\mathbb{B}}_3 \tilde{b}}},$$

which is asymptotically distributed as the standard normal distribution under $H_0$, where $\tilde{N}_{ij} = n / \tilde{p}_{ij}$ for $i$, $j$ = 0, 1, $\tilde{a} = n_x / \left\{ \tilde{p}_{0+}(1 - \tilde{p}_{0+}) \right\}$, $\tilde{b} = n_y / \left\{ \tilde{p}_{+0}(1 - \tilde{p}_{+0}) \right\}$, $\tilde{p}_{01} = \tilde{p}_{10} + \Delta$, $\tilde{p}_{11} = 1 - \tilde{p}_{00} - 2\tilde{p}_{10} - \Delta$, $\tilde{\mathbb{A}}_1 = \sum_{i=1}^{2} \sum_{j=1}^{2} \tilde{N}_{ij}$, $\tilde{\mathbb{A}}_2 = (\tilde{N}_{00} + \tilde{N}_{11})(\tilde{N}_{01} + \tilde{N}_{10}) + 4\tilde{N}_{00}\tilde{N}_{11}$, $\tilde{\mathbb{B}}_1 = \tilde{N}_{00}\tilde{N}_{01}\tilde{N}_{1+} + \tilde{N}_{10}\tilde{N}_{11}\tilde{N}_{0+}$ with $\tilde{N}_{1+} = \tilde{N}_{10} + \tilde{N}_{11}$ and $\tilde{N}_{0+} = \tilde{N}_{00} + \tilde{N}_{01}$, $\tilde{\mathbb{B}}_2 = \tilde{N}_{0+}\tilde{N}_{1+}$, $\tilde{\mathbb{B}}_3 = (\tilde{N}_{00} + \tilde{N}_{10})(\tilde{N}_{01} + \tilde{N}_{11})$, and $\tilde{\mathbb{B}}_4 = \tilde{\mathbb{A}}_1$. Detailed derivation for $T_s(\Delta)$ is presented in Appendix B. The approximate 100(1 − $\alpha$)% confidence limits $\Delta_L$ and $\Delta_U$ for $\Delta$ via score test statistic can be obtained by solving the following equation:

$$T_s(\Delta) = \pm z_{\alpha/2},$$

where the plus and minus signs correspond to the lower limit $\Delta_L$ and the upper limit $\Delta_U$, respectively, and $-1 \le \Delta_L \le \Delta_U \le 1$. These two limits can be obtained by using the secant algorithm (Tango, 1998).

(iii) *Confidence interval based on the Wald-type statistic*

Let $\hat{\Delta} = \hat{p}_{0+} - \hat{p}_{+0}$ be the MLE of $\Delta$. It follows from Chang (2009) that the asymptotic expectation of $\hat{\Delta}$ is given by $E(\hat{\Delta}) \approx \Delta$, and the asymptotic variances of $\hat{p}_{0+}$ and $\hat{p}_{+0}$ can be estimated by

$$\widehat{\text{Var}}(\hat{p}_{0+}) = \frac{1}{N\hat{A}_0}\left[\hat{P}_{EE}\hat{\mathbb{D}}_1 + \hat{P}_{IE}\frac{\hat{\mathbb{D}}_3}{\hat{\mathbb{D}}_2}\right], \textit{ and } \widehat{\text{Var}}(\hat{p}_{+0}) = \frac{1}{N\hat{A}_0}\left[\hat{P}_{EE}\hat{\mathbb{D}}_2 + \hat{P}_{EI}\frac{\hat{\mathbb{D}}_3}{\hat{\mathbb{D}}_1}\right], \qquad (2.3)$$

respectively, where $\hat{\mathbb{D}}_1 = \hat{p}_{0+}(1-\hat{p}_{0+})$, $\hat{\mathbb{D}}_2 = \hat{p}_{+0}(1-\hat{p}_{+0})$, $\hat{\mathbb{D}}_3 = \hat{p}_{00}\hat{p}_{01}\hat{p}_{1+} + \hat{p}_{0+}\hat{p}_{10}\hat{p}_{11}$, and $\hat{A}_0 = \hat{P}_{EE} + \hat{P}_{EI}\hat{P}_{IE}\hat{\mathbb{D}}_3 / (\hat{\mathbb{D}}_1\hat{\mathbb{D}}_2)$. Similar to Chang (2009), it is shown that the covariance of $\hat{p}_{0+}$ and $\hat{p}_{+0}$ is given by $\text{Cov}(\hat{p}_{0+}, \hat{p}_{+0}) = -P_{EE}(p_{00}p_{11} - p_{01}p_{10})/(NA_0)$, which can be estimated by $\widehat{\text{Cov}}(\hat{p}_{0+}, \hat{p}_{+0}) = -\hat{P}_{EE}(\hat{p}_{00}\hat{p}_{11} - \hat{p}_{01}\hat{p}_{10})/(N\hat{A}_0)$. Thus, the asymptotic variance of $\hat{\Delta}$ can be expressed as

$$\widehat{\text{Var}}(\hat{\Delta}) = \widehat{\text{Var}}(\hat{p}_{0+}) + \widehat{\text{Var}}(\hat{p}_{+0}) - 2\frac{\hat{P}_{EE}}{N\hat{A}_0}(\hat{p}_{00}\hat{p}_{11} - \hat{p}_{01}\hat{p}_{10}).$$

An approximate $100(1-\alpha)\%$ CI for $\Delta$ on the basis of the Wald-type statistic $T_{w_1} = (\hat{\Delta} - \Delta)/\sqrt{\widehat{\text{Var}}(\hat{\Delta})}$, which is asymptotically distributed as the standard normal distribution, is given by

$$\left[\max\{-1, \hat{\Delta} - z_{\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{\Delta})}\}, \min\{1, \hat{\Delta} + z_{\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{\Delta})}\}\right],$$

which is denoted as $T_{w1}$-CI. It has been shown that truncating interval to lie within [−1, 1] makes the interval unsatisfactory (Newcombe, 1998a).

On the other hand, the asymptotic variances of $\hat{p}_{0+}$ and $\hat{p}_{+0}$ can be estimated by

$$\widetilde{\text{var}}\left(\hat{p}_{0+}\right)=\frac{1}{N\tilde{A}_0}\left[\hat{P}_{EE}\tilde{\mathbb{D}}_1+\hat{P}_{IE}\frac{\tilde{\mathbb{D}}_3}{\tilde{\mathbb{D}}_2}\right],\ \textit{and}\ \widetilde{\text{var}}\left(\hat{p}_{+0}\right)=\frac{1}{N\tilde{A}_0}\left[\hat{P}_{EE}\tilde{\mathbb{D}}_2+\hat{P}_{EI}\frac{\tilde{\mathbb{D}}_3}{\tilde{\mathbb{D}}_1}\right], \quad (2.4)$$

respectively, where $\tilde{\mathbb{D}}_1=\tilde{p}_{0+}\left(1-\tilde{p}_{0+}\right)$, $\tilde{\mathbb{D}}_2=\tilde{p}_{+0}\left(1-\tilde{p}_{+0}\right)$, $\tilde{\mathbb{D}}_3=\tilde{p}_{00}\tilde{p}_{01}\tilde{p}_{1+}+\tilde{p}_{0+}\tilde{p}_{10}\tilde{p}_{11}$, and $\tilde{A}_0=\hat{P}_{EE}+\hat{P}_{EI}\hat{P}_{IE}\tilde{\mathbb{D}}_3/\left(\tilde{\mathbb{D}}_1\tilde{\mathbb{D}}_2\right)$. Thus, the asymptotic variance of $\hat{\Delta}$ can be estimated by

$$\widetilde{\text{var}}\left(\hat{\Delta}\right)=\widetilde{\text{var}}\left(\hat{p}_{0+}\right)+\widetilde{\text{var}}\left(\hat{p}_{+0}\right)-2\frac{\hat{P}_{EE}}{N\tilde{A}_0}(\tilde{p}_{00}\tilde{p}_{11}-\tilde{p}_{01}\tilde{p}_{10}).$$

An approximate 100(1 − $\alpha$)% CI for Δ on the basis of the Wald-type statistic $T_{w_2}(\Delta)=(\hat{\Delta}-\Delta)/\sqrt{\widetilde{\text{var}}(\hat{\Delta})}$, which is asymptotically distributed as the standard normal distrib tion, can be obtained by solving the following equation:

$$T_{w_2}(\Delta)=\pm z_{\alpha/2},$$

where the plus and minus signs correspond to the lower limit $\Delta_L$ and the upper limit $\Delta_U$, respectively, and $-1\le\Delta_L\le\Delta_U\le 1$, which is denoted as $T_{w2}$-CI.

(iv) *Confidence interval based on the hybrid method*

Let $l_{0+}$ and $l_{+0}$ be the lower limits of the approximate 100(1−*α)%* two-sided CIs for $p_{0+}$ and $p_{+0}$, respectively. By the Central Limit Theorem, we have $l_{0+} = \hat{p}_{0+} - z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{p}_{0+})}$ and $l_{+0} = \hat{p}_{+0} - z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{p}_{+0})}$ Thus, we can obtain

$$\widehat{\mathrm{Var}}_l(\hat{p}_{0+}) = (\hat{p}_{0+} - l_{0+})^2 / z_{\alpha/2}^2, \widehat{\mathrm{Var}}_l(\hat{p}_{+0}) = (\hat{p}_{+0} - l_{+0})^2 / z_{\alpha/2}^2. \qquad (2.5)$$

Similarly, for the upper limits $u_{0+}$ and $u_{+0}$ of $p_{0+}$ and $p_{+0}$, we have

$$\widehat{\mathrm{Var}}_u(\hat{p}_{0+}) = (u_{0+} - \hat{p}_{0+})^2 / z_{\alpha/2}^2, \widehat{\mathrm{Var}}_u(\hat{p}_{+0}) = (u_{+0} - \hat{p}_{+0})^2 / z_{\alpha/2}^2. \qquad (2.6)$$

From Equations (2.5) and (2.6), we observe that the variance estimate $\widehat{\mathrm{Var}}_l(\hat{p}_{0+})$ (or $\widehat{\mathrm{Var}}_l(\hat{p}_{+0})$) is different from $\widehat{\mathrm{Var}}_u(\hat{p}_{0+})$ (or $\widehat{\mathrm{Var}}_u(\hat{p}_{+0})$) when the CI $(l_{0+}, u_{0+})$ (or $(l_{+0}, u_{+0})$) is asymmetric about $\hat{p}_{0+}$ (or $\hat{p}_{+0}$) (Zou, Huang and Zhang, 2009).

Since $\mathrm{Var}(\hat{\Delta}) = \mathrm{Var}(\hat{p}_{0+}) + \mathrm{Var}(\hat{p}_{+0}) - 2\rho\sqrt{\mathrm{Var}(\hat{p}_{0+})\mathrm{Var}(\hat{p}_{+0})}$, where $\rho$ is the correlation coefficient of $\hat{p}_{0+}$ and $\hat{p}_{+0}$, the approximate $100(1-\alpha)\%$ confidence lower and upper limits for Δ based on the Wald-type statistic are given by

$$L = \hat{\Delta} - z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\Delta})}, \quad U = \hat{\Delta} + z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\Delta})}, \qquad (2.7)$$

respectively, where $\widehat{\text{Var}}(\hat{\Delta}) = \widehat{\text{Var}}(\hat{p}_{0+}) + \widehat{\text{Var}}(\hat{p}_{+0}) - 2\hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{p}_{0+})\widehat{\text{Var}}(\hat{p}_{+0})}$, and $\hat{\rho}$ is consistent estimator of $\rho$. Substituting (2.5) and (2.6) into (2.7) yields

$$L = \hat{\Delta} - \sqrt{(\hat{p}_{0+} - l_{0+})^2 + (u_{+0} - \hat{p}_{+0})^2 - 2\hat{\rho}(\hat{p}_{0+} - l_{0+})(u_{+0} - \hat{p}_{+0})}, \qquad (2.8)$$

$$U = \hat{\Delta} - \sqrt{(u_{0+} - \hat{p}_{0+})^2 + (\hat{p}_{+0} - l_{+0})^2 - 2\hat{\rho}(u_{0+} - \hat{p}_{0+})(\hat{p}_{+0} - l_{+0})}, \qquad (2.9)$$

where $\hat{\Delta} = \hat{p}_{0+} - \hat{p}_{+0}$. By Equations (2.8) and (2.9), it is necessary to compute $\hat{\rho}$ in evaluating confidence limits $L$ and $U$. Note that $\rho$ has the following expression

$$\rho = \frac{\text{Cov}(\hat{p}_{0+}, \hat{p}_{+0})}{\sqrt{\text{Var}(\hat{p}_{0+})\text{Var}(\hat{p}_{+0})}} = \frac{P_{EE}(-p_{01}p_{10} + p_{00}p_{11})}{NA_0\sqrt{\text{Var}(\hat{p}_{0+})\text{Var}(\hat{p}_{+0})}}. \qquad (2.10)$$

Hence, $\hat{\rho}$ can be evaluated by replacing $p_{ij}$, $P_{EE}$ and $P_{EI}$ in Equation (2.10) by their corresponding MLEs $\hat{p}_{ij}$ or $\tilde{p}_{ij}$, $\hat{P}_{EE}$ and $\hat{P}_{EI}$. That is, $\rho$ can be estimated by

$$\hat{\rho} = \frac{-\hat{p}_{01}\hat{p}_{10} + \hat{p}_{00}\hat{p}_{11}}{\sqrt{\left\{\hat{\mathbb{D}}_1 + \hat{P}_{IE}\hat{\mathbb{D}}_3 / \left(\hat{P}_{EE}\hat{\mathbb{D}}_2\right)\right\}\left\{\hat{\mathbb{D}}_2 + \hat{P}_{EI}\hat{\mathbb{D}}_3 / \left(\hat{P}_{EE}\hat{\mathbb{D}}_1\right)\right\}}}. \quad (2.11)$$

Also, to obtain confidence limits $L$ and $U$ of $\Delta$ via Equations (2.8) and (2.9), it is necessary to evaluate the confidence limits $l_{0+}$, $l_{+0}$, $u_{0+}$ and $u_{+0}$. To this end, we consider the following two methods.

(A) *The Wilson score confidence interval*

From Table 2 and model assumptions given in Section 2.1, we have $n_{0+} + n_{x0} \sim B(N, P_{E+} p_{0+})$ and $n_{+0} + n_{y0} \sim B(N, P_{+E} p_{+0})$, where $P_{E+} = P_{EE} + P_{EI}$ and $P_{+E} = P_{EE} + P_{IE}$. Thus, $p_{0+}$ and $p_{+0}$ can be estimated by $\breve{p}_{0+} = (n_{0+} + n_{x0})/(n + n_x)$ and $\breve{p}_{+0} = (n_{+0} + n_{y0})/(n + n_y)$, respectively. According to the Central Limit Theorem and Appendix C, $(N\hat{P}_{E+})^{1/2}(\breve{p}_{0+} - p_{0+})/\sqrt{p_{0+}(1-p_{0+})}$ is asymptotically distributed as the standard normal distribution, which implies

$$\mathrm{P}\left(\frac{(N\hat{P}_{E+})^{1/2}(\breve{p}_{0+} - p_{0+})}{\sqrt{p_{0+}(1-p_{0+})}} \le z_{\alpha/2}\right) = \mathrm{P}\left(\frac{(N\hat{P}_{E+})(\breve{p}_{0+} - p_{0+})^2}{p_{0+}(1-p_{0+})} \le z_{\alpha/2}^2\right) = 1-\alpha.$$

Hence, the lower $(l_{0+}^{ws})$ and upper $(u_{0+}^{ws})$ limits of the $100(1-\alpha)\%$ two-sided Wilson score CI for $p_{0+}$ are the smaller and larger roots to the following quadratic equation with respect to parameter $p_{0+} : (N\hat{P}_{E+})(\breve{p}_{0+} - p_{0+})^2/\{p_{0+}(1-p_{0+})\} = z_{\alpha/2}^2$, which yields

$$l_{0+}^{ws} = \breve{p}_{0+} - \frac{z_{\alpha/2}}{\tilde{n}_0}\sqrt{\mathcal{A} + \frac{z_{\alpha/2}^2}{4}} \quad \text{and} \quad u_{0+}^{ws} = \breve{p}_{0+} + \frac{z_{\alpha/2}}{\tilde{n}_0}\sqrt{\mathcal{A} + \frac{z_{\alpha/2}^2}{4}},$$

where $\mathcal{A} = N\hat{P}_{E+}\breve{p}_{0+}(1-\breve{p}_{0+})$, $\breve{p}_{0+} = (n_{0+} + n_{x0} + 0.5z_{\alpha/2}^2)/\tilde{n}_0$ and $\tilde{n}_0 = N\hat{P}_{E+} + z_{\alpha/2}^2$. Similarly, the lower $(l_{+0}^{ws})$ and upper $(u_{+0}^{ws})$ limits of the $100(1-\alpha)\%$ two-sided Wilson score CI for $p_{+0}$ can be expressed as

$$l_{+0}^{ws} = \breve{p}_{+0} - \frac{z_{\alpha/2}}{\tilde{n}_0}\sqrt{\mathcal{B} + \frac{z_{\alpha/2}^2}{4}} \quad \text{and} \quad u_{+0}^{ws} = \breve{p}_{+0} + \frac{z_{\alpha/2}}{\tilde{n}_1}\sqrt{\mathcal{B} + \frac{z_{\alpha/2}^2}{4}},$$

respectively, where $\mathcal{B} = N\hat{P}_{+E}\breve{p}_{+0}\left(1-\breve{p}_{+0}\right)$, $\breve{p}_{+0} = \left(n_{+0} + n_{y0} + 0.5z_{\alpha/2}^2\right)/\tilde{n}_1$ and $\tilde{n}_1 = N\hat{P}_{E+} + z_{\alpha/2}^2$.

(B) *The Agresti-Coull interval*

Following Tang, Li and Tang (2010), we consider the Agresti-Coull confidence intervals for $p_{0+}$ and $p_{+0}$. The lower $\left(l_{0+}^{ac}\right)$ and upper $\left(u_{0+}^{ac}\right)$ limits of the $100(1-\alpha)\%$ two-sided Agresti-Coull CI for $p_{0+}$ are given by

$$l_{0+}^{ac} = \breve{p}_{0+} - z_{\alpha/2}\sqrt{\breve{p}_{0+}\left(1-\breve{p}_{0+}\right)/\tilde{n}_0} \quad \text{and} \quad u_{0+}^{ac} = \breve{p}_{0+} + z_{\alpha/2}\sqrt{\breve{p}_{0+}\left(1-\breve{p}_{0+}\right)/\tilde{n}_0},$$

respectively. Similarly, the lower $\left(l_{+0}^{ac}\right)$ and upper $\left(u_{+0}^{ac}\right)$ limits of the $100(1-\alpha)\%$ two-sided Agresti-Coull CI for $p_{+0}$ are respectively given by

$$l_{+0}^{ac} = \breve{p}_{+0} - z_{\alpha/2}\sqrt{\breve{p}_{+0}\left(1-\breve{p}_{+0}\right)/\tilde{n}_1} \quad \text{and} \quad u_{+0}^{ac} = \breve{p}_{+0} + z_{\alpha/2}\sqrt{\breve{p}_{+0}\left(1-\breve{p}_{+0}\right)/\tilde{n}_1}.$$

(v) *Bootstrap-resampling-based confidence intervals*

Given the observed data $\boldsymbol{D} = \left\{n_{00}, n_{01}, n_{10}, n_{11}, n_{x0}, n_{x1}, n_{y0}, n_{y1}\right\}$, we can obtain the MLEs $\hat{p}_{00}$, $\hat{p}_{01}$, $\hat{p}_{10}$ and $\hat{p}_{11}$ of parameters $p_{00}$, $p_{01}$, $p_{10}$ and $p_{11}$ via Appendix A, and the naive MLEs $\hat{P}_{EE}$, $\hat{P}_{EI}$ and $\hat{P}_{IE}$ of parameters $P_{EE}$, $P_{EI}$ and $P_{IE}$ via $\hat{P}_{EE} = n/N$, $\hat{P}_{EI} = n_x/N$ and $\hat{P}_{IE} = n_y/N$, respectively. Based on $\hat{p}_{ij}\left(i, j=0,1\right)$, $\hat{P}_{EE}$, $\hat{P}_{EI}$, and $\hat{P}_{IE}$, we can generate a Bootstrap data set via the distribution: $\left(n^*_{00}, n^*_{01}, n^*_{10}, n^*_{11}, n^*_{x0}, n^*_{x1}, n^*_{y0}, n^*_{y1}\right)$ ~ Multinomial ($N$; $\hat{P}_{EE}\hat{p}_{00}$, $\hat{P}_{EE}\hat{p}_{01}$, $\hat{P}_{EE}\hat{p}_{10}$, $\hat{P}_{EE}\hat{p}_{11}$, $\hat{P}_{EI}\hat{p}_{0+}$, $\hat{P}_{EI}\hat{p}_{1+}$, $\hat{P}_{IE}\hat{p}_{+0}$, $\hat{P}_{IE}\hat{p}_{+1}$). For the generated Bootstrap sample ($n^*_{00}$, $n^*_{01}$, $n^*_{10}$, $n^*_{11}$, $n^*_{x0}$, $n^*_{x1}$, $n^*_{y0}$, $n^*_{y1}$), we first compute the MLEs $\hat{p}^*_{00}$, $\hat{p}^*_{01}$, $\hat{p}^*_{10}$ and $\hat{p}^*_{11}$ of parameters $p_{00}$, $p_{01}$, $p_{10}$, and $p_{11}$ via Appendix A, and then obtain the estimated value $\hat{\Delta}^*$ of $\Delta$ via $\hat{\Delta}^* = \hat{p}^*_{0+} - \hat{p}^*_{+0}$. Independently repeating the above process $G$ times, we obtain $G$ Bootstrap estimates $\{\hat{\Delta}^*_g : g = 1, 2, \ldots, G\}$. Let $\hat{\Delta}^*_{(1)}, \ldots, \hat{\Delta}^*_{(G)}$ denote the ordered values of $\hat{\Delta}^*_g$'s. Based on these $\hat{\Delta}^*_g$'s, we can construct different Bootstrap-resampling-based CIs for $\Delta$.

(A) *Bootstrap-resampling-based percentile confidence interval*

Following Shao and Tu (1995, p.132), the 100(1 − $\alpha$)% Bootstrap-resampling-based percentile CI for $\Delta$ is $\left(\hat{\Delta}^*_{([G\alpha/2])}, \hat{\Delta}^*_{([G(1-\alpha/2)])}\right)$, where [$a$] represents the integer part of $a$.

(B) *Bootstrap-resamping-based percentile-t confidence interval*

Let $S = (\mathrm{var}(\hat{\Delta}))^{1/2}$, and $S$* be the value of $S$ calculated from the generated Bootstrap sample. The Bootstrap distribution of $\hat{\Delta}^*$ can be defined as $F(x) = \Pr^*\{(\hat{\Delta}^* - \hat{\Delta}) / S^* \le x\}$, where Pr* is the conditional probability distribution given the original samples. Based on the generated $G$ Bootstrap samples, we can obtain $(\hat{\Delta}^* - \hat{\Delta}) / S^*$ and $\{t^*_g = (\hat{\Delta}^*_g - \hat{\Delta}) / S^*_g : g = 1, 2, \ldots, G\}$, where $S^*_g$

is the $g$th Bootstrap replication of $S$. Following Efron and Tibshirani (1993), the $100(1 - \alpha)\%$ Bootstrap-resampling-based percentile-t CI for $\Delta$ is given by $\left(\hat{\Delta} - \chi_{1-\alpha/2}\hat{S}, \hat{\Delta} - \chi_{\alpha/2}\hat{S}\right)$, where $\hat{S} = (\widehat{\text{Var}(\hat{\Delta})})^{1/2}$, $\chi_{\alpha/2}$ and $\chi_{1-\alpha/2}$ are the $100\alpha/2$ and $100(1-\alpha/2)$ percentiles of the empirical distribution of $t_g^*$, respectively.

**3 Monte Carlo simulation studies**

To investigate the performance of the proposed CI estimators of $\Delta$, we computed their empirical coverage probabilities (ECPs), empirical confidence widths (ECWs), and distal and mesial non-coverage probabilities (DNCPs and MNCPs) via extensive Monte Carlo simulation studies. Here, the *empirical coverage probability* was defined as

$$\text{ECP} = \frac{1}{J}\sum_{j=1}^{J} I[\Delta \in \{\Delta_L(\boldsymbol{D}^{(j)}), \Delta_U(\boldsymbol{D}^{(j)})\}],$$

where $\Delta_L(\boldsymbol{D}^{(j)})$ and $\Delta_U(\boldsymbol{D}^{(j)})$ were the lower and upper limits of CI for $\Delta$ based on the $j$th observed sample $\boldsymbol{D}^{(j)} = \left\{n_{00}^{(j)}, n_{01}^{(j)}, n_{10}^{(j)}, n_{11}^{(j)}, n_{x0}^{(j)}, n_{x1}^{(j)}, n_{y0}^{(j)}, n_{y1}^{(j)}\right\}$, which was randomly generated from the following multinomial distribution Mult($N$; $P_{EE}p_{00}$, $P_{EE}p_{01}$, $P_{EE}p_{10}$, $P_{EE}p_{11}$, $P_{EE}p_{0+}$, $P_{EE}p_{1+}$, $P_{IE}p_{+0}$, $P_{IE}p_{+1}$), and $I[\Delta \in \{\Delta_L, \Delta_U\}]$ was an indicator function of the event $[\Delta \in \{\Delta_L, \Delta_U\}]$ which was 1 if $\Delta \in \{\Delta_L, \Delta_U\}$, and 0 otherwise. The *empirical confidence width* was defined as

$$\text{ECW} = \frac{1}{J}\sum_{j=1}^{J} (\Delta_U(\boldsymbol{D}^{(j)}) - \Delta_L(\boldsymbol{D}^{(j)})).$$

When $\Delta > 0$, the *mesial and distal non-coverage probabilities* (Newcombe, 1998b) can be interpreted as the left and right non-coverage probabilities, respectively, which were defined by

$$\text{MNCP} = \frac{1}{J}\sum_{j=1}^{J} I\{\Delta < \Delta_L(\boldsymbol{D}^{(j)})\} \quad \text{and} \quad \text{DNCP} = \frac{1}{J}\sum_{j=1}^{J} I\{\Delta > \Delta_U(\boldsymbol{D}^{(j)})\},$$

respectively. When $\Delta < 0$, the *mesial and distal non-coverage probabilities* can be interpreted as the right and left non-coverage probabilities, respectively. The *ratio of the MNCP to the non-coverage probability* (*NCP*) was defined as

$$\text{RNCP} = \frac{\text{MNCP}}{\text{NCP}} = \frac{\text{MNCP}}{1.0 - \text{ECP}}.$$

Following Newcombe (1998b) and Tang, Li and Tang (2010), an interval can be regarded as *satisfactory* if (a) its ECP is close to the pre-specified 95% confidence level, (b) it possesses shorter interval width, and (c) its RNCP lies in the interval [0.4, 0.6], as *too mesially located* if its RNCP is less than 0.4, and *too distally* if its RNCP is greater than 0.6.

In the first Monte Carlo simulation study, we considered the following parameter settings: (i) $N$ = 20, 30, 50, 80, 100, 150; (ii) $p_{0+}$ was taken to be 0.3, 0.4, 0.6 and 0.7; (iii) $\Delta$ varied from −0.1 to 0.1 with step size being 0.05; (iv) $\rho$ was taken to be $\rho$ = −0.1, 0, 0.1, where $\rho$ was the correlation coefficient between the paired binary outcomes defined by $\rho = (p_{00} - p_{0+}p_{+0}) / (p_{0+}p_{1+}p_{+0}p_{+1})^{1/2}$; (v) ($P_{EE}$, $P_{EI}$, $P_{IE}$) = (0.8, 0.1, 0.1), (0.7, 0.2, 0.1), (0.7, 0.1, 0.2), (0.7, 0.15, 0.15), (0.6, 0.2, 0.2); (vi) the confidence level was set to be $1-\alpha = 95\%$. We generated a total of $J$ = 10000 replications for each combination of parameters. In computing

Bootstrap-resampling CIs, $G = 5000$ Bootstrap samples were generated. For each configuration of parameters *N*, $p_{0+}$, Δ, ρ, $P_{EE}$ $P_{EE}$, $P_{EI}$, $P_{IE}$, the observed data $\boldsymbol{D}^{(j)}$ of the *j*th replication $(j = 1, \ldots, 10000)$ were randomly generated from the multinomial distribution Mult$(N;\ P_{EE}\,p_{00}, P_{EE}\,p_{01}, P_{EE}\,p_{10}, P_{EE}\,p_{11}, P_{EI}\,p_{0+}, P_{EI}\,p_{1+}, P_{IE}\,p_{+0}, P_{IE}\,p_{+1})$ in which $p_{+0} = p_{0+} - \Delta$, $p_{00} = p_{0+}\,p_{+0} + \rho\left(p_{0+}\,p_{1+}\,p_{+0}\,p_{+1}\right)^{1/2}$, $p_{01} = p_{0+}\,p_{+1} - \rho\left(p_{0+}\,p_{1+}\,p_{+0}\,p_{+1}\right)^{1/2}$, $p_{10} = p_{1+}\,p_{+0} - \rho\left(p_{0+}\,p_{1+}\,p_{+0}\,p_{+1}\right)^{1/2}$, $p_{11} = p_{1+}\,p_{+1} + \rho\left(p_{0+}\,p_{1+}\,p_{+0}\,p_{+1}\right)^{1/2}$, $p_{+1} = 1 - p_{0+} + \Delta$ and $p_{1+} = 1 - p_{0+}$. Based on the generated samples $\left\{\boldsymbol{D}^{(j)} : j = 1, \ldots, 10000\right\}$, we calculated the 95% coverage probabilities, expected widths and RNCPs for the settings under consideration. Figures 1-3 presented box plots of ECPs, ECWs and RNCPs of various CIs. Here, each box plot contained 4 (i.e., the number of marginal probability $p_{0+}$) ×5 (i.e., the number of Δ's) ×3 (i.e., the number of ρ's) ×5 (i.e., the number of ($P_{EE}$, $P_{EI}$, $P_{IE}$)'s) = 300 data points.

To study the performance of our proposed CI estimators for Δ under the moderate/large correlation coefficients, we conducted the second simulation study under the following parameter settings: (i) *N* = 20, 30, 50, 80, 100, 150; (ii) $p_{0+} = 0.5$; (iii) $\Delta = -0.05$, 0, 0.05; (iv) $\rho = -0.9$, -0.6, -0.5, -0.1, 0, 0.1, 0.5, 0.6, 0.9. Here, we did not consider other values of $p_{0+}$ because some values of $p_{00}$, $p_{01}$, $p_{10}$ and $p_{11}$ may be negative under $p_{0+} \neq 0.5$. Results were presented in Table 3.

According to Figures 1–3 and Table 3, we obtained the following observations. For small sample sizes (e.g., see Figure 1), we found that the CIs based on the score statistic (i.e., $T_s$), the Wald-type statistic (i.e., $T_{w1}$) and the hybrid method with the Wilson score method (i.e., *MW*) produced deflated coverage probabilities (e.g., their median ECPs were less than 93%). Two bootstrap CIs (i.e., $B_1$ and $B_2$) and the Wald-type CI (i.e., $T_{w2}$) behaved satisfactorily in the sense that their median ECPs were close to the pre-specified confidence level 95%. The CIs based on the likelihood ratio method (i.e., $T_l$) and the hybrid method with the Agresti-Coull interval (i.e., *M A*) always guaranteed their median ECPs at or above the pre-specified confidence level. As sample size increased, median ECPs of all CIs except for $T_l$, $T_s$ and $T_{w1}$ became closer to the pre-specified confidence level. From Figure 2, we observed that median RNCPs of all CIs except for $T_l$ and $T_s$ generally lied in the interval [0.4, 0.6]. This showed that our derived CIs generally exhibited appropriate symmetry. From Figure 3, we observed that the CIs based on $T_l$ and $T_s$ generally yielded shorter median ECWs than other CIs. However, this may be due to their deflation in ECPs. Generally, the larger the sample size the narrower the confidence width. Also, the interval width increased as the proportion of missing observations increased. From Table 3, we found that (i) all mean interval widths decreased as the correlation (i.e., $\rho$) increased, (ii) there was no significant effect of $\rho$ on mean ECPs for CIs of $\Delta$ derived from $T_{w2}$, *MW*, *MA*, $B_1$ and $B_2$ methods, (iii) whilst there was a large effect of $\rho$ on mean ECPs for CIs of $\Delta$ derived from $T_l$, $T_s$ and $T_{w1}$ methods. For moderate values of $p_{0+}$, the mean coverage probabilities were closer to the pre-specified confidence level and the interval widths were generally wider. Finally, we did not observe significant effect of Δ on mean coverage probabilities and interval widths. In view of the above findings, we would recommend the hybrid CI with the wilson score interval (*MW*) and $T_{w2}$-CI as they (i) generally well controlled their coverage probabilities around the pre-chosen confidence level; (ii) consistently yielded shorter interval widths (even for small sample designs); and (iii) usually guaranteed their ratios of the MNCPs to the non-coverage probabilities lying in [0.4, 0.6]. Particularly, if one would like a CI that yields the shortest interval width, the hybrid CI with the Agresti-Coull interval is the optimal choice. If one would like a CI that yields less discrepancy in the ratio of the MNCP to the non-coverage probability, the $T_{w2}$-CI is the desirable candidate.

**4 An illustrative example**

In this section, the neurological study of meningitis patients introduced in Section 1 is used to illustrate the proposed methodologies. In this example, we are interested in CI construction of the difference between the incidence rates of neurological complication before and after the standard treatment. Under the previously given notation, we have $n_{00}=8$, $n_{01}=8$, $n_{10}=3$, $n_{11}=6$, $n_{x0}=2$, $n_{y0}=2$, $n_x=6$, $n_y=2$ and $N$ = 33. We calculated MLEs of the incidence rates of neurological complication before and after the standard treatment via the aforementioned EM algorithm, which were given by $\hat{p}_{0+}=0.6504$ and $\hat{p}_{+0}=0.4821$, respectively. Thus, an estimate of Δ was given by 0.1683. Various 95% CIs for $\Delta = p_{0+} - p_{+0}$ were presented in Table 4. According to Table 4, we observed that the incidence rates of neurological complication before and after the standard treatment were the same since all CIs except for the score-based CI include

0. Since the CIs based on *MA* and $T_{w2}$ methods are the most reliable according to our simulation studies, we have reason to believe that there is no significant difference between the incidence rates of neurological complication before and after the standard treatment at the 5% significance level.

## 5 Conclusion

In this paper, we considered the problem of CI construction for the difference between two correlated proportions in paired-comparison studies with missing observations. Under the assumption of MAR, we derived the score test statistic and proposed eight CI estimators for the difference between two correlated proportions in the presence of incomplete paired binary data based on the likelihood ratio method, the score test method, the Wald-type test method, the hybrid method with the Wilson score and Agresti-Coull intervals, and the Bootstrap-resampling method. Extensive simulation studies were conducted to evaluate the performance of the proposed CIs with respect to their empirical coverage probabilities (ECPs), empirical interval widths (ECWs) and ratios of the mesial non-coverage probabilities to the non-coverage probabilities (RNCPs). Based on our simulation results, we found that the hybrid CI with the wilson score interval (i.e., *MW*) and the Wlad-type CI with the constrained MLE behave satisfactorily for small to moderate sample sizes in the sense that their coverage probabilities could be well controlled around the pre-specified nominal confidence level and their RNCPs could be well controlled in the interval [0.4, 0.6]. Hence, they were recommended for practical applications when the coverage probability is of interest. Unlike the asymptotic score and likelihood ratio CIs, the proposed hybrid CIs possess analytical expressions and are thus recommended due to their computational simplicity. In particular, the hybrid CI with the Agresti-Coull interval is highly recommended in practice. We considered an analogue of the continuity correction of Newcombe (1998c) for the hybrid interval and omitted the corresponding results since the improvement is not significant.

In this paper, it is assumed that the missing mechanism is MAR. When the missing mechanism causing the incompleteness of the data depends on the treatment and the outcome, CI construction for the difference between two correlated proportions is not trivial and is under investigation.





**Acknowledgements**

The authors are grateful to the Editor, Associate Editor and two referees for providing some constructive comments and suggestions. This research was supported by grants from the National Science Fund for Distinguished Young Scholars of China (11225103), and Research Fund for the Doctoral Program of Higher Education of China (20115301110004).

**Appendix A: Maximum likelihood estimators of $p_{ij}$'s**

Let $\hat{p}_{ij}$ be the maximum likelihood estimator (MLE) of $p_{ij}$ for $i$, $j$ = 0, 1. It follows from Campbell (1984) and Chang (2009) that MLEs of $p_{00}$, $p_{01}$, $p_{10}$ and $p_{11}$ satisfy the following equations:

$$\hat{p}_{00} = \left\{ n_{00} + n_{x0}\hat{p}_{00} / \left( \hat{p}_{00} + \hat{p}_{01} \right) + n_{y0}\hat{p}_{00} / \left( \hat{p}_{00} + \hat{p}_{10} \right) \right\} / N,$$

$$\hat{p}_{01} = \left\{ n_{01} + n_{x0} \hat{p}_{01} / \left( \hat{p}_{00} + \hat{p}_{01} \right) + n_{y1} \hat{p}_{01} / \left( \hat{p}_{01} + \hat{p}_{11} \right) \right\} / N,$$

$$\hat{p}_{10} = \left\{ n_{10} + n_{y0} \hat{p}_{10} / \left( \hat{p}_{00} + \hat{p}_{10} \right) + n_{x1} \hat{p}_{10} / \left( \hat{p}_{10} + \hat{p}_{11} \right) \right\} / N,$$

$$\hat{p}_{11} = \left\{ n_{11} + n_{y1} \hat{p}_{11} / \left( \hat{p}_{01} + \hat{p}_{11} \right) + n_{x1} \hat{p}_{11} / \left( \hat{p}_{10} + \hat{p}_{11} \right) \right\} / N.$$

Thus, an EM algorithm for computing MLEs of $p_{ij}$'s can refer to Campbell (1984, p.314) and Chang (2009, p.794).

**Appendix B: Derivation of the score test statistic**

Let $\beta = p_{0+} - p_{+0} - \Delta$. The log-likelihood function of the observed data $\boldsymbol{D}$ can be rewritten as

$$\begin{aligned} l\left(\beta, p_{00}, p_{10}\right) = {} & n_{00} \log p_{00} + n_{01} \log\left(\beta + p_{10} + \Delta\right) + n_{10} \log p_{10} \\ & + n_{11} \log\left(1 - p_{00} - 2p_{10} - \beta - \Delta\right) + n_{x0} \log\left(p_{00} + \beta + p_{10} + \Delta\right) \\ & + n_{x1} \log\left(1 - p_{00} - \beta - p_{10} - \Delta\right) + n_{y0} \log\left(p_{00} + p_{10}\right) \\ & + n_{y1} \log\left(1 - p_{00} - p_{10}\right) + c, \end{aligned} \qquad (A.1)$$

where $c$ is the constant which is independent of parameters $p_{00}$, $p_{10}$ and $\beta$.

According to Equation $(A.1)$, the score function with respect to $\beta$ and the Fisher information matrix with respect to $\beta$, $p_{00}$ and $p_{10}$ under $\beta = 0$ are given by

$$\frac{\partial l}{\partial \beta}\Big|_{\beta=0} = \frac{n_{01}}{p_{10} + \Delta} - \frac{n_{11}}{1 - p_{00} - 2p_{10} - \Delta} + \frac{n_{x0}}{p_{00} + p_{10} + \Delta} - \frac{n_{x1}}{1 - p_{00} - p_{10} - \Delta},$$

and

$$
\boldsymbol{I} = \begin{pmatrix} N_{01} + N_{11} + a & N_{11} + a & N_{01} + 2N_{11} + a \\ N_{11} + a & N_{00} + N_{11} + a + b & 2N_{11} + a + b \\ N_{01} + 2N_{11} + a & 2N_{11} + a + b & N_{01} + N_{10} + 4N_{11} + a + b \end{pmatrix}, \quad (A.2)
$$

respectively, where $N_{ij} = n / p_{ij}$ for $i, j = 0,1$, $a = n_x / \{ p_{0+} (1 - p_{0+}) \}$, $b = n_y / \{ p_{+0} (1 - p_{+0}) \}$, $p_{01} = p_{10} + \Delta$, and $p_{11} = 1 - p_{00} - 2 p_{10} - \Delta$. It follows from Equation (A.2) that the upper left element $I^{11}$ of $\boldsymbol{I}^{-1}$ can be expressed as

$$
I^{11} = \frac{\mathbb{A}_2 + (a+b)\mathbb{A}_1}{\mathbb{B}_1 + \mathbb{A}_1 ab + \mathbb{B}_2 a + \mathbb{B}_3 b},
$$

where $\mathbb{A}_1 = N_{00} + N_{10} + N_{01} + N_{11}$, $\mathbb{A}_2 = (N_{00} + N_{11})(N_{01} + N_{10}) + 4N_{00} + N_{11}$, $\mathbb{B}_1 = N_{00}N_{01}N_{10} + N_{00}N_{01}N_{11} + N_{00}N_{10}N_{11} + N_{01}N_{10}N_{11}$, $\mathbb{B}_2 = (N_{00} + N_{01})(N_{10} + N_{11})$, $\mathbb{B}_3 = (N_{00} + N_{10})(N_{01} + N_{11})$, and $\mathbb{B}_4 = \mathbb{A}_1$. Hence, the score statistic for testing $H_0 : p_{0+} - p_{+0} = \Delta$ is given by

$$
\begin{aligned}
T_s(\Delta) &= \left( \partial l / \partial \beta |_{\beta=0} \right) \sqrt{I^{11}} \,|_{p_{00}=\tilde{p}_{00},\, p_{10}=\tilde{p}_{10},\, p_{01}=\tilde{p}_{01}+\Delta,\, p_{11}=\tilde{p}_{11}} \\
&= \left( \frac{n_{01}}{p_{01}} - \frac{n_{11}}{p_{11}} + \frac{n_{x0}}{p_{0+}} - \frac{n_{x1}}{p_{1+}} \right) \sqrt{\frac{\mathbb{A}_2 + (a+b)\mathbb{A}_1}{\mathbb{B}_1 + \mathbb{A}_1 ab + \mathbb{B}_2 a + \mathbb{B}_3 b}} \,|_{p_{00}=\tilde{p}_{00},\, p_{10}=\tilde{p}_{10},\, p_{01}=\tilde{p}_{10}+\Delta,\, p_{11}=\tilde{p}_{11}} .
\end{aligned}
$$

**Appendix C: Expectations and variances of $\breve{p}_{1+}$ and $\breve{p}_{+1}$**

It can be shown from properties of multinomial distribution that $n_{1+} + n_{x1} \mid n + n_x \sim B(n + n_x, p_{1+})$, $n_{+1} + n_{y1} \mid n + n_y \sim B(n + n_y, p_{+1})$, $n + n_x \mid N \sim B(N, P_{E+})$ and $n + n_y \mid N \sim B(N, P_{+E})$. By the delta method, we have $E\left\{\hat{P}_{E+}^{-1}\right\} \approx P_{E+}^{-1}$ and $E\left\{\hat{P}_{+E}^{-1}\right\} \approx P_{+E}^{-1}$. Then, according to the properties of expectation and variance and the above given expressions, we have

$$
\begin{aligned}
E\left\{\breve{p}_{1+}\right\} &= E_{n+n_x}\left\{E\left[\breve{p}_{1+}\middle| n+n_x\right]\right\} = E_{n+n_x}\left\{E\left[\frac{n_{1+}+n_{x1}}{n+n_x}\middle| n+n_x\right]\right\} = p_{1+}, \\
\operatorname{var}\left\{\breve{p}_{1+}\right\} &= \operatorname{var}_{n+n_x}\left\{E\left[\breve{p}_{1+}\middle| n+n_x\right]\right\} + E_{n+n_x}\left\{\operatorname{var}\left[\breve{p}_{1+}\middle| n+n_x\right]\right\} \\
&= \operatorname{var}_{n+n_x}\left\{p_{1+}\right\} + E_{n+n_x}\left\{\frac{p_{1+}\left(1-p_{1+}\right)}{n+n_x}\right\} \\
&= p_{1+}\left(1-p_{1+}\right)E_{n+n_x}\left\{\frac{1}{n+n_x}\right\} \\
&= \frac{p_{1+}\left(1-p_{1+}\right)}{N}E_{n+n_x}\left\{\frac{1}{\hat{P}_{E+}}\right\} \\
&\approx \frac{p_{1+}\left(1-p_{1+}\right)}{NP_{E+}}.
\end{aligned}
$$

Similarly, we can show that $E\left\{\breve{p}_{+1}\right\} = p_{+1}$ and $\operatorname{var}\left\{\breve{p}_{+1}\right\} \approx p_{+1}\left(1-p_{+1}\right)/\left(NP_{+E}\right)$.

**References**

Bhoj, D. S., Snijders, T. A. B. (1986). Testing equality of correlated proportions with incomplete data on both response. *Psychometrika* 51:579–588.

Campbell, G. (1984). Testing equality of proportions with incomplete correlated data. *Journal of Statistical Planning and Inference* 10:311–321.

Chang, M. (2009). Estimation of multiple response rates in phase II clinical trials with missing observations. *Journal of Biopharmaceutical Statistics* 19:791–802.

Choi, S. C., Stablein, D. M. (1982). Practical tests for comparing two proportions with incomplete data. *Applied Statistics* 31:256–262.

Choi, S. C., Stablein, D. M. (1988). Comparing incomplete paired binomial data under non-random mechanisms. *Statistics in Medicine* 7:929–939.

Ekbohm, G. (1982). On testing the equality of proportions in the paired case with incomplete data. *Psychometrika* 47:115–118.

Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Boca Raton: Chapman & Hall.

Lin, Y., Lipsitz, S., Sinha, D., Gawande, A. A., Regenbogen, S. E., Greenberg, C. C. (2009). Using Bayesian p-values in a $2\times2$ table of matched pairs with incompletely classified data. *Journal of the Royal Statistics Society Series* C 58:237–246.

Newcombe, R. G. (1998a). Two-sided confidence intervals for the single proportion: comparison of seven method. *Statistics in Medicine* 17:857–872.

Newcombe, R. G. (1998b). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 17:873–890.

Newcombe, R. G. (1998c). Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 17:2635–2650.

Pradhan, V., Menon, S., Das, U. (2013). Correlated profile likelihood confidence interval for binomial paired incomplete data. *Pharmaceutical Statistics* 12:48–58.

Shao, J., Tu, D. S. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Tang, M. L., Tang, N. S. (2004). Exact tests for comparing two paired proportions with incomplete data. *Biometrical Journal* 46:72–82.

Tang, M. L., Ling, M. H., Tian, G. (2009). Exact and approximate unconditional confidence intervals for proportion difference in the presence of incomplete data. *Statistics in Medicine* 28:625–641.

Tang, M. L., Li, H. Q., Tang, N. S. (2010). Confidence interval construction for proportion ratio in paired studies based on hybrid method. *Statistical Methods in Medical Research* 21:361–378.

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

Thomson, P. C. (1995). A hybrid paired and unpaired analysis for the comparison of proportions. *Statistics in Medicine* 14:1463–1470.

Zou, G. Y. (2008), Donner, A. (2008). Construction of confidence limits about effect measures: a general approach. *Statistics in Medicine* 27:1693–1702.

Zou, G. Y., Huang, W. Y., Zhang, X. H. (2009). A note on confidence interval estimation for a linear function of binomial proportions. *Computational Statistics and Data Analysis* 53:1080–1085.

Table 1. Neurological complication data from Choi and Stablein (1982).

| | $Y=0$ | $Y=1$ | Missing $Y$ | Total |
|---|---|---|---|---|
| $X=0$ | 8 | 8 | 4 | 20 |
| $X=1$ | 3 | 6 | 2 | 11 |
| Missing $X$ | 2 | 0 | – | 2 |
| Total | 13 | 14 | 6 | 33 |

$X$ = beginning of treatment; $Y$ = end of treatment.

Table 2 Observed frequencies for a matched-pair design with missing observations

| | $Y=0$ | $Y=1$ | Missing $Y$ | Total |
|---|---|---|---|---|
| $X=0$ | $n_{00}$ | $n_{01}$ | $n_{x0}$ | $n_{0+}+n_{x0}$ |
| $X=1$ | $n_{10}$ | $n_{11}$ | $n_x-n_{x0}$ | $n_{1+}+n_x-n_{x0}$ |
| Missing $X$ | $n_{y0}$ | $n_y-n_{y0}$ | $z=0$ | $n_y$ |
| Total | $n_{+0}+n_{y0}$ | $n_{+1}+n_y-n_{y0}$ | $n_x$ | $N=n+n_x+n_y$ |

Table 3 Mean ECPs and ECWs of various confidence intervals with different $\rho$.

| | $\rho$ | $T_l$ | $T_s$ | $T_{w1}$ | $T_{w2}$ | $MW$ | $MA$ | $B_1$ | $B_2$ |
|---|---|---|---|---|---|---|---|---|---|
| ECP | −0.9 | 0.9684 | 0.9668 | 0.9613 | 0.9532 | 0.9547 | 0.9617 | 0.9635 | 0.9434 |
| | −0.6 | 0.9693 | 0.9686 | 0.9581 | 0.9502 | 0.9556 | 0.9562 | 0.9607 | 0.9415 |
| | −0.5 | 0.9690 | 0.9695 | 0.9588 | 0.9509 | 0.9552 | 0.9557 | 0.9612 | 0.9423 |
| | −0.1 | 0.9671 | 0.9680 | 0.9565 | 0.9480 | 0.9524 | 0.9528 | 0.9578 | 0.9417 |
| | 0 | 0.9663 | 0.9677 | 0.9558 | 0.9474 | 0.9516 | 0.9518 | 0.9574 | 0.9414 |
| | 0.1 | 0.9667 | 0.9678 | 0.9569 | 0.9487 | 0.9527 | 0.9532 | 0.9569 | 0.9417 |
| | 0.5 | 0.9422 | 0.9468 | 0.9646 | 0.9559 | 0.9569 | 0.9573 | 0.9616 | 0.9509 |
| | 0.6 | 0.9115 | 0.9221 | 0.9675 | 0.9595 | 0.9563 | 0.9569 | 0.9628 | 0.9551 |
| | 0.9 | 0.9035 | 0.9106 | 0.9955 | 0.9504 | 0.9517 | 0.9521 | 0.9657 | 0.9601 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ECW | −0.9 | 0.3724 | 0.3702 | 0.5187 | 0.5128 | 0.5100 | 0.5108 | 0.4883 | 0.4893 |
| | −0.6 | 0.3562 | 0.3542 | 0.4848 | 0.4793 | 0.4659 | 0.4665 | 0.4579 | 0.4589 |
| | −0.5 | 0.3494 | 0.3482 | 0.4727 | 0.4672 | 0.4516 | 0.4525 | 0.4470 | 0.4477 |
| | −0.1 | 0.3179 | 0.3181 | 0.4175 | 0.4118 | 0.3938 | 0.3942 | 0.3963 | 0.3970 |
| | 0 | 0.3082 | 0.3091 | 0.4015 | 0.3957 | 0.3782 | 0.3787 | 0.3813 | 0.3818 |
| | 0.1 | 0.2972 | 0.2993 | 0.3846 | 0.3785 | 0.3620 | 0.3625 | 0.3653 | 0.3659 |
| | 0.5 | 0.2391 | 0.2466 | 0.3037 | 0.2954 | 0.2863 | 0.2867 | 0.2874 | 0.2874 |
| | 0.6 | 0.2195 | 0.2295 | 0.2791 | 0.2695 | 0.2623 | 0.2628 | 0.2628 | 0.2623 |
| | 0.9 | 0.1366 | 0.1568 | 0.1887 | 0.1633 | 0.1638 | 0.1641 | 0.1620 | 0.1599 |

Table 4 Various 95% CIs for $p_{1+} - p_{+1}$ based on the neurological data set.

| | $T_l$ | $T_s$ | $T_{w1}$ | $T_{w2}$ | $MW$ | $MA$ | $B_1$ | $B_2$ |
|---|---|---|---|---|---|---|---|---|
| Lower | −0.0740 | 0.0301 | −0.0653 | −0.0779 | −0.0908 | −0.0905 | −0.0761 | −0.0856 |
| Upper | 0.3966 | 0.3927 | 0.4019 | 0.3946 | 0.3522 | 0.3518 | 0.3972 | 0.4275 |
| Width | 0.4706 | 0.3626 | 0.4672 | 0.4725 | 0.4430 | 0.4423 | 0.4733 | 0.5131 |

Figure 1: ECPs of various confidence interval estimates for different total sample size (N)

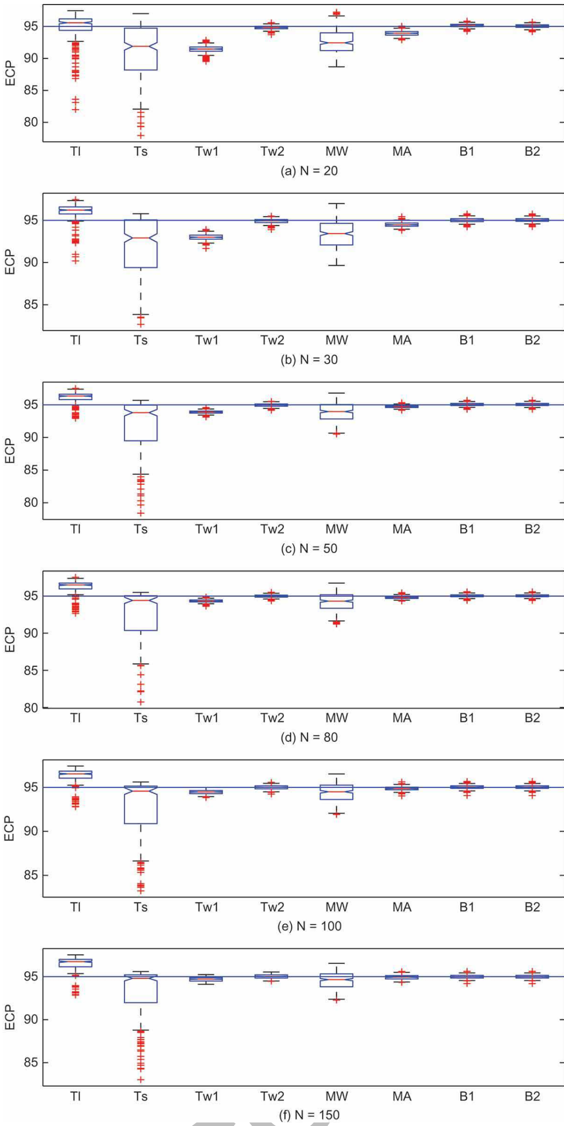

Figure 2: RNCPs of various confidence interval estimates for different total sample size (N)

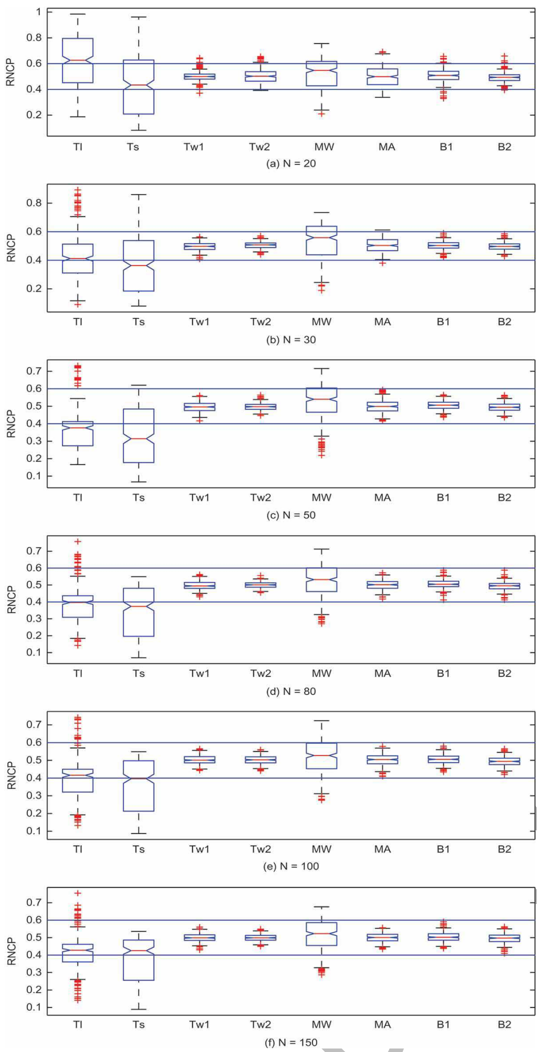

Figure 3: ECWs of various confidence interval estimates for different total sample size (N)