# Exact and Asymptotic Statistical Tests for Proportion Difference in One-to-Two Matched Binary Data

Hui Lin [*], Chong Wang[†]

November 20, 2014

**Abstract** Matched observations with dichotomous responses commonly occur in medical and epidemiological researches. Although standard approaches exist for one-to-one paired binary data analyses, there is not much work for one-to-two or one-to-N matched binary data in the current statistical literature. The existing Miettinen's test assumes that the multiple observations from the same matched set are mutually independent. In this paper, we propose exact and asymptotic tests for one-to-two matched binary data. Our method is markedly different from existing methods in that ours does not rely on a mutual independence assumption. The emphasis on dependence among observations from the same matched set is natural and appealing, as much in human health as it is in veterinary medicine. It can be applied to many types of diagnostic studies with one-to-two matched data structure. Our methods can be generalized to one-to-N matched case in a straightforward manner.

**Keywords:** exact test, asymptotic test, matched binary data, diagnostic studies, pooled sample

## 1 Introduction

Matched observations with dichotomous responses commonly occur in medical and epidemiological researches. Although standard approaches exist for one-to-one paired binary data analyses, not much research on one-to-two or one-to-N matched binary data has been published. Our research was originally motivated by the pooling of diagnostic tests. Often, testing units one-by-one is inefficient, especially when disease prevalence is sufficiently low. The concept of screening pooled samples originated during the second world war to detect syphilis in US soldiers [1]. It has aroused significant amount of attention and been used successfully in various applications. Many studies have demonstrated the successful use of pooling strategies on HIV testing [2] [3] [4] [5]. Budget reduction is an important issue which limits the number of tests so that the derived estimates can be imprecise. One way to overcome budget limitations and improve the accuracy of estimates is pooled testing. Vansteelandt *et al.* [6] showed that a good design can severely reduce cost. An applied example in Vansteelandt *et al.* [6] showed that using test pools with an average of seven units reduced cost by 44 percent with virtually no loss in precision. In some circumstances, the advantages of pooling include earning more accuracy as well[2].

---

[*]hui.lin@pioneer.com, DuPont Pioneer, Johnston, IA 50131,USA

[†]chwang@iastate.edu, Department of Statistics Iowa State University Ames, IA 50011,USA

For the one-to-one case, McNemar [7] developed a test of marginal homogeneity in a $2 \times 2$ table that is applicable to pair-matched observations or a cohort measured twice on a variable with binary outcome. Bennett and Underwood [8] compared exact power with the non-central Chi-square approximation for sample sizes of 10, 20 and 40 and found the Chi-square approximation to be adequate. Miettinen [9] derived the asymptotic power for testing the difference between cases and controls with dichotomous response in the case of one to one and one to R matching. Stephen [10] derived the exact power based on Miettinen's work and compared it to the asymptotic power of the test. However, Miettinen's test assumes that the multiple observations from the same matched set are mutually independent conditioned on the pair. This assumption is hard to hold for pooling test data where the pooled sample is of course dependent of the individual samples being pooled. Furthermore, the independence assumption can be assessed statistically using Fisher's exact test and our data show significant evidence of dependence.

We proposed exact and asymptotic tests for one-to-two matched binary data. Our methods fit a more general situation that does not assume that observations from the same subject are mutually independent. It can be applied to many types of diagnostic studies with one-to-two matched data structures besides dual sample pooling, such as one-to-two case control studies. It is important to understand the properties of matching designs so as to be able to make the best use of them. Our methods can be generalized to one-to-N matched cases. For clarity of presentation we establish basic concepts, terminology and notation in Section 2. We illustrate the exact test procedure and an asymptotic test procedure in Section 3 and Section 4, respectively. In Section 5, we demonstrate the merits of our tests through a simulation study. In Section 6, we applied our methods on two practical situations that fail to have the independence required by Miettinen's test. Discussion follows in Section 7.

## 2  Basic concepts, Terminology and Notation

Assume we have $n$ subjects going through two strategies of test. By saying one-to-two we mean there is one binary observation taken from each subject under strategy 1 and two binary observations taken from the same subject under strategy 2. In the paper, we use upper case letters to denote random variables and lower case letters to denote observed realizations. We denote the set of three observations from subject $j$ and its realization as:

$$(Y_{1j}, Y_{2j1}, Y_{2j2}) \ and \ (y_{1j}, y_{2j1}, y_{2j2})$$

respectively, where $Y_{1j}$ denotes the observation under strategy 1 while $Y_{2j1}$ and $Y_{2j2}$ denote observations under strategy 2 with $j = 1, .., n$.

For the $j^{th}$ matched group a realization $(y_{1j}, y_{2j1}, y_{2j2})$ is obtained for the random response vector $(Y_{1j}, Y_{2j1}, Y_{2j2})$. The value of the response variable $Y$ is either 0 or1. And $p_1 = Pr\{Y_{1j} = 1\}$ (probability a subject under strategy 1 has test result 1) and $p_2 = Pr\{Y_{2j1} = 1\} = Pr\{Y_{2j2} = 1\}$ (probability a subject under strategy 2 has test result 1). The object of the study is to make inferences about

$$\delta = p_1 - p_2,$$

and test the null hypothesis

$$H_0 : \delta = 0$$

We consider the multinomial distribution of the response vector $(X_{1j}, X_{2j})$ where $X_{1j} = Y_{1j}$ and $X_{2j} = Y_{2j1} + Y_{2j2}$. There are six possible realizations and denote $Z_{kl}^{(j)} = I(X_{1j} = k, \ X_{2j} = l)$ with

2

$k = 0, 1$ and $l = 0, 1, 2$ . It is a multinomial distribution with $Z_{kl}^{(j)} \sim multi - Bernoulli(p_{kl})$ where $p_{kl} = E[Z_{kl}^{(j)}]$ is invariant across units denoted by $j$. The cell counts for a total of $n$ sets. $Z_{kl} = \sum_{j=1}^{n} Z_{kl}^{(j)}$ has a multinomial distribution with $Z_{kl} \sim multinomial(n, p_{kl})$ .

Table 1: Outcome for Subject $j$

|  |  | Test 2 |  |
|---|---|---|---|
| Test 1 | $Z_{12}^{(j)}$ | $Z_{11}^{(j)}$ | $Z_{10}^{(j)}$ |
|  | $Z_{02}^{(j)}$ | $Z_{01}^{(j)}$ | $Z_{00}^{(j)}$ |

Table 2: Counting Table for n Sets of Observations

|  |  | Test 2 | | | |
|---|---|---|---|---|---|
|  |  | 2 | 1 | 0 | Total |
| Test 1 | 1 | $Z_{12}$ | $Z_{11}$ | $Z_{10}$ | $n_{1.}$ |
|  | 0 | $Z_{02}$ | $Z_{01}$ | $Z_{00}$ | $n_{0.}$ |
|  | Total | $n_{.2}$ | $n_{.1}$ | $n_{.0}$ | $n$ |

## 2.1 Miettinen Exact Test

Miettinen proposed an exact test for this matching design under the following two assumptions:

1. the $n$ vectors $(Y_{1j}, Y_{2j1}, Y_{2j2})$ are independently and identically distributed, and that

2. $Y_{1j}, Y_{2j1}, Y_{2j2}$ are mutually independent conditionally on $(\pi_1, \pi_2) = (\pi_{1j}, \pi_{2j})$ where $p_1 = E(\pi_1), p_2 = E(\pi_2)$.

Miettinen [9] proposed an exact test based on the multinomial formulation. Conditioning on $S_1 = Z_{10} + Z_{01}$ and $S_2 = Z_{11} + Z_{02}$, $Z_{10}$ and $Z_{11}$ have independent binomial distributions. Under $H_0$,

$$Z_{10} \sim Binomial(S_1, \frac{1}{3});$$

$$Z_{11} \sim Binomial(S_2, \frac{2}{3}).$$

The computation of the p-value for hypothesis testing is: $p = Pr(Z_{10} + Z_{11} \geq z_{10} + z_{11} = v)$ i.e.

$$p = \sum_{k_1 + k_2 \geq v} \begin{pmatrix} s_1 \\ k_1 \end{pmatrix} \left(\frac{1}{3}\right)^{k_1} \left(\frac{2}{3}\right)^{s_1 - k_1} \begin{pmatrix} s_2 \\ k_2 \end{pmatrix} \left(\frac{2}{3}\right)^{k_2} \left(\frac{1}{3}\right)^{s_2 - k_2} \tag{1}$$

When Test 1 and Test 2 results are biologically related, as in a pooling test senario, the assumption of independence between Test 1 and Test 2 may not be reasonable. Paired test analysis methods such as McNemar's test do not generally require independence between paired test results. In the following sections, we discuss statistical tests without the conditional independence assumption.

3

# 3 Random Exact Test

## 3.1 Test Statistic

Let $R_{kl}^{(j)} \mid Z_{kl}^{(j)} \sim Bin(Z_{kl}^{(j)}, \frac{1}{2})$ and $R_{kl} = \sum_{j=1}^{n} R_{kl}^{(j)}$. The marginal probability is
$Pr\{R_{kl}^{(j)} = 1\} = Pr\{R_{kl}^{(j)} = 1 \mid Z_{kl}^{(j)} = 1\}Pr\{Z_{kl}^{(j)} = 1\} = \frac{p_{kl}}{2}$. So for $k \neq k'$ or $l \neq l'$,

$$Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 2\} = Pr\{Z_{k'l'}^{(j)} = 1, R_{kl}^{(j)} = 1\} = 0 \tag{2}$$

$$
\begin{aligned}
Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 1\} &= Pr\{Z_{k'l'}^{(j)} = 1, R_{kl}^{(j)} = 0\} + Pr\{Z_{k'l'}^{(j)} = 0, R_{kl}^{(j)} = 1\} \\
&= Pr\{Z_{k'l'}^{(j)} = 1\}Pr\{R_{kl}^{(j)} = 0 \mid Z_{k'l'}^{(j)} = 1\} + Pr\{R_{kl}^{(j)} = 1\}Pr\{Z_{k'l'}^{(j)} = 0 \mid R_{kl}^{(j)} = 1\} \\
&= Pr\{Z_{k'l'}^{(j)} = 1\} + Pr\{R_{kl}^{(j)} = 1\} = p_{k'l'} + \frac{p_{kl}}{2}
\end{aligned}
\tag{3}
$$

$$Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 0\} = 1 - Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 1\} = 1 - (p_{k'l'} - \frac{p_{kl}}{2}) \tag{4}$$

Then we have $\sum_{j=1}^{n}(Z_{k'l'}^{(j)} + R_{kl}^{(j)}) = Z_{k'l'} + R_{kl} \sim Bin(n, p_{k'l'} + \frac{p_{kl}}{2})$, for any $(k, l) \neq (k', l')$.

$$
\begin{aligned}
\delta &= E(X_{1j}) - \frac{E(X_{2j})}{2} \\
&= (p_{12} + p_{11} + p_{10}) - (p_{12} + p_{02} + \frac{1}{2}p_{11} + \frac{1}{2}p_{01}) \\
&= p_{10} - \frac{1}{2}p_{01} + \frac{1}{2}p_{11} - p_{02}
\end{aligned}
\tag{5}
$$

Denote $S = Z_{10} + R_{11} + Z_{02} + R_{01}$ and $p_s = p_{10} + \frac{p_{11}}{2} + p_{02} + \frac{p_{01}}{2}$. Under $H_0$:
$p_{10} + \frac{1}{2}p_{11} = p_{02} + \frac{1}{2}p_{01}$, we have $Z_{10} + R_{11} \mid S \sim Bin(S, \frac{1}{2})$. A two-sided Random Exact Test is done through the following three steps:

1. Random sample $r_{11} \mid \sim Bin(z_{11}, \frac{1}{2})$ and $r_{01} \sim Bin(z_{01}, \frac{1}{2})$

2. Denote $s_1 = max(z_{10} + r_{11}, z_{02} + r_{01})$, $s_2 = min(z_{10} + r_{11}, z_{02} + r_{01})$ and
   $s = z_{10} + r_{11} + z_{02} + r_{01}$.

3. Calculate p-value by $Pr\{x \leq s_2 \text{ or } x \geq s_1\}$ with $x \sim Bin(s, \frac{1}{2})$.

Due to the randomization of $r_{11}$, the procedure can give different answers for the exact same data. We can avoid the arbitrariness of randomization while keeping the beautiful theory of these procedures by a simple change of viewpoint to what is called a "fuzzy p-value" advanced by Geyer & Meeden (2005) [11]. Different from conventional p-values, fuzzy p-values are random variables interpreted as p-values. In terms of the random exact test illustrated above, $r_{11}$ is called a latent variable and the p-value calculated in step 3 is refered to as a latent p-value. The latent p-value would be a p-value if the values of the latent variable were observed. The exact test employing the notion of a fuzzy p-value uses simulations of the latent under the null hypothesis. It provides an expression of both the strength and the uncertainty of the evidence against the null hypothsis.

## 3.2 Power of the Random Exact Test

For fixed $\delta$, $p_1$, $p_{12}$ and $p_{11}$,

$$p_{01} = 2 * p_1 * (1 - p_1) - p_{11} \tag{6}$$

$$p_{02} = p_1^2 - p_{12} \tag{7}$$

$$p_{10} = p_1 + \delta - p_{12} - p_{11} \tag{8}$$

$$p_{00} = (1 - p_1)^2 - (p_1 + \delta - p_{12} - p_{11}) \tag{9}$$

We have shown that $Z_{10} + R_{11} \sim Bin(n, p_{10} + \frac{p_{11}}{2})$ and $Z_{02} + R_{01} \sim Bin(n, p_{02} + \frac{p_{01}}{2})$.

$$p_s = p_{10} + \frac{p_{11}}{2} + p_{02} + \frac{p_{01}}{2} = 2p_1 + \delta - p_{11} - 2p_{12} \tag{10}$$

$$\frac{p_{10} + \frac{p_{11}}{2}}{p_{10} + \frac{p_{11}}{2} + p_{02} + \frac{p_{01}}{2}} = \frac{p_1 - p_{12} - \frac{p_{11}}{2} + \delta}{2p_1 + \delta - p_{11} - 2p_{12}} \equiv q \tag{11}$$

Then $S \sim Bin(N, p_s)$ and $Z_{10} + R_{11} \mid S \sim Bin(S, q)$. The unconditional power can be obtained as the expectation of the conditional power. The power expression of the exact binomial test is ( 12).

$$
\begin{aligned}
&Pr\{Z_{10} + R_{11} \leq u_{\alpha/2} \ or \ Z_{10} + R_{11} \geq u_{1-\alpha/2}\} \\
&= \sum_{S=0}^{n} \binom{n}{S} p_s^S (1 - p_s)^{n-S} \sum_{\substack{Z_{10}+R_{11} \leq l_{\alpha/2}, \\ Z_{10}+R_{11} \geq u_{\alpha/2}}} \binom{S}{Z_{10}+R_{11}} q^{Z_{10}+R_{11}} (1 - q)^{S-(Z_{10}+R_{11})}
\end{aligned} \tag{12}
$$

where $l_{\alpha/2} \doteq max\{n \mid \sum_{x=0}^{n} \binom{S}{x}(\frac{1}{2})^S \leq \frac{\alpha}{2}\}$ and $u_{\alpha/2} \doteq min\{n \mid \sum_{x=n}^{S} \binom{S}{x}(\frac{1}{2})^S \leq \frac{\alpha}{2}\}$.

## 4  Asymptotic Test

Denote $T^{(j)} = (Z_{10}^{(j)} + \frac{Z_{11}^{(j)}}{2}) - (Z_{02}^{(j)} + \frac{Z_{01}^{(j)}}{2})$.

$$E[T^{(j)}] = p_{10} + \frac{p_{11}}{2} - p_{02} - \frac{p_{01}}{2} = \delta \tag{13}$$

$$
\begin{aligned}
Var[T^{(j)}] &= E[T^{(j)2}] - \{E[T^{(j)}]\}^2 \\
&= E[\{(Z_{10}^{(j)} + \frac{Z_{11}^{(j)}}{2}) - (Z_{02}^{(j)} + \frac{Z_{01}^{(j)}}{2})\}^2] - \delta^2 \\
&= E[(Z_{10}^{(j)} + \frac{Z_{11}^{(j)}}{2})^2 + (Z_{02}^{(j)} + \frac{Z_{01}^{(j)}}{2})^2 - 2(Z_{10}^{(j)} + \frac{Z_{11}^{(j)}}{2})(Z_{02}^{(j)} + \frac{Z_{01}^{(j)}}{2})] - \delta^2
\end{aligned} \tag{14}
$$

Since at most one of $\{Z_{10}^{(j)}, Z_{11}^{(j)}, Z_{02}^{(j)}, Z_{01}^{(j)}\}$ is non-zero, $Z_{kl}^{(j)} Z_{k'l'}^{(j)} = 0$ if $k \neq k'$ or $l \neq l'$. Also we have $Z_{kl}^{(j)2} = Z_{kl}^{(j)}$. Then ( 14) can be written as ( 15).

$$Var[T^{(j)}] = E[Z_{10}^{(j)2} + \frac{Z_{11}^{(j)2}}{4} + Z_{02}^{(j)2} + \frac{Z_{01}^{(j)2}}{4})] - \delta^2 = p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2 \tag{15}$$

Since observations from different subjects are independent, we have:

$$\mu = E \sum_{j=1}^{n} T^{(j)} = n\delta \tag{16}$$

$$\sigma^2 = Var \sum_{j=1}^{n} T^{(j)} = n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2) \tag{17}$$

Under $H_0$, By CLT, $\frac{\sum_{j=1}^{n} T^{(j)}}{\sqrt{Var(\sum_{j=1}^{n} T^{(j)})}} = \frac{(Z_{10} + \frac{Z_{11}}{2}) - (Z_{02} + \frac{Z_{01}}{2})}{\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4})}}$ is asymptotic standard normal when $n$ is large. The asymptotic test is to compare the following test statistics to $N(0,1)$.

$$\frac{(Z_{10} + \frac{Z_{11}}{2}) - (Z_{02} + \frac{Z_{01}}{2})}{\sqrt{z_{10} + \frac{z_{11}}{4} + z_{02} + \frac{z_{01}}{4}}} \tag{18}$$

When $\delta \neq 0$, $\frac{(Z_{10} + \frac{Z_{11}}{2}) - (Z_{02} + \frac{Z_{01}}{2}) - n\delta}{\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2)}}$ is asymptotic standard normal . The power with respect to $\delta$ is a function of the mean and variance of the test statistic

$$\beta = 2\Phi(\frac{\phi_{\alpha/2}\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4})} - n\delta}{\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2)}}) \tag{19}$$

where $\phi_{\alpha/2}$ is the $\alpha/2$ lower quantile of standard normal distribution and $\Phi(\cdot)$ is the cumulative density function of standard normal distribution. It is time consuming to compute ( 12) and ( 19) for large n. Therefore, in the following simulation study, we estimate the power of exact and asymptotic tests through Monte Carlo sampling.

## 5    Simulation

We conducted a Monte Carlo study to examine type one error levels and power of the proposed statistical tests. In particular, we took $n = 10, 20, 30, 50, 100, 200, 300$. Since the exact power depend on the individual binomial or multinomial parameters, the arbitrary choice of a further parameter is necessary. We gave arbitrary values for $p_1$, $p_{12}$ , $p_{11}$ and $\delta$, then the rest parameters are determined by solving ( 6) - ( 9). We consider different parameterizations according to ( 20) - ( 24).

$$\delta = 0.05 * (h - 1), \ where \ h = 1, 2, 3, 4, 5 \tag{20}$$

For each $\delta$ value in ( 20), we simulate sample from four different settings:

$$Setting \ 1: \ p_1 = 0.3, p_{12} = 0.01, p_{11} = 0.01 \tag{21}$$

$$Setting \ 2: \ p_1 = 0.3, p_{12} = 0.08, p_{11} = 0.15 \tag{22}$$

$$Setting \ 3: \ p_1 = 0.4, p_{12} = 0.15, p_{11} = 0.24 \tag{23}$$

$$Setting \ 4: \ p_1 = 0.4, p_{12} = 0.11, p_{11} = 0.03 \tag{24}$$

Note that the $4^{th}$ setting can only get 3 $\delta$ values (i.e $\delta = 0, 0.05, 0.1$) due to the support of parameter ( $[0, 1]$). For each case, $M = 2000$ simulations were performed. Function "rmultinom()" in R is used to simulate multinomial samples. As mentioned, it is computationally infeasible to
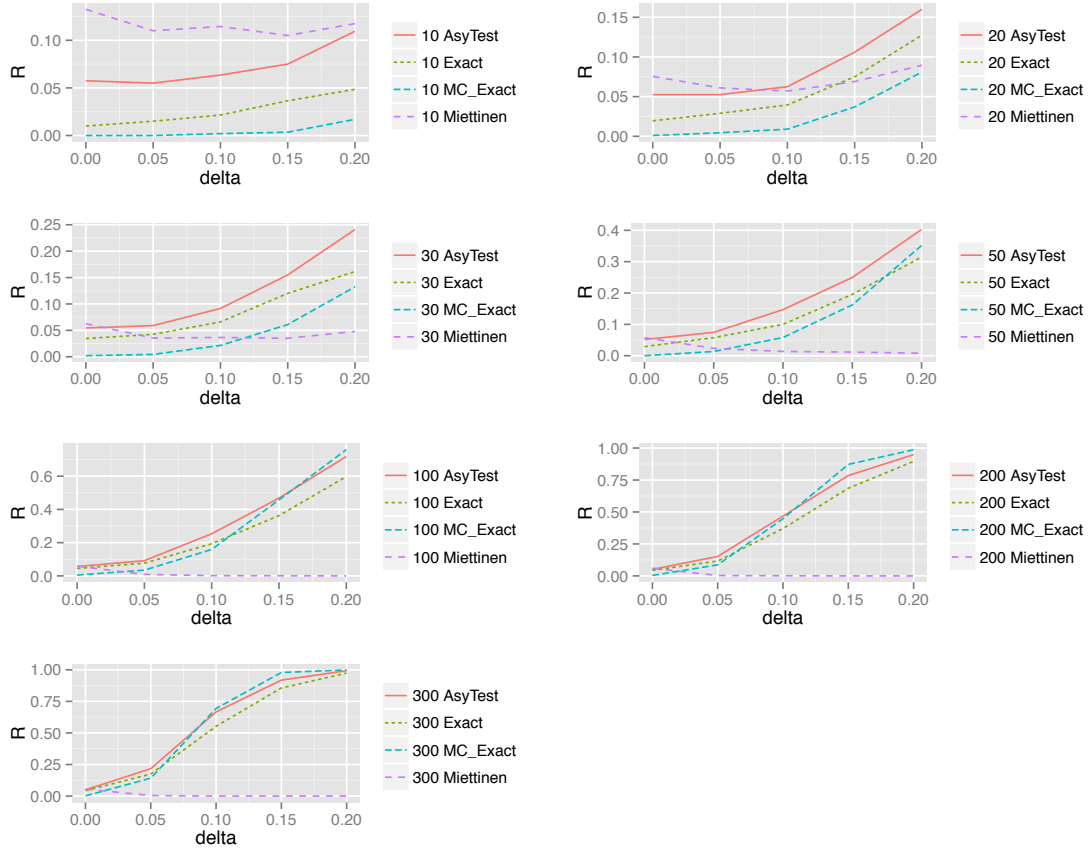
6

Figure 1: Comparison of exact test power and asymptotic test power for setting 1 of one-to-two case. The numbers in the legend indicate the sample size. AsyTest indicates asymptotic test; Exact indicates exact test.

calculate ( 12) for large $n$. Therefore, the power for each test on each sample set was estimated with 2000 simulations.

The resulting power values of the exact test and the asymptotic test for different parameterizations are shown in Figure 1 to 4. From the results we can see that the Miettinen's Test does not work here. The asymptotic test consistently dominates the others for all settings. Though for small sample size ($n \leq 30$) the performance drops, our proposed tests perform well as long as the size is larger than 50. As $\delta$ increases, the power increases steadily for the exact binomial test and asymptotic test as we would expect. Miettnen's test nearly has no power even the sample size is large except for setting 3. The reason may be that the design of parameterization for setting 3 approximate the mutually independent situation the most. A full
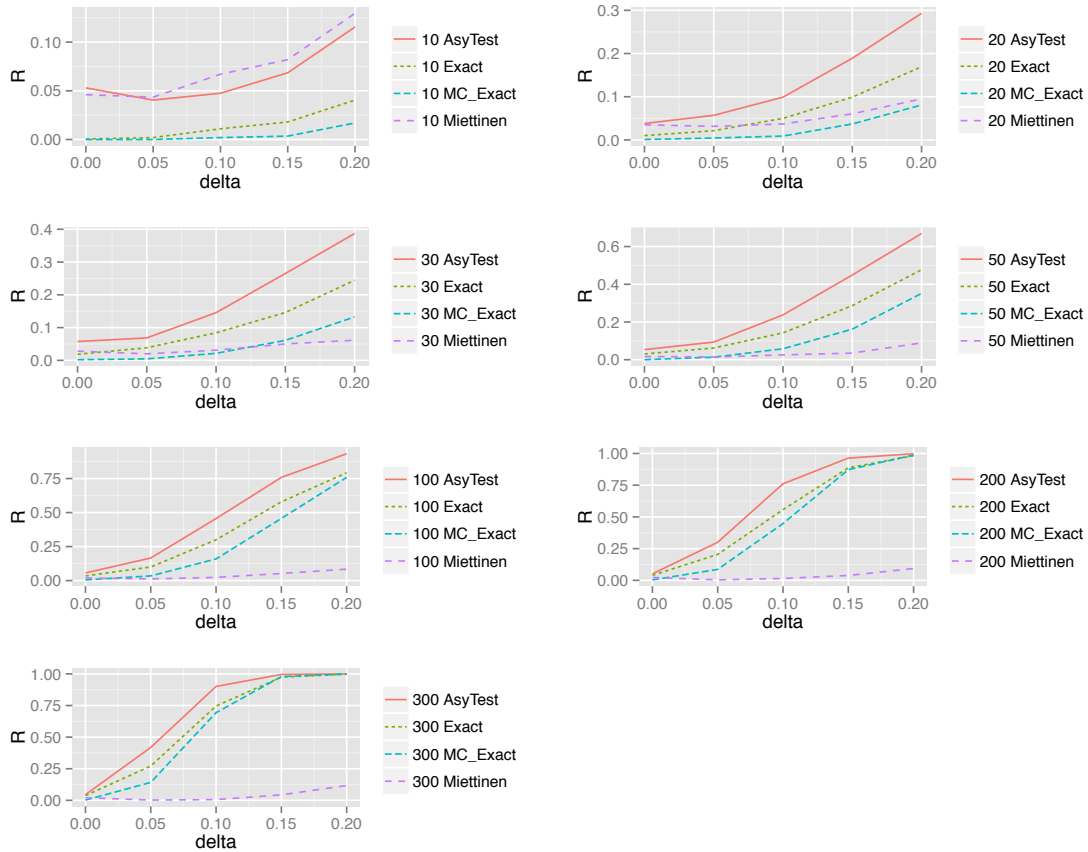
Figure 2: Comparison of exact test power and asymptotic test power for setting 2 of one-to-two case. The numbers in the legend indicate the sample size. AsyTest indicates asymptotic test; Exact indicates exact test.

table of parameterization strategy for setting 3 under non-hypothesis is shown in table **??**.

# 6  Application Examples

## 6.1  Dual Sample Pooling Test

Salmonella enteric serovar Enteritidis (SE) has emerged in the past 30 years as a leading cause of human salmonellosis in the United States [12, 13]. If SE is isolated from the environment of chicken houses, then eggs from SE-positive houses must be tested. Testing eggs for SE requires a large sample size as only a small proportion are contaminated in an infected flock. Therefore,
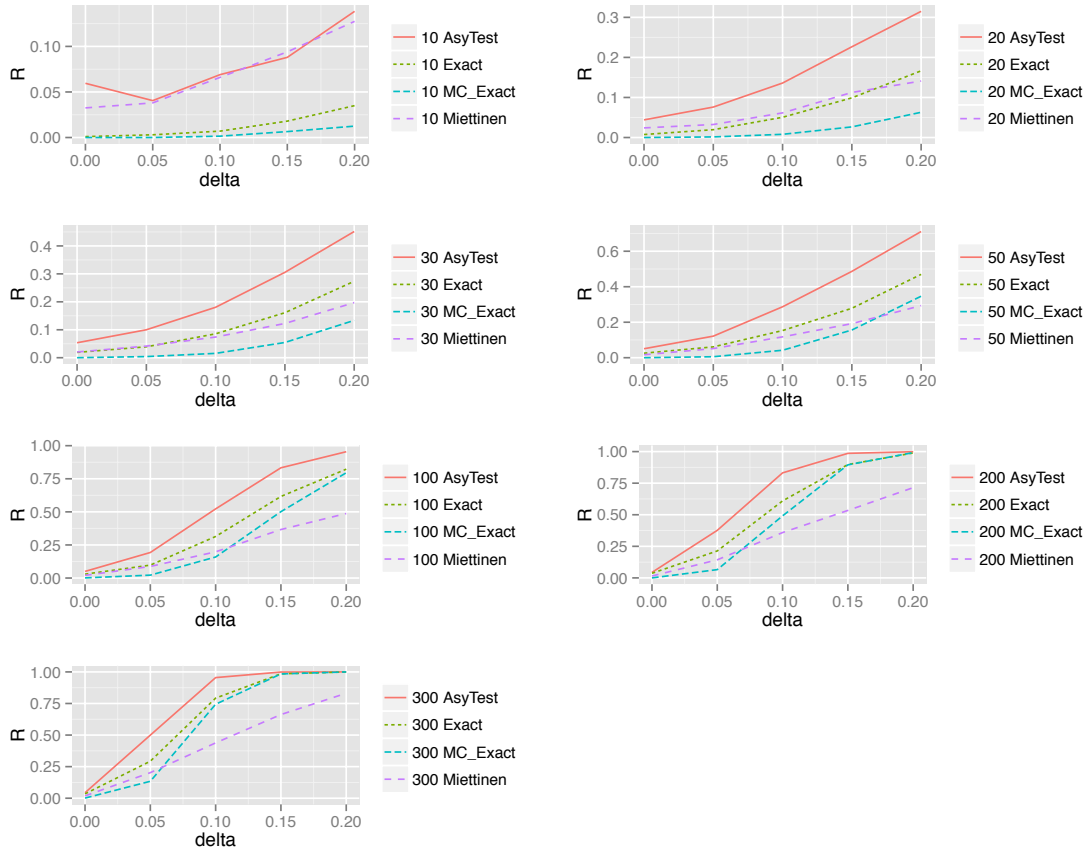
Figure 3: Comparison of exact test power and asymptotic test power for setting 3 of one-to-two case. The numbers in the legend indicate the sample size. AsyTest indicates asymptotic test; Exact indicates exact test.

environmental sampling is the primary means by which flocks are monitored for SE. Environmental (or egg) testing has traditionally been carried out using bacterial culturing which is the standard by which all other tests are compared. Bacteriological culturing typically requires 5 to 7 days before results are obtained. Real-time polymerase chain reaction (RT PCR) is one testing method that has been developed to decrease the time required for testing. The cost of testing associated with the implementation of U.S. Food and Drug Administration (FDA) 's Final Rule has placed a substaincial burden on producers. Sample pooling is one strategy to reduce costs and labor. The aim of the study is to examine the validity of an SE-specific RT PCR in pooled samples. The provisionally approved National Poultry Improvement Plan (NPIP) modified semisolid Rappaport-Vassiliadis (MSRV) method as the gold standard. RT PCR results from pool
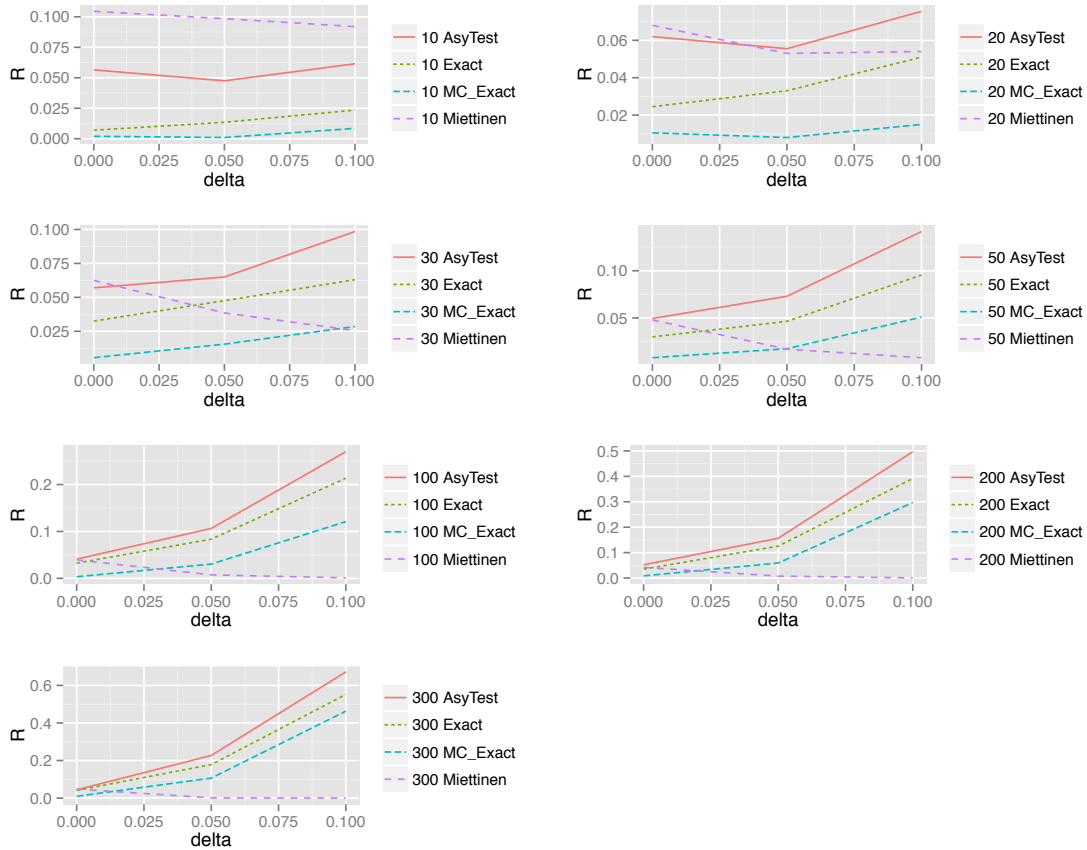
Figure 4: Comparison of exact test power and asymptotic test power for setting 4 of one-to-two case. The numbers in the legend indicate the sample size. AsyTest indicates asymptotic test; Exact indicates exact test.

sizes of two were compared with single sample testing. A total of 208 environmental field samples were collected from three commercial layer houses on the same site. Houses were previously found to be positive for SE by culture at the ISU VDL. Each house contained twelve rows of cages with three tiers of cages within each row. Flocks within each house consisted of adult laying hens. Gauze drag swabs pre-soaked with sterilized milk mile were used to sample egg belt sections from each tier of cages within each row and from fecal material on support beams directly under the cage section sampled. Samples were taken every fifty feet along the length of the house. Swabs were put into Whirl-Pak bags and transported on ice to the Iowa State University Veterinary Diagnostic Laboratory for testing. After incubation, 1 ml aliquots were removed from the enrichment broth of field environmental samples for RT PCR analysis. Sets of pooled samples

were prepared from these aliquots so that each individual sample was represented once and randomly assigned to a pooled set of 2 samples (208 individual, 104 pools of two). In this example, the pooling test is test 1 and the single test is test 2 with n=104. The counting results are presented in Table 7 .

Table 3: Counting Table for Dual Pooling Test

|  |  | Test 2 | | | |
|---|---|---|---|---|---|
|  |  | 2 | 1 | 0 | Total |
| Test 1 | 1 | 0 | 7 | 0 | 7 |
|  | 0 | 0 | 0 | 97 | 97 |
|  | Total | 0 | 7 | 97 | 104 |

A Fisher's exact test for independence results in a p-value of $4.707 \times 10^{-11}$, indicating convincing evidence of dependency between tests in the table. Thus Miettinen's test should not be applied in this situation because it is derived under the independence assumption. The probability that the fuzzy p-value is less than 0.05 is only 0.06. The median fuzzy p-value is 0.25 and the 95% quantile is 1. The result indicates no evidence against $H_0$.
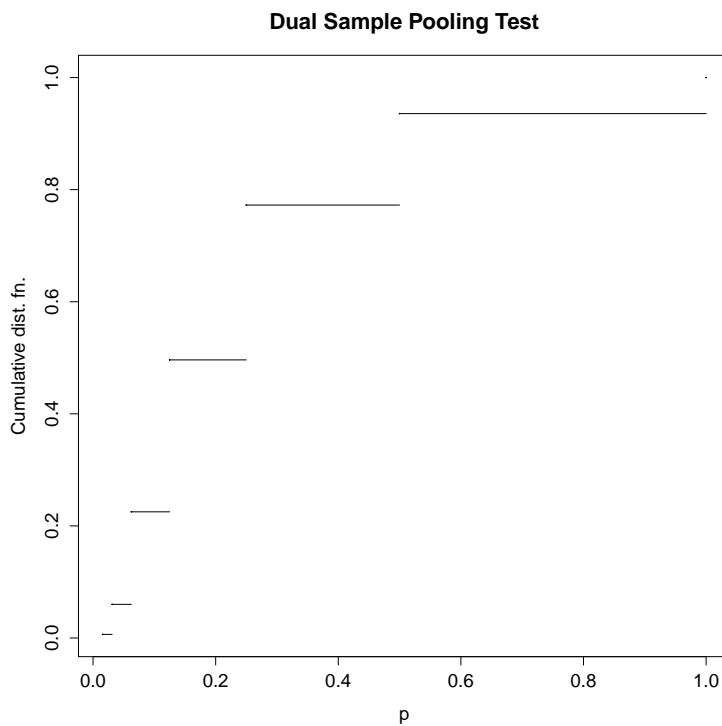


Figure 5: Cumulative distribution functions of the fuzzy p-values for Dual Sample Pooling Test based on 2000 iterations

11

## 6.2 Pen-based oral fluid specimens for influenza A virus detection

Christa K. Goodell et al. used a matched design in their influenza A virus (IVA) monitoring study. For IAV detection, the traditional ante mortem Nasal Swabs (NS) specimen is difficult and expensive to get because it is necessary to select, restrain, and swab individual pigs. Alternatively, oral fluids (OF), a specimen new to swine diagnostics but well-characterized in human diagnostics, is easy to collect because pigs naturally investigate their environment by chewing. The question is to compare the probability of detecting IAV in OF and NS specimens collected from vaccinated pigs. IAV vaccinated pigs were inoculated with subtypes H1N1 or H3N2. Pen-based oral fluid samples were collected day post inoculation. There were two pigs in each pen. The OF and NS samples were tested in the laboratory with results to be either negative or positive. Each OF sample from one pen matches with two NS samples from two individual pigs in the same pen. The data are as follows:

Table 4: Counting Table for influenza A virus detection

|  |  | NS | | | |
|---|---|---|---|---|---|
|  |  | 2 | 1 | 0 | Total |
| OF | 1 | 114 | 28 | 29 | 171 |
|  | 0 | 2 | 7 | 42 | 51 |
|  | Total | 116 | 35 | 81 | 222 |

A Fisher's exact test for independence results in a p-value $< 2.2 \times 10^{-16}$, indicating convincing evidence of dependency between tests in the table. Thus Miettinen's test should not be applied in this situation because it is derived under the independence assumption. The whole distribution of the fuzzy p-value is concentrated below 0.05. The asymptotic test p-value is $2.71 \times 10^{-9}$. There is very strong evidence for difference between positive rates of the two tests. OF is better than SN in terms of both convenience and sensitivity. This example is not a pooling test as the first example, however, the data also has a matched scheme.

# 7 Discussion and Conclusion

The simulation results of above work show that Miettinen's test performs poorly when the multiple observations from the same matched set are dependent. Except for very small numbers of matched sets, in general, the results support that both exact and asymptotic test have good power and control type one error well. Asymptotic test out performs the exact test by effectiveness and computational speed. The estimated power for the asymptotic test based on 2000 simulated data sets is very close to the calculated results from the power function. The tests proposed in the present work have rather wide applicability in medical and other research. Both the exact and the asymptotic versions of our proposed statistical tests can be generalized from 1-to-2 to 1-to-N matched data. A related question arise: does the exact and asymptotic test remain accurate for $N > 2$? It is a question worthy of future investigation.
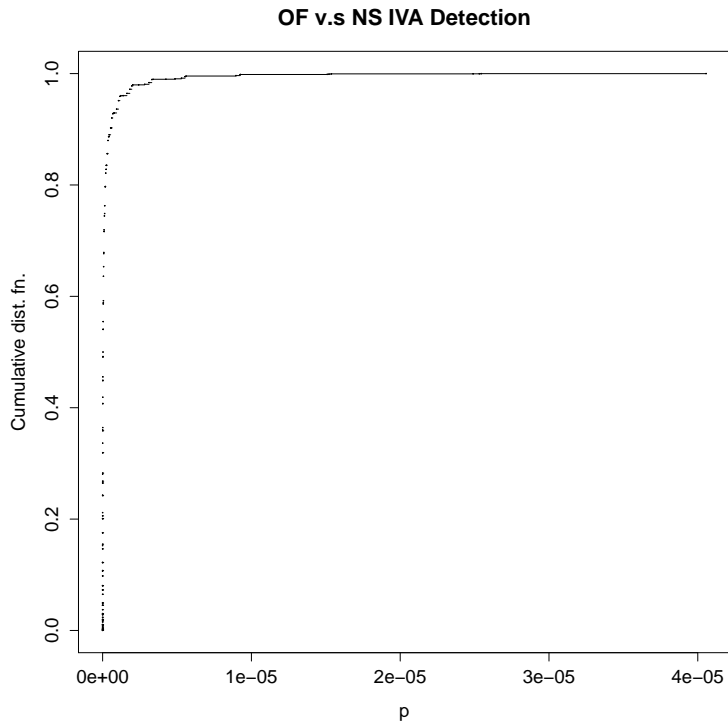
**OF v.s NS IVA Detection**



Figure 6: Cumulative distribution functions of the fuzzy p-values for Pen-based oral fluid specimens for influenza A virus detection based on 2000 iterations

# References

[1] Dorfman, R. 1943, The Detection of Defective Members of large Populations , Annals of Mathematical Statistics. 14, 436-440

[2] Eugene Litvak, Xin M. Tu, Marcello Pagano, 1994, Screening for the Presence of a Disease by Pooling Sera Samples , Journal of the American Statistical Association, Vol. 89, No. 426, pp. 424-434

[3] Tu XM, Litvak E, Pagano M, 1994, Studies of AIDS and HIV surveillance. Screening tests: can we get more by doing less? ,Stat Med. 1994 Oct 15-30;13(19-20):1905-19

[4] STEFANOS A. ZENIOS AND LAWRENCE M. WEIN , 1998, pooled testing for hiv prevalence estimation: exploiting the dilution effect Statist. Med. 17, 1447-1467

[5] Wesley O. Johnsona; Joseph L. Gastwirth , 2000, Dual group screening ,Journal of Statistical Planning and Inference 83 pp. 449-473

[6] S. Vansteelandt, E. Goetghebeur, T. Verstraeten, 2000, Regression Models for Disease Prevalence with Diagnostic Tests on Pools of Serum Samples ,Biometrics, vol 56, No. 4,pp. 1126-1133

13

[7] McNemar, Q. 1947, Note on the sampling error of the differences between correlated proportions of percentages, Psychometrika 12, 153-157

[8] Bennett, B. M. and Underwood, R. E. (1970). On McNemar's test for the 2$\times$2 table and its power function. Biometrics 26, 339-343.

[9] Olli S. Miettinen, 1969, Individual Matching with Multiple Controls in the Case of All-or-None Responses, Biometrics, Vol. 25, No. 2, pp. 339-355

[10] Stephen W. Duffy, 1984, Asymptotic and Exact Power for the McNemar Test and Its Analogue with R Controls Per Case, Biometrics, Vol. 40, No. 4, pp. 1005-1015

[11] Charles J, Geyer and Glen D. Meeden, Fuzzy and Randomized Confidence Intervals and P-values, . Statistical Science, Vol. 20, No. 4, 358-366. 2005

[12] Angulo, F.J. and D.L. Swerdlow. Epidemiology of human Salmonella enteric server Enteritidis in the United States. Iowa State University Press, Ames, 1999

[13] Patrick, M.E., P.M. Adcock, T.M. Gomez, S.F.Altekruse, B.H.Holland, R.V. Tauxe and D.L. Swerdlow. Salmonella enteritidis infections, United States, 1985-1999. Emerging Infectious Diseases 10:1-7. 2004