

Article

Statistical Tests for Proportion Difference in One-to-Two Matched Binary Diagnostic Data: Application to Environmental Testing of Salmonella in the United States

Hui Lin ¹, Adam Zhu ^{1,2} and Chong Wang ^{1,*} 

¹ Department of Statistics, Iowa State University, Ames, IA 50011, USA; hlin@iastate.edu (H.L.); aczhu@iastate.edu (A.Z.)

² Ames High School, Ames, IA 50010, USA

* Correspondence: chwang@iastate.edu

Abstract: Pooled sample testing is an effective strategy to reduce the cost of disease surveillance in human and animal medicine. Testing pooled samples commonly produces matched observations with dichotomous responses in medical and epidemiological research. Although standard approaches exist for one-to-one paired binary data analyses, there is not much work on 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 methods are markedly different from the previous studies in that we do not rely on the mutual independence assumption. The emphasis on the interdependence of observations within a matched set is inherent and attractive in both human health and veterinary medicine. It can be applied to all kinds of diagnostic studies with a one-to-two matched data structure. Our methods can be generalized to the one-to-N matched case. We discuss applications of the proposed methods to the environmental testing of salmonella in the United States.

Keywords: disease surveillance; production animals; exact test; asymptotic test; matched binary data; diagnostic testing; sample pooling

MSC: 62F03



Citation: Lin, H.; Zhu, A.; Wang, C. Statistical Tests for Proportion Difference in One-to-Two Matched Binary Diagnostic Data: Application to Environmental Testing of Salmonella in the United States. *Mathematics* **2024**, *12*, 741. <https://doi.org/10.3390/math12050741>

Academic Editor: Yuhlong Lio

Received: 13 January 2024

Revised: 25 February 2024

Accepted: 29 February 2024

Published: 1 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pooled sample testing is an effective strategy to reduce the cost of disease surveillance in human and animal medicine. Testing pooled samples commonly produces matched observations with dichotomous responses in medical and epidemiological research. Although standard approaches exist for one-to-one paired binary data analyses, there is not much work on one-to-two or one-to-N matched binary data in the current statistical literature.

Our research was originally motivated by the pooling of diagnostic tests. Often, testing the units one-by-one is inefficient, especially when the prevalence is sufficiently small. The concept of screening pooled samples originated during the second world war to detect syphilis among US soldiers [1]. Since then, it has aroused significant attention and has been used successfully in various applications. Many studies have demonstrated the successful use of the pooling strategy in HIV testing [2–5]. However, budget reduction is an important issue that limits the number of tests, causing the derived estimates to be imprecise. One way to overcome the budget limitation and improve the accuracy of the estimates is pooled testing. Vansteelandt et al. [6] showed that a good design can severely reduce cost. An instance of practical applications in Vansteelandt et al. [6] demonstrated that using data from a study on sample pools with an average size of seven reduced cost

to 44% of the original price with virtually no loss in precision. In some circumstances, the advantages of pooling go beyond reducing the cost but also earning more accuracy [2].

In the case of one-to-one matching, McNemar in 1947 [7] developed a test of marginal homogeneity in a 2×2 table that is applicable to pair-matched observations or a cohort measured twice on a variable with a binary outcome. Bennett and Underwood (1970) conducted a comparison of exact power with the non-central Chi-square approximation for sample sizes of 10, 20, and 40 and found the approximation to be adequate [8]. Miettinen [9] derived the asymptotic power for testing the difference between cases and controls with a dichotomous response in the case of one-to-one and one-to-R matching. Stephen derived the exact power based on Miettinen's work.

Using Miettinen's work as a basis, Stephen (1984) calculated the exact power of the test [10]. Additionally, Stephen compared the asymptotic power and exact power of Miettinen's test. However, Miettinen's test assumes that the multiple observations from the same matched set are mutually independent. It is clear that this assumption does not hold for pooling test data because the pooled sample is dependent on the individual samples being pooled.

In this paper, we propose exact and asymptotic tests for one-to-two matched binary data. Our methods fit more realistic situations without assuming that observations from the same subject are mutually independent. This method is suitable for various types of diagnostic studies that use a one-to-two matched data structure, including but not limited to dual-sample pooling, one-to-two case-control studies, and so on. Our methods can be generalized to one-to-N matched cases. For clarity of presentation, we explain basic concepts, terminologies, and notations in Section 2. We illustrate the exact test procedure and asymptotic test procedure in Sections 3 and 4. In Section 5, we show the merits of our tests by presenting a simulation study we conducted, where our tests outperformed Miettinen's test. In Section 6, we present the application of our methods to two practical situations that fail to obtain the independence required by Miettinen's test. The results are presented in Section 6, followed by a discussion in Section 7.

2. Basic Concepts, Terminologies, and Notations

2.1. Joint Counting Table for Two Diagnostic Testing Strategies

Consider a research scenario involving n subjects undergoing two different diagnostic testing strategies. In the "one-to-two" scheme, each subject yields one binary observation under diagnostic testing Strategy 1 and two binary observations under diagnostic testing Strategy 2. We denote the set of three observed values from subject $j, j = 1, 2, \dots, n$ as

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

where Y_{1j} is the observed value under Strategy 1, and Y_{2j1} and Y_{2j2} are from Strategy 2. This set of three values are matched by subject $j, j = 1, 2, \dots, n$. In this paper, uppercase letters represent random variables, while lowercase letters denote their corresponding realizations. For example, for the j -th subject, $(y_{1j}, y_{2j1}, y_{2j2})$ represents a realization obtained for the random response vector $(Y_{1j}, Y_{2j1}, Y_{2j2})$. The value of the binary response variables is either 0 or 1. Denote the probability that a random observation under diagnostic testing Strategy 1 takes value 1 to be p_1 and the probability that a random observation under diagnostic testing Strategy 2 takes value 1 to be p_2 , i.e., $p_1 = Pr\{Y_{1j} = 1\}$ and $p_2 = Pr\{Y_{2j1} = 1\} = Pr\{Y_{2j2} = 1\}, j = 1, 2, \dots, n$. The parameter of interest for the statistical inference is

$$\delta = p_1 - p_2.$$

For a hypothesis testing question, a null hypothesis of interest is

$$H_0 : \delta = p_1 - p_2 = 0,$$

that is, there is no difference between the two diagnostic testing strategies in the probability of observing value 1.

Now, consider the number of value 1s observed for subject j , (X_{1j}, X_{2j}) , where $X_{1j} = Y_{1j}$ and $X_{2j} = Y_{2j1} + Y_{2j2}$, multinomial distributions of the response vector, for diagnostic testing Strategy 1 and Strategy 2, respectively. The variable X_{1j} can take the value 0 or 1 and X_{2j} can take the value 0,1, or 2. Usually, the data from such an experiment over n subjects are summarized into a counting table, such as Table 1, where $Z_{kl}^{(j)}, k = 0, 1, l = 0, 1, 2$ are the number of subjects, i.e., counts, for the combinations of $(X_{1j}, X_{2j}), j = 1, 2, \dots, n$ diagnostic testing observations. There are six possible realizations for the realization of (X_{1j}, X_{2j}) for each $j = 1, 2, \dots, n$. Denote $Z_{kl}^{(j)} = I(X_{1j} = k, X_{2j} = l)$ with $k = 0, 1$ and $l = 0, 1, 2$ for subject j . The counts over all n subjects falling in the table's cells, defined as $Z_{kl} = \sum_{j=1}^n Z_{kl}^{(j)}$, are multinomially distributed with

$$(Z_{00}, Z_{01}, Z_{02}, Z_{10}, Z_{11}, Z_{12}) \sim \text{multinomial}(n, p_{00}, p_{01}, p_{02}, p_{10}, p_{11}, p_{12}).$$

Here, the probabilities $p_{kl} = P(Z_{kl}^{(j)} = 1, k = 0, 1, l = 0, 1, 2)$ do not depend on subject $j, j = 1, 2, \dots, n$. The joint counting table for two diagnostic testing strategies over n subjects is illustrated in Table 1.

Table 1. Joint counting table for two diagnostic testing strategies over n subjects.

		Diagnostic Testing Strategy 2			Total
		2	1	0	
Strategy 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

2.2. Miettinen's Exact Test

Miettinen [9] 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;
2. Y_{1j}, Y_{2j1}, Y_{2j2} are mutually independent conditionally on (p_1, p_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 ,

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

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

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

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

In situations where the outcomes of diagnostic testing Strategy 1 and Strategy 2 are biologically dependent, such as in the context of a pooling test scenario, the assumption of independence between Test 1 and Test 2 becomes untenable. Traditional paired test analysis methodologies, exemplified by McNemar's test, typically do not necessitate the assumption of independence between paired test results. In the subsequent sections, we present novel statistical tests that do not rely on the assumption of independence.

3. Randomized Exact Test

In this section, we propose a randomized exact test for one-to-two matched binary diagnostic data without assuming independence between the diagnostic testing strategies from the same subject.

3.1. Test Statistic

Following the model structure described in Section 2.1, for any j ,

$$p_1 = Pr\{Y_{1j} = 1\} = P(Z_{12}^{(j)} = 1 \text{ or } Z_{11}^{(j)} = 1 \text{ or } Z_{10}^{(j)} = 1) = p_{12} + p_{11} + p_{10},$$

and

$$p_2 = Pr\{Y_{2j1} = 1\} = P(Z_{12}^{(j)} = 1 \text{ or } Z_{02}^{(j)} = 1) + \frac{1}{2}P(Z_{11}^{(j)} = 1 \text{ or } Z_{01}^{(j)} = 1) \\ = p_{12} + p_{02} + \frac{1}{2}p_{11} + \frac{1}{2}p_{01}.$$

Thus, the null hypothesis of interest $H_0 : \delta = p_1 - p_2 = (p_{12} + p_{11} + p_{10}) - (p_{12} + p_{02} + \frac{1}{2}p_{11} + \frac{1}{2}p_{01}) = p_{10} - \frac{1}{2}p_{01} + \frac{1}{2}p_{11} - p_{02} = 0$ is equivalent to $H_0: p_{10} + \frac{1}{2}p_{11} = p_{02} + \frac{1}{2}p_{01}$.

We propose a randomized test by considering variables $R_{kl}^{(j)} | Z_{kl}^{(j)} \sim Bin(Z_{kl}^{(j)}, \frac{1}{2})$ and $R_{kl} = \sum_{j=1}^n R_{kl}^{(j)}$. The marginal probability of $R_{kl}^{(j)}$ is $Pr\{R_{kl}^{(j)} = 1\} = Pr\{R_{kl}^{(j)} = 1 | Z_{kl}^{(j)} = 1\}Pr\{Z_{kl}^{(j)} = 1\} = \frac{p_{kl}}{2}$. So, for $k \neq k'$ or $l \neq l'$, the distribution of $Z_{k'l'}^{(j)} + R_{kl}^{(j)}$ is

$$Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 2\} = Pr\{Z_{k'l'}^{(j)} = 1, R_{kl}^{(j)} = 1\} = 0; \\ Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 1\} = Pr\{Z_{k'l'}^{(j)} = 1, R_{kl}^{(j)} = 0\} + Pr\{Z_{k'l'}^{(j)} = 0, R_{kl}^{(j)} = 1\} \\ = Pr\{Z_{k'l'}^{(j)} = 1\}Pr\{R_{kl}^{(j)} = 0 | Z_{k'l'}^{(j)} = 1\} + Pr\{R_{kl}^{(j)} = 1\}Pr\{Z_{k'l'}^{(j)} = 0 | R_{kl}^{(j)} = 1\} \\ = Pr\{Z_{k'l'}^{(j)} = 1\} \cdot 1 + Pr\{R_{kl}^{(j)} = 1\} \cdot 1 = p_{k'l'} + \frac{p_{kl}}{2}; \\ Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 0\} = 1 - Pr\{Z_{k'l'}^{(j)} + R_{kl}^{(j)} = 1\} = 1 - (p_{k'l'} + \frac{p_{kl}}{2}).$$

Then, we have $\sum_{j=1}^n (Z_{k'l'}^{(j)} + R_{kl}^{(j)}) = Z_{k'l'} + R_{kl} \sim Bin(n, p_{k'l'} + \frac{p_{kl}}{2})$. 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 $Z_{10} + R_{11} | S \sim Bin(S, \frac{1}{2})$. A two-sided Randomized Exact Test can be performed through the following three steps:

1. Randomly sample $r_{11} | z_{11} \sim Bin(z_{11}, \frac{1}{2})$ and $r_{01} | z_{01} \sim Bin(z_{01}, \frac{1}{2})$.
2. Calculate $s_1 = \max(z_{10} + r_{11}, z_{02} + r_{01})$, $s_2 = \min(z_{10} + r_{11}, z_{02} + r_{01})$, and $s = z_{10} + r_{11} + z_{02} + r_{01}$.
3. Calculate the p -value as $Pr\{x \leq s_2 \text{ or } x \geq s_1\}$ with $x \sim Bin(s, \frac{1}{2})$.

Due to the randomization of r_{11} in the randomized test, the procedure can give different answers for the same data. The arbitrariness of randomization can be avoided while keeping the beautiful theory of these procedures through a simple change of viewpoints to what is called "fuzzy p -value", advanced by Geyer and Meeden (2005) [11]. In contrast to traditional p -values, fuzzy p -values are random variables interpreted as p -values. In terms of the randomized exact test illustrated above, r_{11} is called a latent variable; hence, the p -value calculated in step 3 is named the latent p -value. The latent p -value would be the p -value if the latent variable was observed. The exact test employing the notion of a fuzzy p -value uses simulations of the latent variable under the null hypothesis. It provides an expression of both the strength and the uncertainty of the evidence against the null hypothesis. In practice, the randomized exact test procedure described above is repeated a large number of times to obtain an empirical distribution of the fuzzy p -value for inference.

3.2. Power of Randomized Exact Test

In this section, we derive the power of the randomized exact test as a function of $\delta = p_1 - p_2$, p_2 , p_{12} , and p_{11} . Following the model structure described in Section 2.1, we first express the parameters p_{01} , p_{02} , p_{10} , and p_{00} as functions of δ , p_1 , p_{12} , and p_{11} ,

$$\begin{aligned}
 p_{01} &= 2p_2(1 - p_2) - p_{11}; \\
 p_{02} &= p_2^2 - p_{12}; \\
 p_{10} &= p_2 + \delta - p_{12} - p_{11}; \\
 p_{00} &= (1 - p_2)^2 - (p_2 + \delta - p_{12} - p_{11}).
 \end{aligned}$$

We have shown in Section 3 that $Z_{10} + R_{11} \sim \text{Bin}(n, p_{10} + \frac{p_{11}}{2})$ and $Z_{02} + R_{01} \sim \text{Bin}(n, p_{02} + \frac{p_{01}}{2})$. Furthermore, note that

$$\begin{aligned}
 p_s &= p_{10} + \frac{p_{11}}{2} + p_{02} + \frac{p_{01}}{2} = 2p_2 + \delta - p_{11} - 2p_{12}; \\
 \frac{p_{10} + \frac{p_{11}}{2}}{p_{10} + \frac{p_{11}}{2} + p_{02} + \frac{p_{01}}{2}} &= \frac{p_2 - p_{12} - \frac{p_{11}}{2} + \delta}{2p_2 + \delta - p_{11} - 2p_{12}} \equiv q.
 \end{aligned} \tag{1}$$

Then, $S \sim \text{Bin}(N, p_s)$ and $Z_{10} + R_{11} \mid S \sim \text{Bin}(S, q)$. The unconditional power can be obtained as the expectation of the conditional power. We derive the power expression of the exact binomial test as a function of q defined in expression 1 as

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

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

4. Asymptotic Test

In this section, we propose an asymptotic version of the test which avoids the randomness in the randomized exact test proposed above. Let us denote $T^{(j)} = (Z_{10}^{(j)} + \frac{Z_{11}^{(j)}}{2}) - (Z_{02}^{(j)} + \frac{Z_{01}^{(j)}}{2})$. Then

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

Since at most one of $\{Z_{10}^{(j)}, Z_{11}^{(j)}, Z_{02}^{(j)}, Z_{01}^{(j)}\}$ is non-zero, we have $Z_{kl}^{(j)} Z_{k'l'}^{(j)} = 0$ if $k \neq k'$ or $l \neq l'$. Furthermore, we have $Z_{kl}^{(j)2} = Z_{kl}^{(j)}$. Then, we get the variance expression:

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

Since observations from different subjects are independent, we have

$$\begin{aligned}
 \mu &= E \sum_{j=1}^n T^{(j)} = n\delta, \\
 \sigma^2 &= Var \sum_{j=1}^n T^{(j)} = n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2).
 \end{aligned}$$

By the central limit theorem (CLT), under the null hypothesis H_0 , $\frac{\sum_{j=1}^n T^{(j)}}{\sqrt{\text{Var}(\sum_{j=1}^n T^{(j)})}} = \frac{(Z_{10} + \frac{Z_{11}}{2}) - (Z_{02} + \frac{Z_{01}}{2})}{\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4})}}$ is asymptotically standard normal when n is large. The asymptotic test is to compare the following test statistics to a standard normal distribution $N(0, 1)$:

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

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 asymptotically standard normal. The power with respect to the effect size δ is a function of the mean and variance of the test statistic

$$\beta = 2\Phi\left(\frac{\phi_{\alpha/2} \sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4})} - n\delta}{\sqrt{n(p_{10} + \frac{p_{11}}{4} + p_{02} + \frac{p_{01}}{4} - \delta^2)}}\right), \tag{3}$$

where $\phi_{\alpha/2}$ is the $\alpha/2$ lower quantile of the standard normal distribution and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

5. Simulation

In this section, we present the simulation studies we conducted to validate our proposed methods and compare them to Miettinen’s exact test.

5.1. The Simulation Setting

The simulation results were based on $n = 10, 20, 30, 50, 100, 200, 300$. We set various values for p_1, p_{12} , and p_{11} and set δ as

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

We study the following four different settings:

Setting 1 : $p_1 = 0.3, p_{12} = 0.01, p_{11} = 0.01;$

Setting 2 : $p_1 = 0.3, p_{12} = 0.08, p_{11} = 0.15;$

Setting 3 : $p_1 = 0.4, p_{12} = 0.15, p_{11} = 0.24;$

Setting 4 : $p_1 = 0.4, p_{12} = 0.11, p_{11} = 0.03.$

For each case, $M = 2000$ simulations were performed. Function “`rmultinom()`” in R was used to simulate the multinomial samples.

It is unprofitable to compute expression (2) for n with large values. Therefore, the power for each test was estimated using 2000 simulations. The powers for the exact test in expression (2) and asymptotic test in expression (3) were also calculated. Since the exact power is dependent on the individual binomial or multinomial parameters, the arbitrary choice of a further parameter is necessary. The type I error rate for the simulations was set to be 0.05.

5.2. Simulation Results

The resulting power of the exact test and the asymptotic test for each of the four settings are plotted in Figures 1–4. In the figures, the horizontal axis is for the effect size δ and the vertical axis is for R, the rejection rate. The curves represent the methods: AsyTest—the asymptotic test; Exact—the exact test; MC Exact—the exact test power using Monte Carlo simulation; and Miettinen—Miettinen’s exact test.

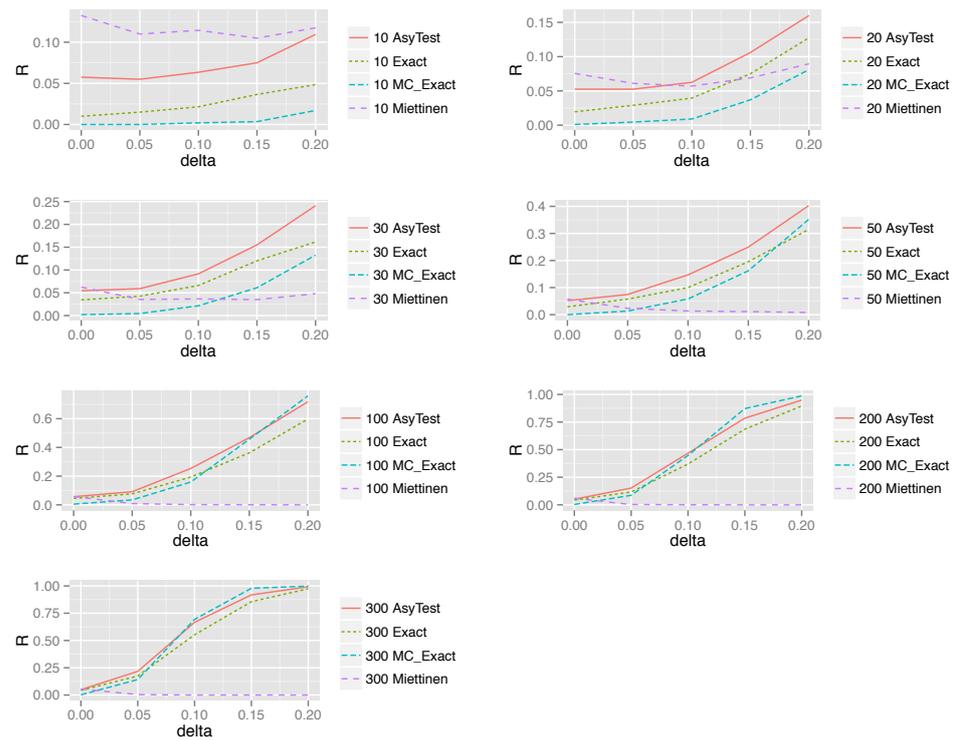


Figure 1. Comparison of exact test power and asymptotic test power for simulation setting 1 of one-to-two case. In the figure, the horizontal axis is for the effect size δ and the vertical axis is for R , the rejection rate. The curves represent the methods: AsyTest—the asymptotic test; Exact—the exact test; MC Exact—the exact test power using Monte Carlo simulation; and Miettinen—Miettinen’s exact test.

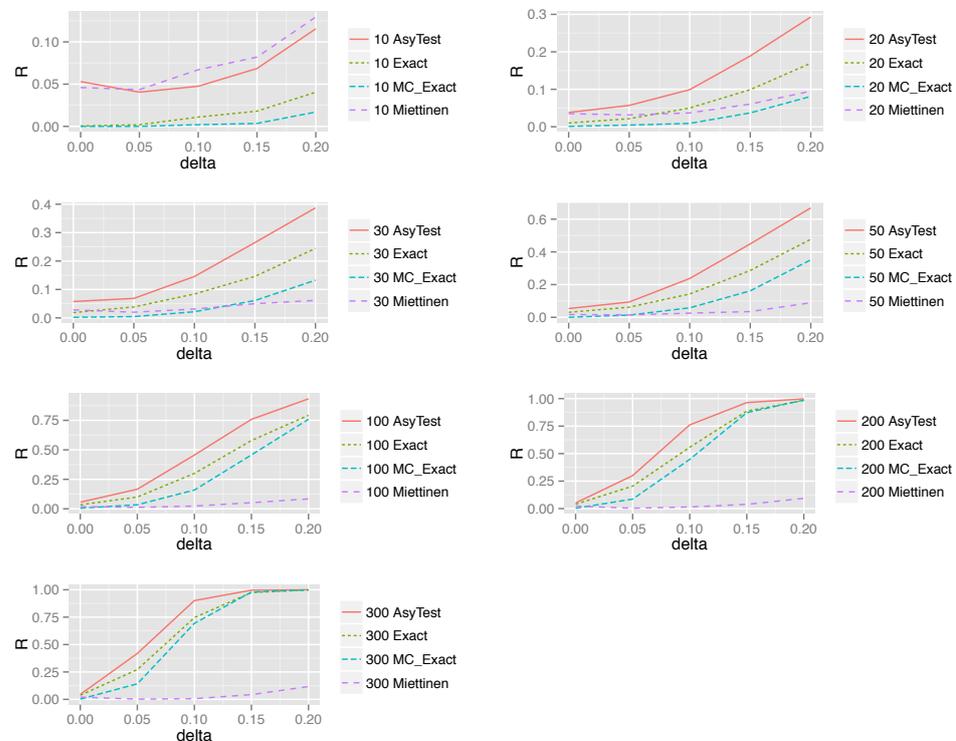


Figure 2. Comparison of exact test power and asymptotic test power for simulation setting 2 of one-to-two case. In the figure, the horizontal axis is for the effect size δ and the vertical axis is for R , the rejection rate. The curves represent the methods: AsyTest—the asymptotic test; Exact—the exact test; MC Exact—the exact test power using Monte Carlo simulation; and Miettinen—Miettinen’s exact test.

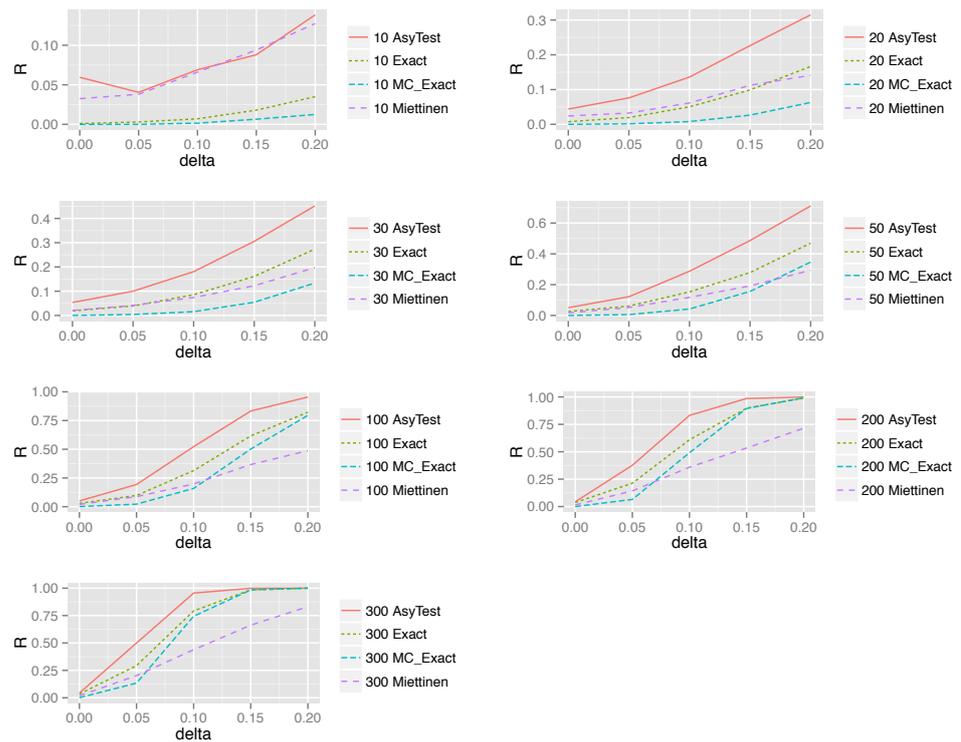


Figure 3. Comparison of exact test power and asymptotic test power for simulation setting 3 of one-to-two case. In the figure, the horizontal axis is for the effect size δ and the vertical axis is for R , the rejection rate. The curves represent the methods: AsyTest—the asymptotic test; Exact—the exact test; MC Exact—the exact test power using Monte Carlo simulation; and Miettinen—Miettinen’s exact test.

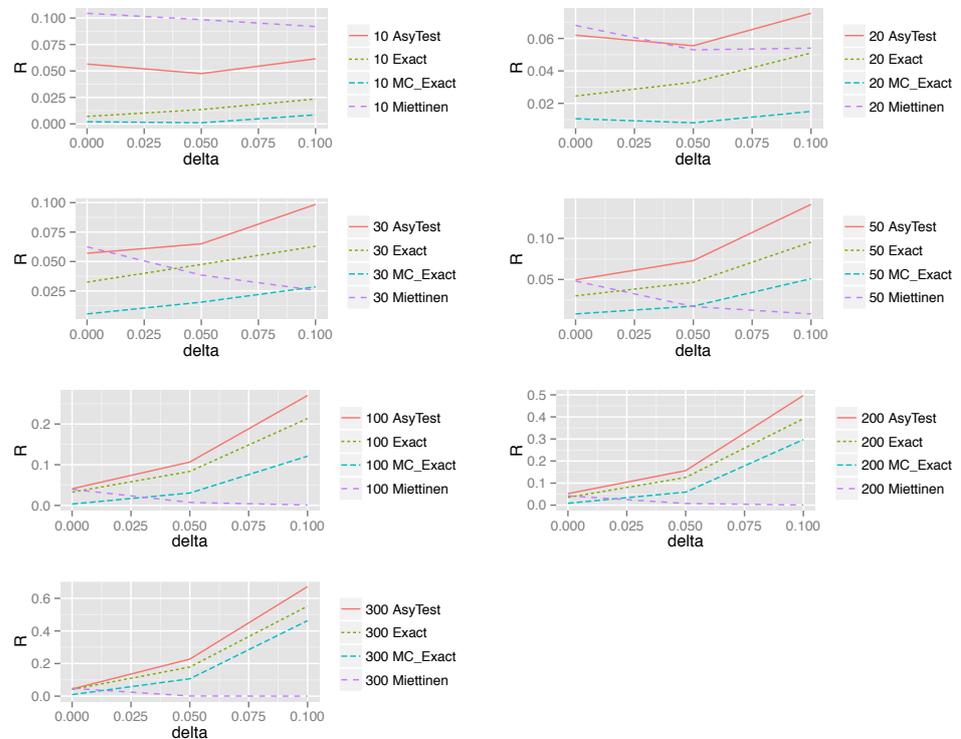


Figure 4. Comparison of exact test power and asymptotic test power for simulation setting 4 of one-to-two case. In the figure, the horizontal axis is for the effect size δ and the vertical axis is for R , the rejection rate. The curves represent the methods: AsyTest—the asymptotic test; Exact—the exact test; MC Exact—the exact test power using Monte Carlo simulation; and Miettinen—Miettinen’s exact test.

When the effect size $\delta = 0$, the null hypothesis is true and the rejection rate is the type I error of the methods being compared. When the effect size $\delta > 0$, the null hypothesis is false and the rejection rate is the power of the methods. Figures 1–4 show that our proposed methods control the type I error rates very well (around or below the desired level 0.05) and outperform Miettinen’s exact test in almost all combinations of sample size, effect size, and parameter setting. The power increases steadily as the effect size increases for the exact binomial test and asymptotic test as we would expect. Miettinen’s test performs poorly when the assumption of mutual independence does not hold, especially in terms of power. The asymptotic test consistently dominates the other for all settings.

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, environmental sampling is the primary means by which flocks are monitored for SE. Environmental (or egg) testing has traditionally been carried out using bacterial culture, which is the standard that all other tests are compared to. 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. Testing costs are high because of the implementation of the U.S. Food and Drug Administration (FDA)’s Final Rule. Sampling pooling is one strategy to reduce costs and labor. The aim of the study was 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 was used as the gold standard. RT PCR results from a pool size of two were compared with a 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 through culture at the Iowa State University (ISU) Veterinary Diagnostic Laboratory (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 skim milk, 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 in 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. A set of pooled samples were prepared from these aliquots so that each individual sample was represented once and randomly assigned to a pooled set of two samples (208 individual, 104 pools of two). In the example, the pooled test was test 1, and the single test was test 2 in our model with $n = 104$. The counting results are presented in Table 2.

Table 2. Counting table for dual pooling test.

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

Fisher’s exact test for independence resulted in a p -value of 4.707×10^{-11} , indicating convincing evidence of dependency between the two tests in the table. Thus, Miettinen’s

test should not be applied in this situation because it is derived under the independence assumption. The cumulative distribution function of the fuzzy p -values for the dual sample pooling test is shown in Figure 5. The probability that the fuzzy p -value is less than 0.05 was only 0.0613. The median fuzzy p -value was 0.25. The result indicates very weak evidence against H_0 .

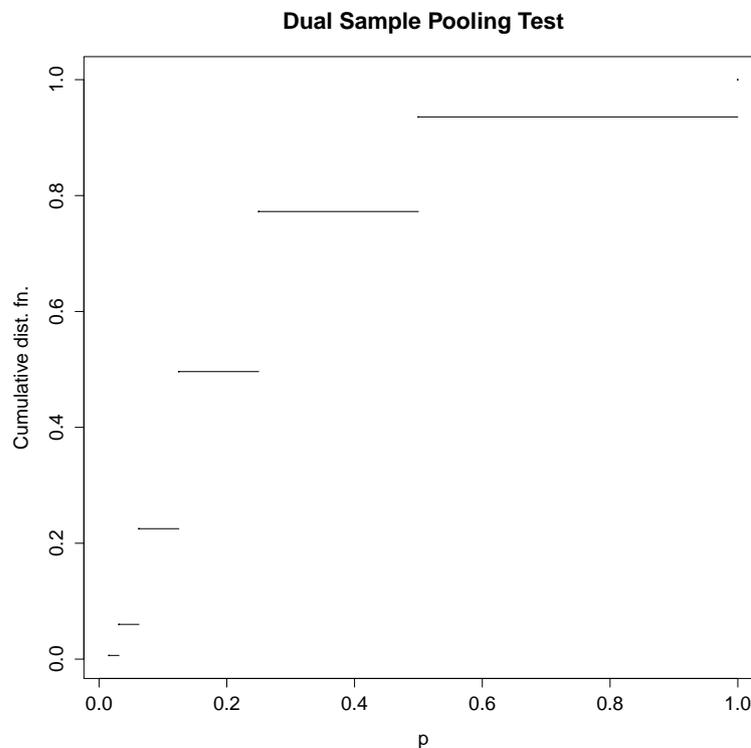


Figure 5. Cumulative distribution function of the fuzzy p -values for the dual sample pooling test.

6.2. Pen-Based Oral Fluid Specimens for Influenza A Virus Detection

Christa K. Goodell et al. used the matched design in their influenza A virus (IVA) monitoring study. For IAV detection, the traditional ante mortem specimen nasal swab (NS) is hard and expensive to get because it is necessary to select, restrain, and swab individual pigs. Alternatively, oral fluid (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 on the day post inoculation. The OF and NS samples were tested in the laboratory with results to be either negative or positive. Each OF sample from one pen was matched with two NS samples from two individual pigs in the same pen. The data are presented as follows in Table 3:

Table 3. Counting table for influenza A virus detection.

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

Fisher's exact test for independence resulted 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 due to the violation of the independence assumption. The cumulative distribution function of the fuzzy p -values for the dual sample pooling test is shown in Figure 6. The whole distribution of the fuzzy p -value is concentrated below 0.05. This provides very strong evidence that there is a difference between the positive rates of the two tests.

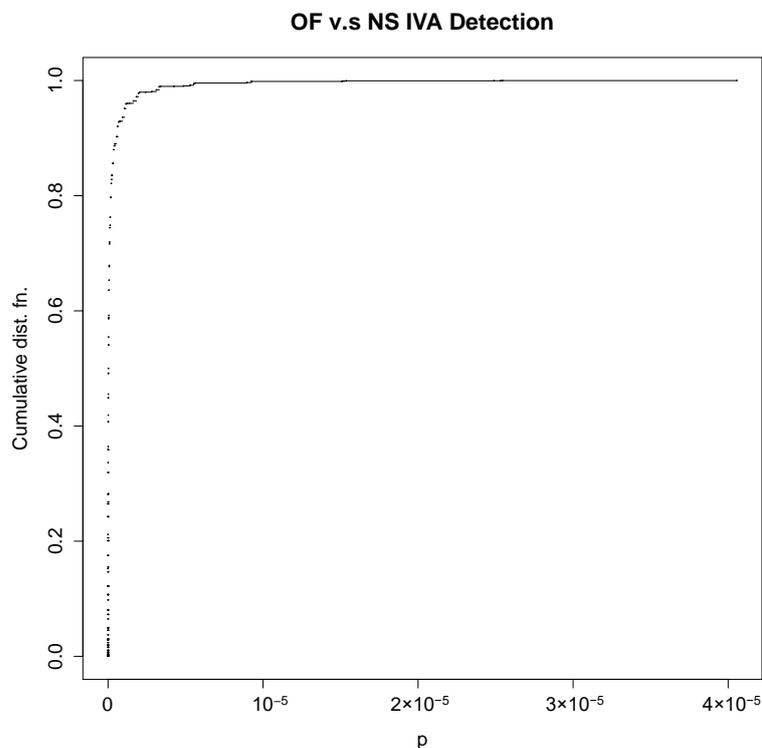


Figure 6. Cumulative distribution function of the fuzzy p -values for the pen-based oral fluid specimens for influenza A virus detection.

7. Discussion and Conclusions

Sample pooling is a very common practice in disease surveillance for both animals and human beings. In this paper, we propose novel statistical tests for proportion difference in one-to-two matched binary data. The results show that both tests we propose are valid, whereas Miettinen's test performs poorly when the multiple observations from the same matched set are dependent. The asymptotic test takes less than one millisecond on a PC equipped with an INTEL Xeon® X5482 Quad-Core Processor 3.2 GHz and 4 Gb RAM and outperforms the exact test in computational speed. The asymptotic test is also more friendly to practical users as it avoids the randomness of a randomized test. 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, e.g., oral fluid testing in animals, environmental testing of salmonella, and COVID-19 testing in humans. 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.

Author Contributions: Conceptualization, H.L. and C.W.; methodology, H.L. and C.W.; software, H.L.; validation, H.L., A.Z. and C.W.; formal analysis, H.L.; data curation, H.L.; writing—original draft preparation, H.L.; writing—review and editing, H.L., A.Z. and C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dorfman, R. The Detection of Defective Members of large Populations. *Ann. Math. Stat.* **1943**, *14*, 436–440. [[CrossRef](#)]
2. Litvak, E.; Tu, X.M.; Pagano, M. Screening for the Presence of a Disease by Pooling Sera Samples. *J. Am. Stat. Assoc.* **1994**, *89*, 424–434. [[CrossRef](#)]
3. Zenios, S.A.; Wein, L.M. Pooled testing for hiv prevalence estimation: Exploiting the dilution effect. *Stat. Med.* **1998**, *17*, 1447–1467. [[CrossRef](#)]
4. Tu, X.M.; Litvak, E.; Pagano, M. Studies of AIDS and HIV surveillance. Screening tests: Can we get more by doing less? *Stat. Med.* **1994**, *13*, 1905–1919. [[CrossRef](#)] [[PubMed](#)]
5. Johnson, W.O.; Gastwirth, J.L. Dual group screening. *J. Stat. Plan. Inference* **2000**, *83*, 449–473. [[CrossRef](#)]
6. Vansteelandt, S.; Goetghebeur, E.; Verstraeten, T. Regression Models for Disease Prevalence with Diagnostic Tests on Pools of Serum Samples. *Biometrics* **2000**, *56*, 1126–1133. [[CrossRef](#)] [[PubMed](#)]
7. McNemar, Q. Note on the sampling error of the differences between correlated proportions of percentages. *Psychometrika* **1947**, *12*, 153–157. [[CrossRef](#)] [[PubMed](#)]
8. Bennett, B.M.; Underwood, R.E. On McNemar’s test for the 2×2 table and its power function. *Biometrics* **1970**, *26*, 339–343. [[CrossRef](#)]
9. Miettinen, O.S. Individual Matching with Multiple Controls in the Case of All-or-None Responses. *Biometrics* **1969**, *25*, 339–355. [[CrossRef](#)] [[PubMed](#)]
10. Duffy, S.W. Asymptotic and Exact Power for the McNemar Test and Its Analogue with R Controls Per Case. *Biometrics* **1984**, *40*, 1005–1015. [[CrossRef](#)]
11. Geyer, C.J.; Meeden, G.D. Fuzzy and Randomized Confidence Intervals and p -values. *Stat. Sci.* **2005**, *20*, 358–366. [[CrossRef](#)]
12. Angulo, F.J.; Swerdlow, D.L. *Epidemiology of Human Salmonella Enteric Server Enteritidis in the United States*; Iowa State University Press: Ames, IA, USA, 1999.
13. Patrick, M.E.; Adcock, P.M.; Gomez, T.M.; Altekruze, S.F.; Holland, B.H.; Tauxe, R.V.; Swerdlow, D.L. Salmonella enteritidis 364 infections, United States, 1985–1999. *Emerg. Infect. Dis.* **2004**, *10*, 1–7. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.