Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v19n1p79

Improved Confidence Intervals for Mixed Paired and Unpaired Designs

By Kathman et al.

March 15, 2026

Electronic Journal of Applied Statistical Analysis
Vol. 19, Issue 01, March 2026, 79-96
DOI: 10.1285/i20705948v19n1p79

Improved Confidence Intervals for Mixed Paired and Unpaired Designs

Jacob Kathman*[a], William Staab[b], Matt Jester[c], and Scott Richter[d]

[a] *Wake Forest University, Department of Mathematics, 127 Manchester Hall, PO Box 7388, Winston-Salem, NC 27109*
[b] *University of North Carolina at Wilmington, Department of Mathematics, Sartarelli Hall, Office 2007G, 601 South College Road, Wilmington, NC 28403-5970*
[c, d] *University of North Carolina at Greensboro, Department of Mathematics and Statistics, 116 Petty Building, PO Box 26170, Greensboro, NC 27402-6170*

March 15, 2026

Several statistics are considered for creating confidence intervals for the difference of two proportions in the mixed paired and unpaired design. Five new estimators are introduced and compared with the estimators proposed by Derrick (2015). An interval incorporating the adjustment of Agresti and Caffo (2000) is the recommended estimator for comparing proportions in the mixed paired and unpaired design.

keywords: mixed paired/unpaired design, proportion difference, confidence interval.

1 Introduction

There are many scenarios in which one has a combination of paired and unpaired data. Generally, this type of design arises when one wants to do a paired test, but there are some subjects who do not have suitable pairs. It is beneficial to use all of the data for a more complete study, especially when there are a limited number of subjects that meet the study's criteria. Rather than discarding some of the data, unpaired subjects can be randomly assigned to one of the treatments to create a mixed paired/unpaired two-sample design.

For example, a previous clinical trial involved the effect on visual acuity score of two different types of laser eye surgery Dubnicka et al. (2002). Patients who had both eyes

*Corresponding author: sjricht2@uncg.edu.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

eligible for treatment had one eye randomly assigned to each method in a pair-matched design. Patients with only one eligible eye were randomly assigned to one of the methods in a completely randomized design.

A second example from Derrick et al. (2015) used data collected from a Seasonal Affective Disorder (SAD) support group to see if the quality of life of people with this disorder was different at two different times in the year. Subjects were asked to respond with a binary 'yes' or 'no' to whether they were happy with their life at that time or not. Although membership in the group was fairly consistent between the two times, there was also some natural turnover. This left the researchers with some unpaired data in which people were only present for the first or last measurement.

In the case of a quantitative response, several different methods have been proposed for making use of all the data in a mixed paired and unpaired design. Bhoj (1978, 1984, 1989) investigated a statistic that was a linear combination of paired and unpaired t statistics, $t_{combined} = \lambda t_{paired} + (1 - \lambda)t_{unpaired}$ for testing the null hypothesis that the difference of means is zero, and compared to previously proposed statistics. They found that there was no single statistic that was optimal for all cases. However, the Z_b statistic in Bhoj (1989) was the recommendation for the homoscedastic case when nothing is known about the correlation, ρ.

All of the statistics considered by Bhoj (1978, 1984, 1989) were parametric statistics. Dubnicka et al. (2002) proposed weighted and unweighted combinations of the Wilcoxon signed-rank statistic and the Wilcoxon-Mann-Whitney statistic $W_{combined} = \lambda W_{paired} + (1 - \lambda)W_{unpaired}$. They suggested these rank based tests should be more robust than statistics based on means. Magel and Fu (2014) suggested first standardizing W_{paired} and $W_{unpaired}$ and then combining the statistics, which showed slightly higher power in some cases.

Johnson and Richter (2022) proposed permutation versions of the Dubnicka et al. (2002) and Magel and Fu (2014) tests. They performed an extensive simulation study to compare various tests. No single test was always the best. The parametric methods such as the Z_b statistic in Bhoj (1989) were generally best for normally distributed data, while the non-parametric tests were better for non-normal data. The rank-based methods were generally recommended as the default as they were most powerful for non-normal distributions and were nearly as powerful as parametric methods for normal distributions.

While extensive work has been done on methods for comparing means, there is less literature on tests for comparing two sample proportions of a dichotomous dependent variable. Choi and Stablein (1982) proposed several statistics for comparing proportions in the mixed paired and unpaired design and performed a simulation study to compare the statistics. They ultimately recommended a statistic that used an estimator that combined the observations from the paired and unpaired samples. Tang and Tang (2004) developed exact tests for comparing two paired proportions with incomplete data, extending McNemar-type procedures to situations in which some pairs are missing observations. Although their framework focused on incomplete pairing rather than an explicit mixture of paired and unpaired subjects, their methodological goal was to make use of all available data. Their exact inference approach provides a useful complement to the approximate methods of Choi and Stablein (1982) and informs later developments such as Derrick et al. (2015).

Derrick et al. (2015) proposed combined test statistics and related interval estimators for the mixed paired and unpaired design. These methods used estimators similar to Choi and Stablein (1982), which combined observations from the paired and unpaired samples. Derrick et al. (2015) found that their z_6 statistic, which used individual sample standard deviations to estimate the standard error, had inflated Type 1 error rates, which would correspond to undercoverage in the confidence interval estimator.

When considering confidence intervals for a difference in proportions, $\pi_1 - \pi_2$, Agresti and Caffo (2000) noted that the Wald confidence interval,

$$(p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, \tag{1}$$

where $\hat{\pi}_1 = p_1$ and $\hat{\pi}_2 = p_2$ are the sample proportions, had poor coverage, especially in extreme cases when n was small or the π values were near the boundaries of 0 or 1. They proposed a simple adjustment that improved the coverage.

Derrick's interval based on the z_6 statistic is a natural extension of the Wald interval as it also uses the sample proportions in the standard error estimate. We propose that applying the Agresti-Caffo adjustment to the z_6 estimator will improve its coverage.

Derrick also proposed a test statistic, z_8, which used a pooled estimator for the common population proportion in the standard error estimate. The test based on this statistic did not have the same Type I error problems with the parameters he used in his simulation as did the test based on z_6. However, Derrick did not consider any population proportion values close to 0 or 1. Therefore, while Derrick recommended the z_8 estimator, its performance was not analyzed for extreme cases with π_1 less than 0.15. Thus, we compare its performance to that of the adjusted z_6 estimator for these more extreme cases.

Agresti (2003) suggested a paired data adjustment similar to the Agresti-Caffo adjustment. We compare the coverage of this adjusted estimator to the unadjusted estimator, and also investigate adding this to the adjusted z_6 estimator. Finally, we propose a new estimator based on a combined test statistic analogous to the statistics of Bhoj (1978, 1984, 1989) and Dubnicka et al. (2002). We perform a simulation study to evaluate and compare the properties of these estimators.

2 Methods

Table 1: Notation summary for the adjustments made to statistics.

z	No Adjustments
z'	Agresti-Caffo Adjusted
z''	Both Adjustments (Unpaired and Paired) Applied

The mixed paired and unpaired design can be displayed as separate tables, as shown in Table 2 and Table 3. Table 2 organizes all of the counts of the unpaired data into n_1 total observations for sample 1 and n_2 total observations for sample 2. Table 3 organizes all of the counts of the paired data into categories that sum to n_{12} total observations.

Table 2: Response frequencies for two independent samples with a dichotomous response.

	Response		Total
	1	0	
Sample 1	e	f	n_1
Sample 2	g	h	n_2

Table 3: Response frequencies for two paired samples with a dichotomous response.

	Response Sample 2		Total
Response Sample 1	1	0	
1	a	b	m
0	c	d	$n_{12} - m$
Total	k	$n_{12} - k$	n_{12}

2.1 Unpaired Data Only

The most common interval estimator with unpaired data for the proportion difference is based on the statistic

$$z_1 = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}, \tag{2}$$

where π_1 and π_2 are estimated by $\hat{\pi}_1 = p_1 = \frac{e}{n_1}$ and $\hat{\pi}_2 = p_2 = \frac{g}{n_2}$.

The $(1 - \alpha) * 100\%$ confidence interval for $\pi_1 - \pi_2$, based on z_1, is

$$(p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}, \tag{3}$$

where $Z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the N(0,1) distribution.

2.2 Paired Data Only

An interval estimator with paired data for the proportion difference is based on the statistic (Agresti (2003))

$$z_3 = \frac{p_{12} - p_{21} - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_{1+}(1-\pi_{1+})+\pi_{+1}(1-\pi_{+1})-2(\pi_{11}\pi_{22}-\pi_{12}\pi_{21})}{n_{12}}}}, \tag{4}$$

where π_{1+}, π_{+1}, π_{11}, π_{12}, π_{21}, and π_{22} are estimated by $\hat{\pi}_{1+} = p_{1+} = \frac{a+b}{n_{12}}$, $\hat{\pi}_{+1} = p_{+1} = \frac{a+c}{n_{12}}$, $\hat{\pi}_{11} = p_{11} = \frac{a}{n_{12}}$, $\hat{\pi}_{12} = p_{12} = \frac{b}{n_{12}}$, $\hat{\pi}_{21} = p_{21} = \frac{c}{n_{12}}$, and $\hat{\pi}_{22} = p_{22} = \frac{d}{n_{12}}$, respectively.

The $(1-\alpha) * 100\%$ confidence interval for $\pi_1 - \pi_2$, based on z_3, is

$$p_{12} - p_{21} \pm Z_{\alpha/2}\sqrt{\frac{p_{1+}(1-p_{1+}) + p_{+1}(1-p_{+1}) - 2(p_{11}p_{22} - p_{12}p_{21})}{n_{12}}}. \tag{5}$$

2.3 Derrick's Wald Interval Using Both Paired and Unpaired Data

There are several statistics proposed by Derrick et al. (2015) that use all of the data in the paired/unpaired design. The test statistic is

$$z_6 = \frac{\bar{p}_1 - \bar{p}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_{12}+n_1} + \frac{\pi_2(1-\pi_2)}{n_{12}+n_2} - 2r_1(\frac{\sqrt{\pi_1(1-\pi_1)}\sqrt{\pi_2(1-\pi_2)}n_{12}}{(n_{12}+n_1)(n_{12}+n_2)})}}, \tag{6}$$

where π_1 and π_2 are estimated by $\hat{\pi}_1 = \bar{p}_1 = \frac{a+b+e}{n_{12}+n_1}$ and $\hat{\pi}_2 = \bar{p}_2 = \frac{a+c+g}{n_{12}+n_2}$, and r_1 is Pearson's phi correlation coefficient.

The corresponding $(1-\alpha) * 100\%$ confidence interval for $\pi_1 - \pi_2$, based on z_6, is

$$(\bar{p}_1 - \bar{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_{12}+n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_{12}+n_2} - 2r_1(\frac{\sqrt{\bar{p}_1(1-\bar{p}_1)}\sqrt{\bar{p}_2(1-\bar{p}_2)}n_{12}}{(n_{12}+n_1)(n_{12}+n_2)})}. \tag{7}$$

2.4 Agresti-Caffo Adjustment

Agresti and Caffo (2000) noted that "a nominal 95% interval for the difference of proportions has actual coverage probability below 0.93 in 88% of the cases with a standard Wald interval." However, with an adjusted interval based on adding two success and two failures to the data (for 95% confidence intervals), the coverage probability was below 0.93 only 1% of the time. The adjusted z_1 interval estimator (z_1') is

$$(\tilde{p}_1 - \tilde{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}, \tag{8}$$

where π_1 and π_2 are estimated by $\hat{\pi}_1 = \tilde{p}_1 = \frac{e+1}{n_1+2}$ and $\hat{\pi}_2 = \tilde{p}_2 = \frac{g+1}{n_2+2}$.

We applied this adjustment to the z_6 estimator. The adjusted z_6 estimator (z_6') using the add two successes and two failures adjustment is the same as that in equation 6, except that π_1 and π_2 are estimated by $\hat{\pi}_1 = \tilde{p}_1 = \frac{a+b+e+1}{n_{12}+n_1+2}$ and $\hat{\pi}_2 = \tilde{p}_2 = \frac{a+c+g+1}{n_{12}+n_2+2}$.

The corresponding $(1-\alpha)*100\%$ confidence interval for $\pi_1 - \pi_2$ is:

$$(\tilde{p}_1 - \tilde{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_{12}+n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_{12}+n_2+2} - 2r_1(\frac{\sqrt{\tilde{p}_1(1-\tilde{p}_1)}\sqrt{\tilde{p}_2(1-\tilde{p}_2)}n_{12}}{(n_{12}+n_1+2)(n_{12}+n_2+2)})}. \quad (9)$$

2.5 Derrick's Interval using the pooled proportion estimator

The last estimator considered for confidence intervals is based on Derrick's z_8 statistic. Under the null hypothesis, $H_0 : \pi_1 = \pi_2 = \pi$, the test statistic for z_8 is:

$$z_8 = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\pi(1-\pi)}{n_{12}+n_1} + \frac{\pi(1-\pi)}{n_{12}+n_2} - 2r_1(\frac{\sqrt{\pi(1-\pi)}\sqrt{\pi(1-\pi)}n_{12}}{(n_{12}+n_1)(n_{12}+n_2)})}}, \quad (10)$$

where π is estimated by $\hat{\pi} = \bar{p} = \frac{(n_1+n_{12})\bar{p}_1 + (n_2+n_{12})\bar{p}_2}{2n_{12}+n_1+n_2}$.

The corresponding $(1-\alpha)*100\%$ confidence interval for $\pi_1 - \pi_2$ is:

$$(\bar{p}_1 - \bar{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n_{12}+n_1} + \frac{\bar{p}(1-\bar{p})}{n_{12}+n_2} - 2r_1(\frac{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\bar{p}(1-\bar{p})}n_{12}}{(n_{12}+n_1)(n_{12}+n_2)})}. \quad (11)$$

3 Simulation Design

To consider how these statistics perform over a wide range of scenarios, a simulation study was performed. The simulation design follows that of Derrick et al. (2015). To generate the unpaired data, random samples of size n_1 and n_2 from a N(0,1) distribution were generated. These observations were then converted into binary '0' and '1' using critical values $C_{\pi i}$ of the normal distribution based on the known population proportions π_1 and π_2. If the number generated from the N(0,1) distribution was less than the critical value, it was assigned a '1', otherwise the number was assigned '0'.

The process to generate paired data followed a similar process. First, two random samples of size n_{12} from a N(0,1) distribution were generated. These are denoted by X_{ij} for $i = (1, 2)$ and $j = (1, 2, ..., n_{12})$. The X_{ij} values were then transformed into correlated normal bivariates Y_{ij} as follows:

$$Y_{1j} = \sqrt{\frac{1+\rho}{2}}X_{1j} + \sqrt{\frac{1-\rho}{2}}X_{2j}, \quad (12)$$

$$Y_{2j} = \sqrt{\frac{1+\rho}{2}}X_{2j} - \sqrt{\frac{1-\rho}{2}}X_{1j}, \tag{13}$$

where ρ is the Pearson correlation between the two random variables and $(Y_1, Y_2) \sim BVN(\mu, \Sigma)$ where $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

The normal deviates for the paired data are then compared to the critical values $C_{\pi i}$ and are converted to '0' and '1' in the same way as the unpaired data. The simulation parameters are listed in Table 4.

Table 4: Simulation Parameters

π_1	0.05, 0.15, 0.30, 0.50
π_2	0.05, 0.15, 0.30, 0.50
n_1	10, 30, 50, 100
n_2	10, 30, 50, 100
n_{12}	10, 30, 50, 100
ρ	0, 0.25, 0.50, 0.75

The simulation was designed to evaluate various configurations across a wide range of scenarios, including proportions ranging from $\pi = 0.05$ to $\pi = 0.50$ and sample sizes from $n_i = 10$ to $n_i = 100$. Therefore, our simulation design examined cases with both equal and unequal group proportions and equal and unequal sample sizes. Our study considered many of the parameters proposed by Derrick et al. (2015). However, we included the lower value of $\pi = 0.05$ because, as stated in Agresti and Caffo (2000), the coverage of the Wald intervals decreased substantially in cases with low success rates, particularly when π was less than 0.10. Varying ρ is considered as it is known to have an impact on paired sample tests Derrick et al. (2015).

10,000 random samples were generated for each combination of parameters. For each sample, we calculated the 95% confidence interval for the true difference in proportions. The coverage was determined as the proportion of the 10,000 intervals that contained the true proportion difference. We measured interval width by averaging the interval widths over 10,000 intervals. This process was repeated for all possible parameter combinations, resulting in one data point for each combination.

All simulations were performed using R.

4 Results

4.1 Coverage Comparisons

We compared the coverage of the interval estimators described in Section 2. The data points in these plots are the estimated coverage for each combination of simulation parameters.

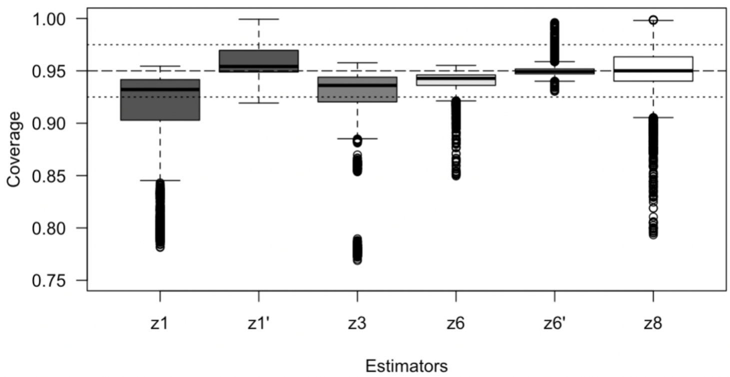


Figure 1: Estimated coverage of 95 percent confidence intervals.

Bradley (1978) proposed a criteria that for a test to be Type I error robust under H_0 for $\alpha = 0.05$, the Type 1 error rate should be between 0.025 and 0.075. As Type I error rate is analogous to confidence interval coverage, this translates to robust coverage for a 95% confidence interval having coverage between 0.925 and 0.975. These bounds on robust coverage are shown by dashed lines around 0.95 in Figure 1. All four non-adjusted estimators, z_1, z_3, z_6, and z_8, exhibited substantial undercoverage. In particular, we note that Derrick's recommended z_8 experiences undercoverage with the inclusion of the $\pi_{1,2} = 0.05$ parameter, meaning it does not provide robust coverage in cases with proportions less than 0.15. It is also important to note that the z_6 estimator was affected by the Agresti-Caffo (2000) adjustment in a similar way as the z_1 estimator. Both adjusted intervals eliminated the undercoverage seen in the non-adjusted intervals but had some tendency to over-cover.

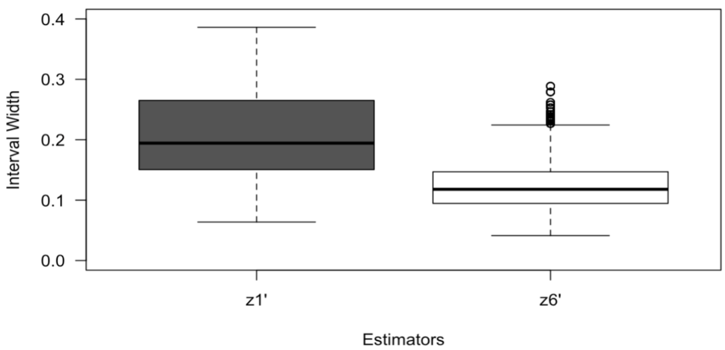


Figure 2: Mean interval width of z_1' and z_6' estimators.

Comparing the interval widths of the two adjusted tests in Figure 2, it is evident that the z'_6 estimator is more precise than the z'_1 estimator. Furthermore, while both adjusted tests had coverage within the acceptable range for almost all cases, as seen in Figure 1, the z'_6 had more cases with coverage closer to 0.95 and fewer cases of overcoverage than z'_1. Thus, while the z'_6 estimator is slightly more complicated to compute, it provided a better estimator than the z'_1 estimator.

4.2 z_6 and z'_6 Coverage for Specific Settings

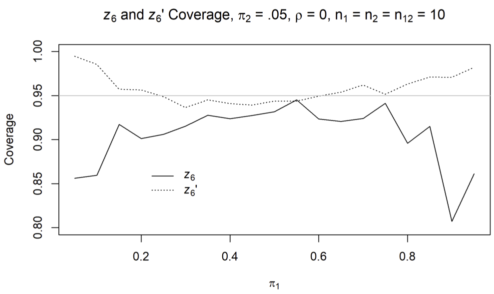
Agresti and Caffo (2000) found that the worst coverage for the Wald interval occurred in cases where the sample size was small and π was close to 0 or 1. Figure 3 shows coverage for the z_6 and z'_6 estimators over π_1 values ranging from 0.05 to 0.95. The parameters π_2, n_1, n_2, n_{12}, and ρ were kept constant over the entire range of π_1 values for each plot.

When $\pi_2 = 0.05$ (Figure 3(a,b,c)) and π_1 approached 0 or 1, the coverage of the z_6 interval also approached 0. However, the z'_6 estimator had coverage close to the nominal level, even in the cases near the boundaries. Figure 3(a,b,c) kept π_2 constant at 0.05 and the sample sizes n_1, n_2, n_{12} constant at 10 but varied ρ from 0 to 0.25 to 0.75. There were no significant changes in the coverage between the different ρ values, so the correlation of the paired data did not appear to have an impact on the coverage of the confidence intervals. When π_2 was increased to 0.5 in Figure 3(d,e), there was no drop off in coverage for either of the confidence intervals. However, the z'_6 interval coverage was closer to the desired 0.95 level than z_6 across the span of π_1 values.

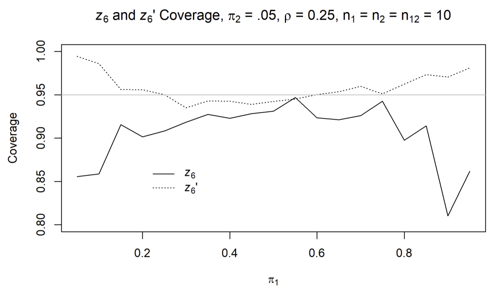
In Figure 4, $\pi_2 = 0.3$ and $\rho = 0.25$ for all of the plots, while different combinations of sample sizes were considered. When all of the sample sizes were equal to 10 (Figure 4(a)), coverage was near 0.95 across all π_1 values for z'_6. However, the coverage for z_6 was never as close to the desired 0.95 coverage as z'_6. The improved performance of z'_6 over z_6 occurred because the adjustment brings the proportions closer together, enabling a better estimate of the true proportion difference, even with small sample sizes. When n_1 was increased to 50 while $n_2 = n_{12} = 10$ (Figure 4(b)), there was fairly constant coverage for both the z_6 and z'_6 estimators, but the z'_6 estimator remained closer to the desired 0.95 coverage. When n_2 was increased to 50 and $n_1 = n_{12} = 10$ (Figure 4(c)), coverage for z_6 was consistently lower than that for z'_6. When n_{12} was increased to 50 and $n_1 = n_2 = 10$ (Figure 4(d)), there was consistent coverage around 0.95 for both the z_6 and z'_6 estimators. The same result was seen when all of the sample sizes were increased to 30 (Figure 4(e)).

5 Paired Data Adjustment

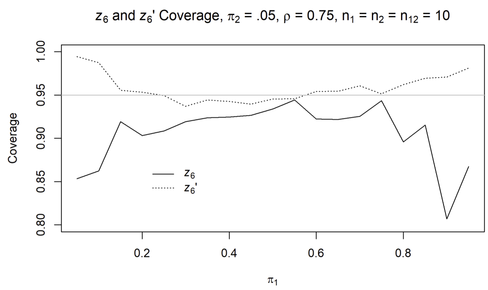
Agresti (2003) suggested an adjustment to the z_3 estimator discussed in Section 2.2. Specifically, this adjustment involves adding one to each of the cells in Table 2 before computing proportions for the confidence interval. The adjustment was applied to both z_3, which only uses paired data, as well as to z'_6 to create a fully adjusted estimator.



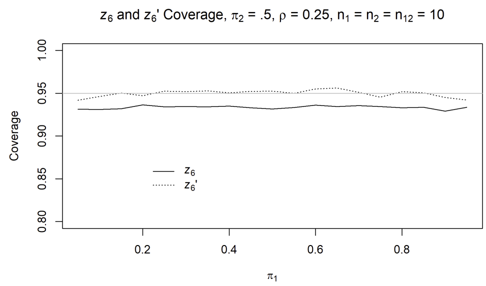
(a)



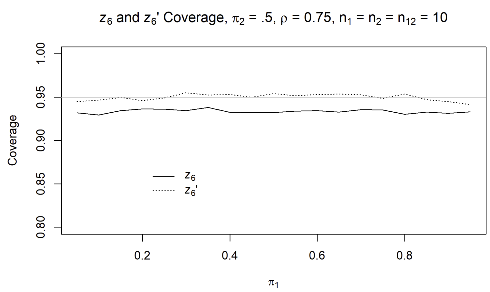
(b)



(c)

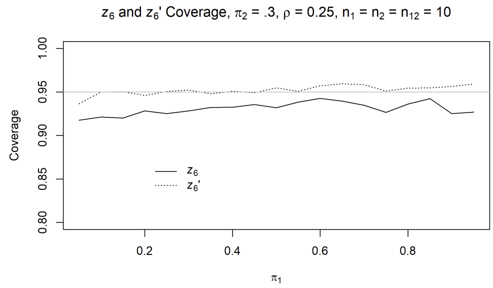


(d)

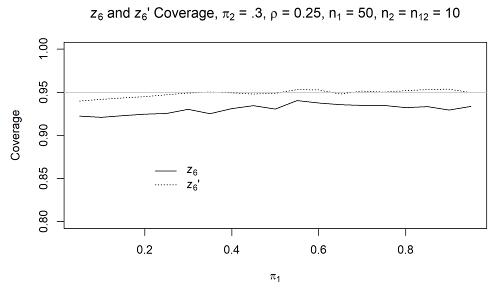


(e)

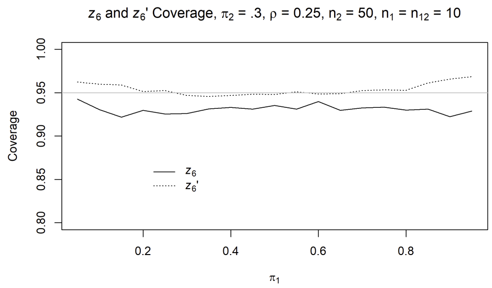
Figure 3: Coverage for z_6 and z_6' estimators for π_1 ranging between 0.05 and 0.95, $n_1 = n_2 = n_{12} = 10$, $\pi_2 = 0.05, 0.5$, and $\rho = 0, 0.25, 0.75$.



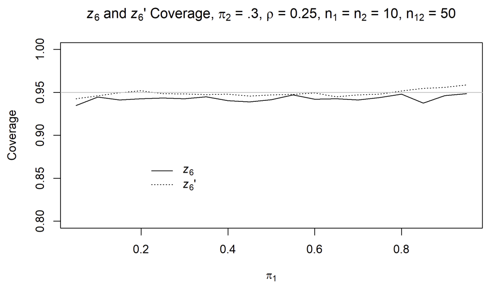
(a)



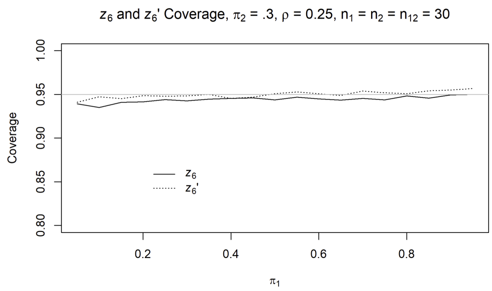
(b)



(c)



(d)



(e)

Figure 4: Coverage for z_6 and z_6' estimators for π_1 ranging between 0.05 and 0.95 and $\pi_2 = 0.3$, $\rho = 0.25$, and different combinations sample sizes of 10, 30, and 50.

5.1 Paired Data Adjustment to z_3

The adjustment is applied to the proportions from equation (4):

$$\begin{aligned} \tilde{p}_{1+} &= \tfrac{a+b+2}{n_{12}+4}, \quad \tilde{p}_{+1} = \tfrac{a+c+2}{n_{12}+4}, \\ \tilde{p}_{11} &= \tfrac{a+1}{n_{12}+4}, \quad \tilde{p}_{12} = \tfrac{b+1}{n_{12}+4}, \\ \tilde{p}_{21} &= \tfrac{c+1}{n_{12}+4}, \quad \tilde{p}_{22} = \tfrac{d+1}{n_{12}+4}. \end{aligned}$$

Therefore, the confidence interval based on the adjusted z_3 estimator (z_3') becomes:

$$\tilde{p}_{12} - \tilde{p}_{21} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}_{1+}(1-\tilde{p}_{1+}) + \tilde{p}_{+1}(1-\tilde{p}_{+1}) - 2(\tilde{p}_{11}\tilde{p}_{22} - \tilde{p}_{12}\tilde{p}_{21})}{n_{12}+4}}. \tag{14}$$

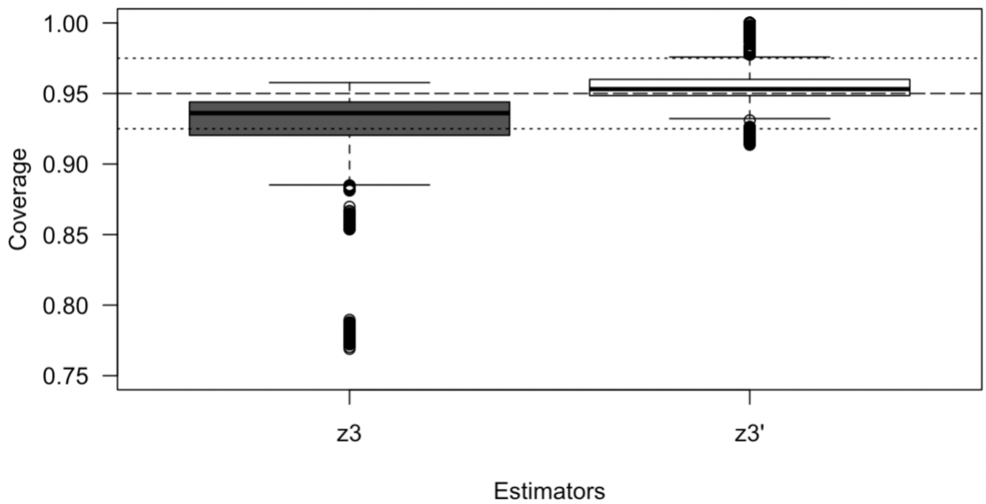


Figure 5: Confidence interval coverage for the z_3 and z_3' estimators.

The adjustment to the paired data for the z_3 estimator eliminated all of the cases where the coverage was below 0.9. Figure 5 shows that with the adjustment, nearly all cases considered fell within the acceptable range of coverage. This is analogous to the results of the adjustment that was used on the unpaired data for the z_1 estimator.

5.2 Paired Data Adjustment to z_6'

We next applied the additional paired adjustment to the (already) adjusted z_6'. This created new estimators where:

$$\check{p_1} = \frac{a+b+e+1+2}{n_{12}+n_1+2+4}, \tag{15}$$

$$\check{p_2} = \frac{a+c+g+1+2}{n_{12}+n_2+2+4}. \tag{16}$$

Substituting these estimators in equation (6), results in the statistic z_6''. Then the confidence interval for $\pi_1 - \pi_2$ based on the z_6'' statistic is

$$\check{p_1} - \check{p_2} \pm Z_{\alpha/2}\sqrt{\frac{\check{p_1}(1-\check{p_1})}{n_{12}+n_1+6} + \frac{\check{p_2}(1-\check{p_2})}{n_{12}+n_2+6} - 2r_1(\frac{\sqrt{\check{p_1}(1-\check{p_1})}\sqrt{\check{p_2}(1-\check{p_2})}(n_{12}+4)}{(n_{12}+n_1+6)(n_{12}+n_2+6)}}). \tag{17}$$

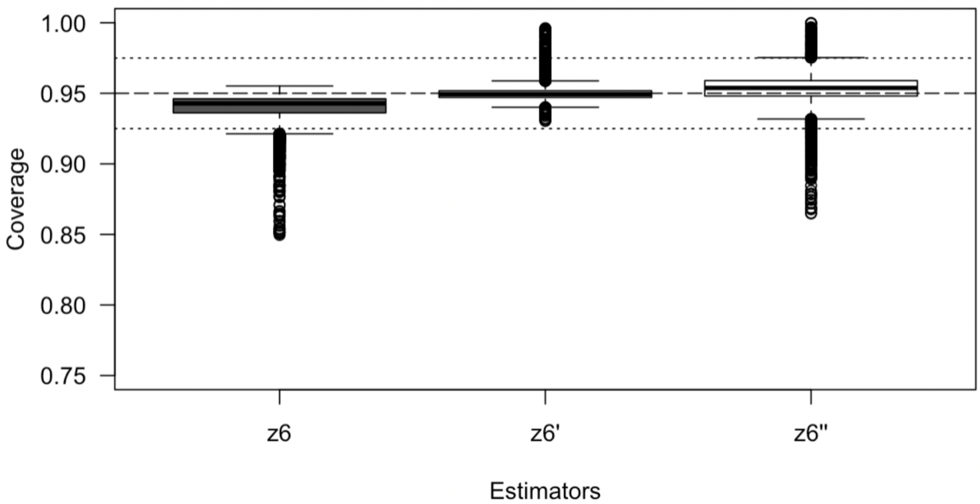
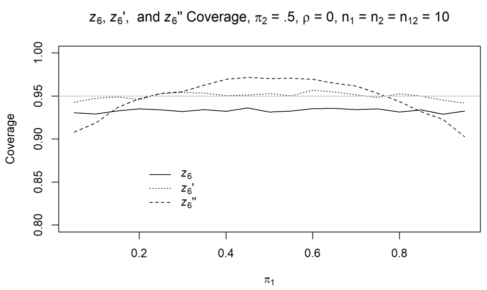


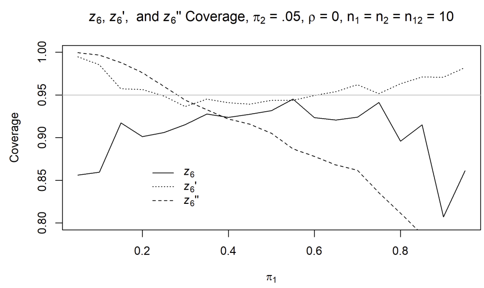
Figure 6: Confidence interval coverage of the z_6, z_6', and z_6'' estimators.

The adjustment used to create z_6'' resulted in more undercoverage than seen in z_6' (Figure 6). Similar to the results for z_3', the adjustment raised the mean coverage compared to its non-adjusted counterpart, though it did not eliminate many of the undercoverage situations that z_6' corrected.

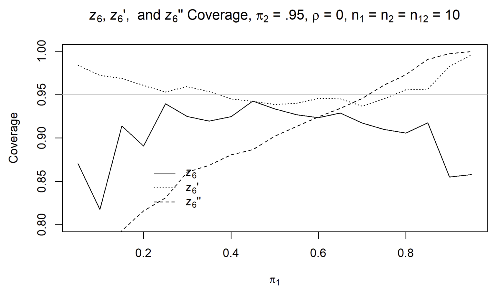
5.3 Investigating the poor coverage cases of z_6''



(a)



(b)



(c)

Figure 7: Coverage for z_6, z_6', and z_6'' estimators for π_1 ranging between 0.05 and 0.95, $n_1 = n_2 = n_{12} = 10$, $\pi_2 = 0.05, 0.5, 0.95$, and $\rho = 0$.

To further understand where poor coverage was happening with the z_6'' estimator, we plotted coverage for z_6'' against z_6 and z_6' in Figure 7. As shown in Figure 7(a), when $\pi_2 = 0.5$, $\rho = 0$, and $n_1 = n_2 = n_{12} = 10$, z_6'' had acceptable coverage between $\pi_1 \approx 0.2$ and $\pi_1 \approx 0.8$. When $\pi_2 = 0.05$, $\rho = 0$, and $n_1 = n_2 = n_{12} = 10$ (Figure 7(b)), there was acceptable coverage until the difference between π_1 and π_2 was greater than 0.3. The same pattern was seen when $\pi_2 = 0.95$, $\rho = 0$, and $n_1 = n_2 = n_{12} = 10$ (Figure 7(c)). More generally, we saw that z_6'' had acceptable coverage when the difference of π values was 0.3 or less.

6 Linear Combination Statistic

Bhoj (1978) proposed a statistic that was a linear combination of paired and unpaired t statistics

$$t_{combined} = \lambda t_{paired} + (1 - \lambda) t_{unpaired}. \tag{18}$$

We investigate paired and unpaired adjustments to a similar linear combination of the statistics z_3 and z_1

$$z_{combined} = \frac{1}{2}z_1 + \frac{1}{2}z_3. \tag{19}$$

We chose an equal weighting for the $z_{combined}$ statistic for mathematical convenience. Since both z_1 and z_3 asymptotically follow a N(0,1) distribution, if it is assumed that these statistics are independent, then it can be shown that the distribution of $z_{combined}$ is $N(0, \frac{1}{\sqrt{2}})$.

A confidence interval based on $z_{combined}$ can be derived using the probability statement

$$P\left(-Z_{\alpha/2} \leq \frac{1}{2}z_1 + \frac{1}{2}z_3 \leq Z_{\alpha/2}\right) = 1 - \alpha. \tag{20}$$

The $z_{combined}$ confidence interval is

$$\frac{B(p_{1u} - p_{2u}) + A(p_{12} - p_{21})}{A + B} \pm \frac{2AB(Z_{\alpha/2})}{A + B}, \tag{21}$$

where

$$A = \sqrt{\frac{p_{1u}(1 - p_{1u})}{n_1} + \frac{p_{2u}(1 - p_{2u})}{n_2}}, \tag{22}$$

$$B = \sqrt{\frac{p_{1+}(1 - p_{1+}) + p_{+1}(1 - p_{+1}) - 2(p_{11}p_{22} - p_{12}p_{21})}{n_{12}}}, \tag{23}$$

$$\begin{aligned} p_{1+} &= \tfrac{a+b}{n_{12}}, & p_{+1} &= \tfrac{a+c}{n_{12}}, \\ p_{11} &= \tfrac{a}{n_{12}}, & p_{12} &= \tfrac{b}{n_{12}}, \\ p_{21} &= \tfrac{c}{n_{12}}, & p_{22} &= \tfrac{d}{n_{12}}, \\ p_{1u} &= \tfrac{e}{n_1}, & p_{2u} &= \tfrac{g}{n_2}. \end{aligned}$$

Next, we considered adding the unpaired (z_1) adjustment recommended by Agresti and Caffo (2000) to $z_{combined}$. The unpaired adjusted $z'_{combined}$ confidence interval is:

$$\frac{B(\tilde{p}_{1u} - \tilde{p}_{2u}) + \tilde{A}(p_{12} - p_{21})}{\tilde{A} + B} \pm \frac{2\tilde{A}B(Z_{\alpha/2})}{\tilde{A} + B}, \tag{24}$$

where $\tilde{p}_{1u} = \frac{e+1}{n_1+2}$, $\tilde{p}_{2u} = \frac{g+1}{n_2+2}$, $\tilde{A} = \sqrt{\frac{\tilde{p}_{1u}(1-\tilde{p}_{1u})}{n_1+2} + \frac{\tilde{p}_{2u}(1-\tilde{p}_{2u})}{n_2+2}}$, and $Z_{\alpha/2}$ represents the $1 - \alpha/2$ quantile of the $N(0, \frac{1}{\sqrt{2}})$ distribution.

Finally, the paired (z_3) adjustment was added to $z'_{combined}$ making it fully adjusted for both unpaired and paired data ($z''_{combined}$). The $z''_{combined}$ confidence interval is:

$$\frac{\tilde{B}(\tilde{p}_{1u} - \tilde{p}_{2u}) + \tilde{A}(\tilde{p}_{12} - \tilde{p}_{21})}{\tilde{A} + \tilde{B}} \pm \frac{2\tilde{A}\tilde{B}(Z_{\alpha/2})}{\tilde{A} + \tilde{B}}, \tag{25}$$

where $\tilde{p}_{12} = \frac{b+1}{n_{12}+4}$, $\tilde{p}_{21} = \frac{c+1}{n_{12}+4}$, $\tilde{B} = \sqrt{\frac{\tilde{p}_{1+}(1-\tilde{p}_{1+})+\tilde{p}_{+1}(1-\tilde{p}_{+1})-2(\tilde{p}_{11}\tilde{p}_{22}-\tilde{p}_{12}\tilde{p}_{21})}{n_{12}+4}}$, and $Z_{\alpha/2}$ represents the $1-\alpha/2$ quantile of the $N(0, \frac{1}{\sqrt{2}})$ distribution.

7 Comparison of Statistics that use both Paired and Unpaired Data

We compared all estimators discussed thus far that use both paired and unpaired data.

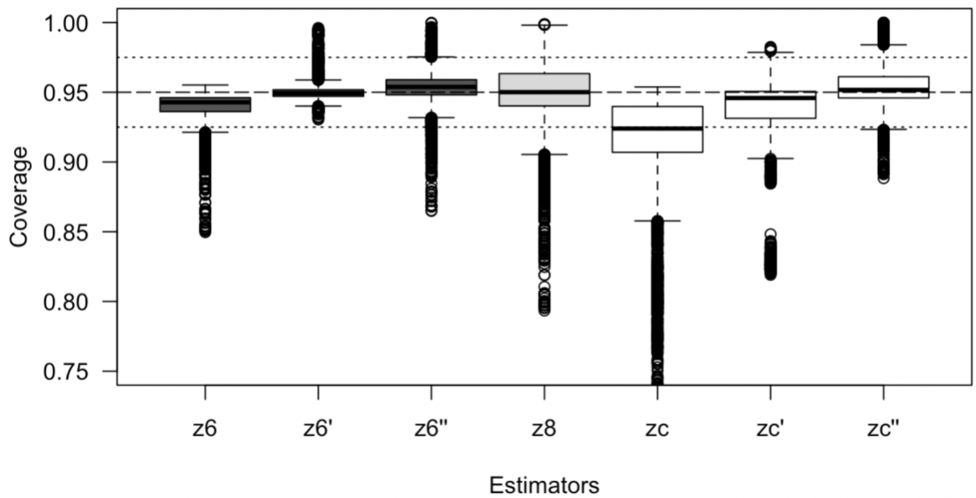


Figure 8: Confidence interval coverage of the z_6, z_6', z_6'', z_8, $z_{combined}$, $z'_{combined}$, $z''_{combined}$ estimators.

Table 5: Percent coverage for intervals using all data. Overcoverage for the 95% confidence interval is > 0.975, adequate coverage is between 0.925 and 0.975, and undercoverage is < 0.925.

Percent Coverage	z_6	z_6'	z_6''	z_8	$z_{combined}$	$z'_{combined}$	$z''_{combined}$
Over 0.975	0	1.71	4.66	13.09	0	0.34	11.87
Between 0.925 and 0.975	89.82	98.29	91.78	73.09	48.29	80.13	84.47
Below 0.925	10.18	0	3.56	13.82	51.71	19.53	3.66

Figure 8 shows that $z'_{combined}$ had improved coverage compared to $z_{combined}$. Furthermore, $z''_{combined}$ had improved coverage compared to $z'_{combined}$. As seen in Table 5, z_6' had the highest percentage of cases with acceptable coverage with 98.29% of the confidence intervals having coverage between 0.925 and 0.975. Furthermore, z'_6 did not have any intervals with coverage below 0.925 and had coverage above 0.975 only 1.71% of the time. While $z''_{combined}$, had less acceptable coverage than z'_6, it is possible that further refinements, such as using unequal weights, could improve its coverage.

8 Conclusion

Situations where a mixture of paired and unpaired data occur are common in many research settings. Researchers often discard one portion of data to be able to estimate the difference in proportions, but this approach can lead to a significant loss of precision. To address this, we proposed several new interval estimators that incorporated both paired and unpaired data.

The interval based on z'_6, incorporating the adjustment of Agresti and Caffo (2000), was found to be the best performing, with acceptable coverage for nearly all cases investigated, while providing more precise intervals than intervals that discard paired or unpaired data.

The adjustment for paired data (z'_3) improved the coverage of z_3 and $z'_{combined}$, but it worsened the coverage performance of z'_6. Agresti and Caffo (2000) noted that there is an optimal number of observations to add to the data, and over adjusting can lead to confidence intervals under-performing. For mixed paired and unpaired data, it appears that combining the adjustments that were optimal for unpaired-only and paired-only data over adjusts. Future research is needed to investigate the optimal adjustment for mixed paired and unpaired data.

z'_6 provides the most acceptable coverage, with robust coverage 98.29% of the time according to our simulation, while providing greater precision than interval estimators that discard data. Thus, z'_6 is recommended as the confidence interval of choice when estimating the difference in proportions in mixed paired and unpaired designs.

Acknowledgement

We would like to thank NSF for funding with grant DMS − 2244160 and the opportunity to do research at the University of North Carolina at Greensboro. We would also like to thank Dr. Sat Gupta for his leadership of this REU Program.

References

Agresti, A. (2003). *Categorical Data Analysis. 2nd Edition.* John Wiley & Sons, Inc.

Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4):280–288.

Bhoj, D. S. (1978). Testing equality of means of correlated variates with missing observations on both responses. *Biometrika*, 65(1):225–228.

Bhoj, D. S. (1984). On difference of means of correlated variates with incomplete data on both responses. *Journal of Statistical Computation and Simulation*, 19(4):275–289.

Bhoj, D. S. (1989). On comparing correlated means in the presence of incomplete data. *Biometrical journal*, 31(3):279–288.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2):144–152.

Choi, S. and Stablein, D. (1982). Practical tests for comparing two proportions with incomplete data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 31(3):256–262.

Derrick, B., Dobson-McKittrick, A., Toher, D., and White, P. (2015). Test statistics for comparing two proportions with partially overlapping samples. *Journal of Applied Quantitative Methods*, 10(3).

Dubnicka, S. R., Blair, R. C., and Hettmansperger, T. P. (2002). Rank-based procedures for mixed paired and two-sample designs. *Journal of Modern Applied Statistical Methods*, 1(1):6.

Johnson, E. N. and Richter, S. J. (2022). Permutation tests for mixed paired and two-sample designs. *Computational Statistics*, 37(2):739–750.

Magel, R. C. and Fu, R. (2014). Proposed nonparametric test for the mixed two-sample design. *Journal of statistical Theory and Practice*, 8:221–237.

Tang, M.-L. and Tang, N.-S. (2004). Exact tests for comparing two paired proportions with incomplete data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 46(1):72–82.