

# Non Iterative Confidence Interval Methods for the Difference between Two Proportions

J Reed III

## Citation

J Reed III. *Non Iterative Confidence Interval Methods for the Difference between Two Proportions*. The Internet Journal of Epidemiology. 2007 Volume 5 Number 2.

## Abstract

The construction of a confidence interval for the difference between two independent proportions is one of the most basic analyses in statistical inference. The Wald asymptotic methods, with and without a continuity correction have less than nominal coverage probability characteristics but continue to be used in spite of this known poor behavior. For the equal sample size case, the Wald-z coverage probability is subnominal overall. The Wald-c coverage probability always exceeds the nominal level and has interval width larger than Wald-z. For the unequal sample size case, Wald-z is always subnominal while Wald-c exceeds the nominal level. Newcombe's hybrid method and the Agresti-Caffo methods have coverage probabilities that are near nominal for either the equal or unequal sample sizes. Considering the coverage probability criterion, Newcombe's hybrid method and Agresti-Caffo method demonstrate superior coverage properties.

## INTRODUCTION

In analyzing the results of studies we often encounter the problem of comparing two binomial success probabilities  $p_1$  and  $p_2$ ,  $p_1 > 0$  and  $p_2 > 0$ . Implicit in this comparison is the independent observations  $X_1 \sim B(n_1, p_1)$  and  $X_2 \sim B(n_2, p_2)$ . The most common comparison is the hypothesis  $H_0: p_1 = p_2$  versus  $H_a: p_1 \neq p_2$ . Accompanying this test is the construction of a confidence interval for the difference between  $p_1$  and  $p_2$ . The most commonly described textbook method is the Wald-z. Occasionally, a continuity corrected version is given (Wald-c). Nearly all introductory statistics textbooks include the Wald-z confidence interval and issue a warning - usually in a footnote - when not to use it.

The general problems associated with constructing a confidence interval for the difference between two independent proportions are similar to those associated with constructing a confidence interval for a single proportion - overshoot and poor coverage probability. Despite these shortcomings, the Wald-z and Wald-c methods continue to dominate in textbooks. The purpose of this paper is to review the coverage probability functions of the Wald methods and a set of alternative non-iterative methods for computing a confidence interval for the difference between two independent proportions.

## METHODS

The primary criteria for evaluating a confidence interval method is the coverage probability function. This coverage probability for the difference between two independent proportions,  $C(\pi_1, \pi_2 | n_1, n_2, \mathbb{I})$ , is found by fixing  $n_1, n_2, \pi_1$ , and  $\pi_2$ , then computing the confidence interval for each  $x_i = 0, \dots, n_i$  for  $i = 1, 2$ . The coverage probability is then defined by:

$$C(\pi_1, \pi_2 | n_1, n_2, \mathbb{I}) = \sum \Pr(X_1 = x_1 | n_1, \pi_1) \Pr(X_2 = x_2 | n_2, \pi_2) \mathbb{I}(\pi_1, \pi_2 | x_1, x_2, n_1, n_2, \mathbb{I}).$$

$$\text{If } (\pi_1 - \pi_2) \in [\text{LB}(x_1, x_2, n_1, n_2, \mathbb{I}), \text{UB}(x_1, x_2, n_1, n_2, \mathbb{I})], \\ \mathbb{I}(\pi_1, \pi_2 | x_1, x_2, n_1, n_2, \mathbb{I}) = 1 \text{ and } 0 \text{ otherwise.}$$

The Wald-z and Wald-c confidence interval lower upper bounds for the difference between two independent proportions are defined as (See Table 1 for a typical data structure):

### Figure 1

Wald-z	$\text{LB} = (p_1 - p_2) - z_{\alpha/2} \sqrt{(ac/m^3 + bd/n^3)}$
	$\text{UB} = (p_1 - p_2) + z_{\alpha/2} \sqrt{(ac/m^3 + bd/n^3)}$
Wald-c	$\text{LB} = (p_1 - p_2) - [z_{\alpha/2} \sqrt{(ac/m^3 + bd/n^3)} + (1/m + 1/n)/2]$
	$\text{UB} = (p_1 - p_2) + [z_{\alpha/2} \sqrt{(ac/m^3 + bd/n^3)} + (1/m + 1/n)/2]$

Figure 2

Table 1: Methods for calculation of confidence intervals for the difference between independent proportions.

	Sample 1	Sample 2	
+	a	b	$p_1 = a/m,$
-	c	d	$p_2 = b/n,$
Total	m	n	$\theta = \pi_1 - \pi_2$ and $\theta' = p_1 - p_2$

Figure 1 shows the 95% confidence interval coverage probability function for the Wald-z and Wald-c methods as a function of  $\pi_1, \pi_1 \in [0,1]$  for  $n_1 = n_2 = 20$  and  $p_2 = 0.3$ . The sawtooth appearance of the coverage functions is due to the discontinuities for values of  $p_1 - p_2$  corresponding to any lower or upper limits in the set of confidence intervals. Like the one sample cousin, the Wald-z coverage probability curve is subnominal and less than 0.95 overall. The Wald-c coverage probability always exceeds 0.95 overall with interval widths larger than Wald-z.

Figure 3

Figure 1: Coverage probabilities for nominal 95% Wald-z and Wald-c as a function of p when p=0.3 with n=n=20.

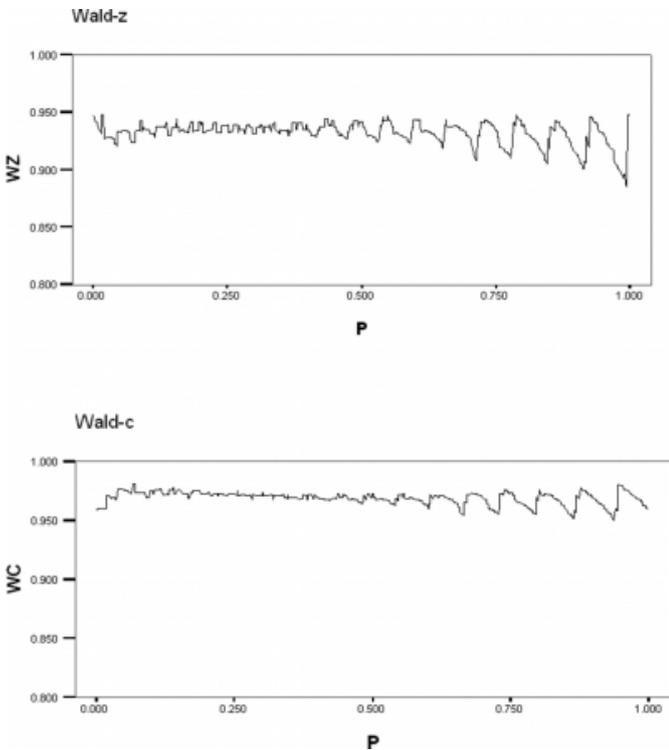
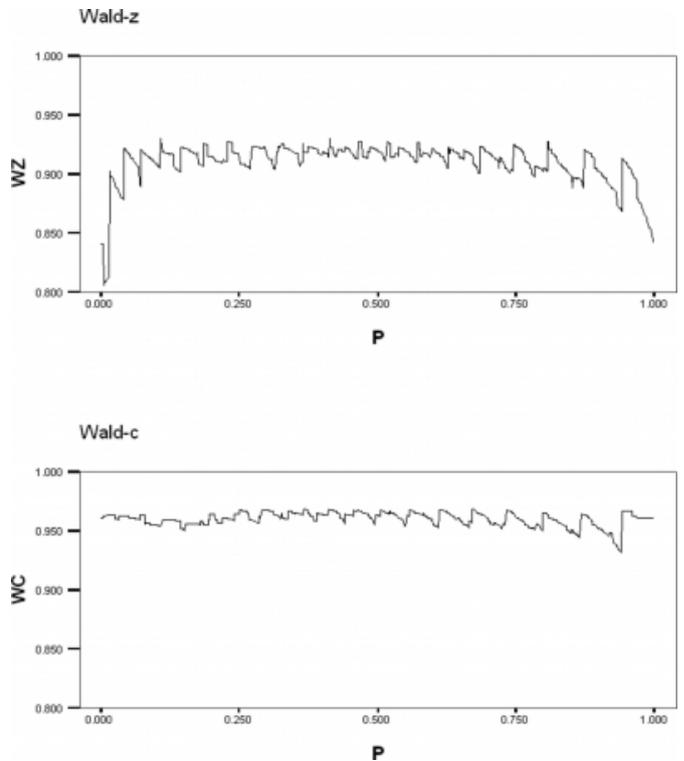


Figure 2 shows the 95% confidence interval coverage probability function for the Wald-z and Wald-c methods as a function of  $\pi_1, \pi_1 \in [0,1]$  for  $n_1 = 20, n_2 = 10$  and  $p_2 = 0.2$ . The Wald-z coverage probability curve is subnominal for differences in proportions near 0 and 1 and less than 0.95

overall.

Figure 4

Figure 2: Coverage probabilities for nominal 95% Wald-z and Wald-c as a function of p when p=0.3 with n=20, n=10.



Beal, (Beal, 1987) evaluated asymptotic methods for computing confidence interval for the differences of two independent proportions. All involved identifying the interval within which  $(\hat{\theta} - \theta')^2 \leq z^2 V(\hat{\theta}, \theta')$ , where  $\hat{\theta} = p_1 - p_2$ , and  $V(\hat{\theta}, \theta') = u \{ 4u(1-u) \} = \pi_1(1-\pi_1)/m + \pi_2(1-\pi_2)/n$ . The Haldane (H) and Jeffreys-Perks (JP) method provides non-degenerative confidence intervals for all values of  $p_1$  and  $p_2$  unlike Wald-z or Wald-c. H and JP generally performed better than the Wald-z and Wald-c and of the two, JP was preferred (Beal, 1987, Radhakrishna, et al, 1992). The Haldane and Jeffreys-Perks lower and upper limits are defined by:

Figure 5

H  $LB = \theta^* - w$  and  $UB = \theta^* + w$ ,  
 where  $\theta^* = (\hat{\theta} + z^2 v (1 - 2\psi')) / (1 + z^2 u)$ ,  
 $w = [z^2 (1 + z^2 u)] \sqrt{u [4\psi'(1 - \psi') - \theta^2] + 2v(1 - 2\psi')\theta^* + 4z^2 u^2 (1 - \psi')\psi' + z^2 v^2 (1 - 2\psi')^2}$   
 $\psi' = (a/m + b/n)/2, u = (1/m + 1/n)/4$ , and  $v = (1/m - 1/n)/4$

JP  $LB = \theta^* - w$  and  $UB = \theta^* + w$ ,  
 Where  $\psi'$  from the Haldane method is:  $\psi' = [(a+0.5)/(m+1) + (b+0.5)/(n+1)]/2$

Newcombe compared eleven methods for estimating the difference between independent proportions (Newcombe, 1998). Similar to the single proportion, the virtues of Wald-z and Wald-c methods are in their simplicity but overshoot and inappropriate intervals are still common. The Haldane and Jeffreys-Perks methods attempt to overcome the overshoot and inappropriate intervals while maintaining closed-form tractability. Newcombe concluded that both H and JP were improvements over the Wald-z and Wald-c methods, but both were still inadequate. Newcombe recommended a hybrid method based on Wilson's score method for a single proportion without continuity correction (NS). The LB and UB for the NS method are:

Figure 6

NS

$$LB = (p_1 - p_2) - \delta, \text{ where}$$

$$\delta = \sqrt{\{(a/m - l_1)^2 + (u_2 - b/n)^2\}} = z_{\alpha/2} \sqrt{\{l_1(1 - l_1)/m + u_2(1 - u_2)/n\}}$$

$$UB = (p_1 - p_2) + \epsilon, \text{ where}$$

$$\epsilon = \sqrt{\{(u_1 - a/m)^2 + (b/n - l_2)^2\}} = z_{\alpha/2} \sqrt{\{u_1(1 - u_1)/m + l_2(1 - l_2)/n\}}, \text{ and}$$

$l_1, l_2, u_1, u_2$  are the lower and upper bounds for the two proportions  $p_1$  and  $p_2$  using Wilson's score method.

Agresti and Coull's adjustment to the Wald method for a single proportion adds  $t/2$  successes and  $t/2$  failures (Agresti and Coull, 1998). Agresti and Caffo later suggest that by adding two successes and two failures (total) to the two-sample method would improve the simple Wald method (Agresti and Caffo, 2000). This is an adjustment that adds one pseudo observation of each type to each sample. For instance, for sample  $i$ ,  $p_i = (r_i + 1)/(n_i + 2)$ .

For  $p_1 - p_2$ , Carlin and Lewis (Carlin and Lewis, 1996) considered independent uniform prior distributions for  $p_1$  and  $p_2$ . The posterior distribution of  $p_i$  is  $\pi$  with mean of  $(x_i + 1)/(n_i + 2)$  and variance  $p_i(1 - p_i)/(n_i + 3)$ . Then, a crude normal approximation for the distribution of the difference of the posterior  $\pi$  variates leads to a closed form confidence interval.

Figure 7

CL

$$LB = (p_1 - p_2) - z_{\alpha/2} \sqrt{\{p_1(1 - p_1)/(n_1 + 3) + p_2(1 - p_2)/(n_2 + 3)\}}$$

$$UB = (p_1 - p_2) + z_{\alpha/2} \sqrt{\{p_1(1 - p_1)/(n_1 + 3) + p_2(1 - p_2)/(n_2 + 3)\}}$$

Zou and Donner (Zou and Donner, 2004) proposed using Fisher's z transformation to the estimator  $\pi'$ , where  $\pi' = p_1 - p_2$ . By letting  $\pi' = \sqrt{(ac/m^3 + bd/n^3)}$ , the resulting interval for  $\pi'$  is given by:

Figure 8

FZ

$$UB = (\exp(2u) - 1) / (\exp(2u) + 1)$$

$$LB = (\exp(2l) - 1) / (\exp(2l) + 1)$$

where  $l$  and  $u$  are given by:  $\frac{1}{2} \log\left[\frac{(1 + \pi')}{(1 - \pi')}\right] \pm z_{\alpha/2} \pi' / (1 - \pi'^2)$

RESULTS

Figure 3 shows the 95% confidence interval coverage probability function for the Newcombe NS, Haldane, Jeffreys-Perks, and Agresti-Caffo, Carlin-Lewis and Fisher's-Z methods as a function of  $\pi_1, \pi_2 \in [0, 1]$  for  $n_1 = n_2 = 20$  and  $p_2 = 0.3$ . The NS, Agresti-Caffo and Fisher's-Z methods demonstrate coverage probabilities that are near nominal over  $\pi_1 \in [0, 1]$ . Table 2 details the corresponding minimum, maximum, mean, and standard deviation for the coverage probability of these confidence interval methods. The Newcombe NS, Fisher's-Z, Agresti-Caffo and Jeffreys-Perks methods are near nominal while the Wald-z is subnominal.

Figure 9

Figure 3: Coverage probabilities for nominal 95% Newcombe NS, Haldane, Jeffreys-Perks, and Agresti-Caffo as a function of  $p$  when  $p=0.3$  with  $n=n=20$ .

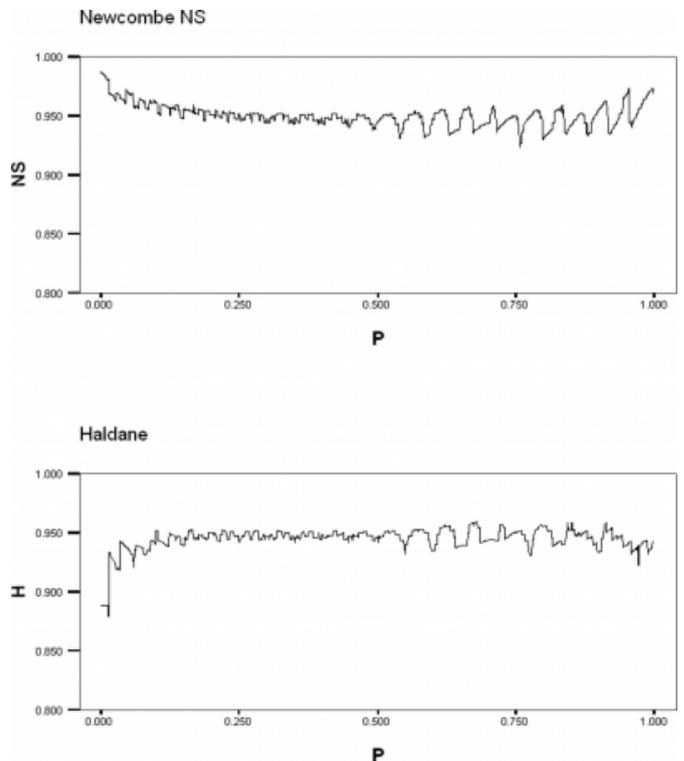


Figure 10

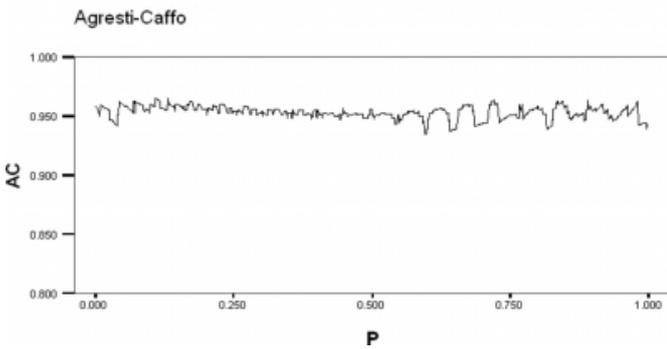
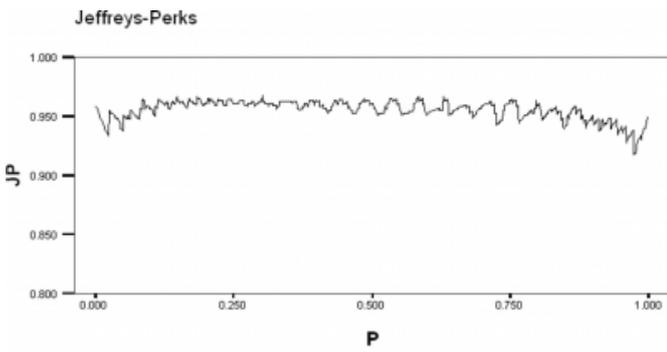


Figure 11

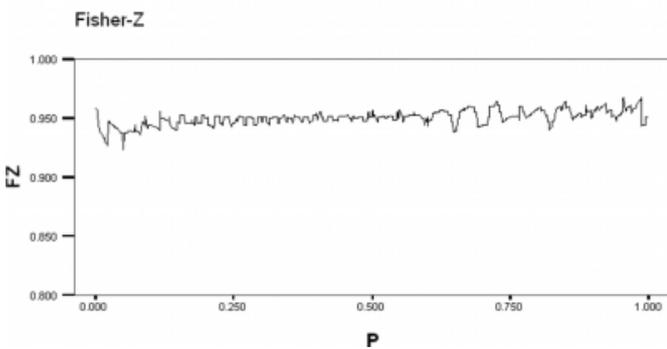
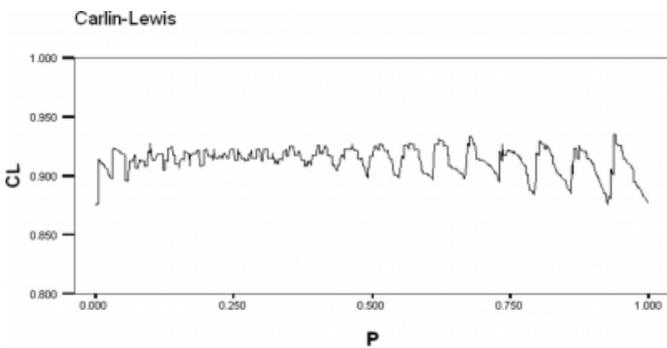


Figure 12

Table 2: Mean Coverage probabilities for nominal 95% Wald-z and Wald-c as a function of p when p=0.3 with n=n=20.

	N	Minimum	Maximum	Mean	Std Deviation
Wald-Z	999	.885	.948	.93169	.009865
Wald-C	999	.949	.981	.96886	.005363
Newcombe-NS	999	.925	.987	.95027	.009071
Haldane	999	.879	.959	.94483	.009359
Jeffreys-Perks	999	.918	.967	.95512	.008197
Agresti-Caffo	999	.935	.965	.95318	.005210
Carlin-Lewis	999	.876	.935	.91294	.011029
Fisher's-Z	999	.924	.967	.95013	.006149

Figure 4 shows the 95% confidence interval coverage probability function for the Newcombe NS, Haldane, Jeffreys-Perks, Agresti-Caffo, Carlin-Lewis and Fisher's Z methods as a function of  $\pi_1$ ,  $\pi_1 \in [0,1]$  for  $n_1 = 20$ ,  $n_2 = 10$  and  $p_2 = 0.2$ . In the unequal sample size situation, Newcombe NS and Agresti-Caffo coverage probability functions are near nominal over  $\pi_1 \in [0, 1]$ . The other methods are subnominal. Table 3 details the corresponding minimum, maximum, mean, and standard deviation for the coverage probability of these confidence interval methods. The Newcombe NS and Agresti-Caffo are nominal. Fisher's-z and the Jeffreys-Perks are equivalent while the Wald-z and Carlin-Lewis methods are subnominal.

Figure 13

Figure 4: Coverage probabilities for nominal 95% Newcombe NS, Haldane, Jeffreys-Perks, and Agresti-Caffo as a function of p when p=0.3 with n=20, n=10.

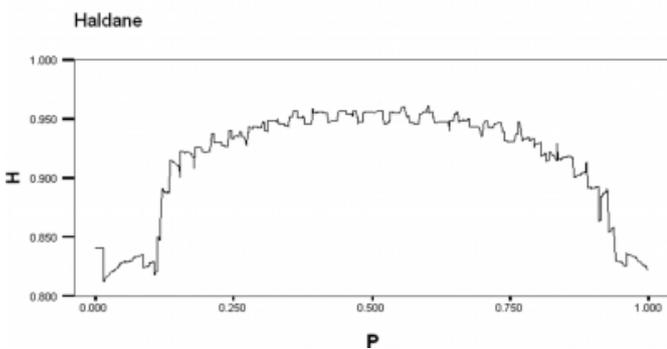
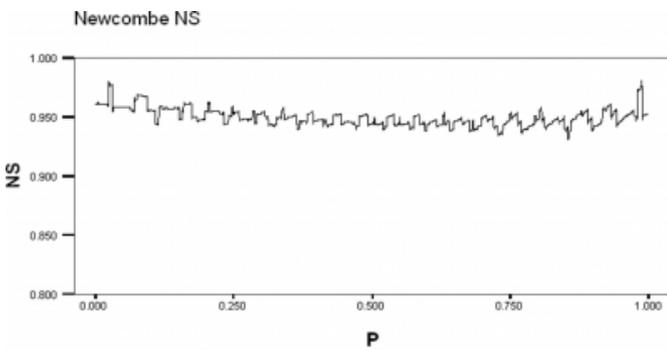


Figure 14

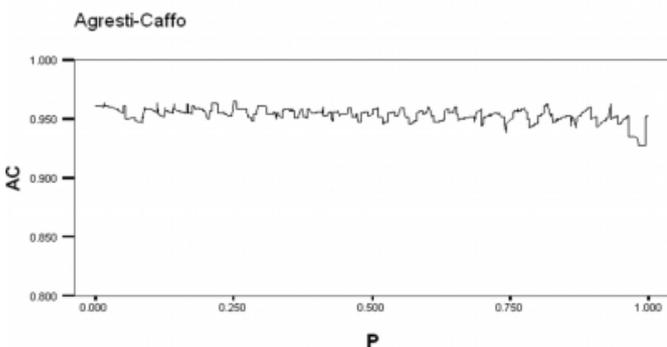
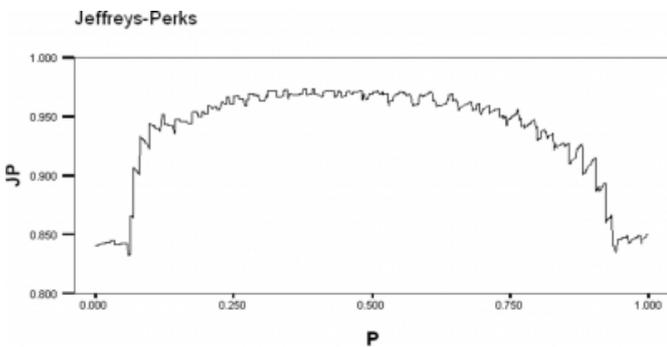


Figure 15

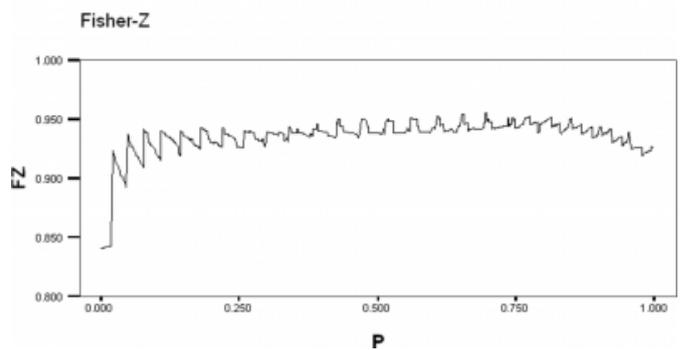
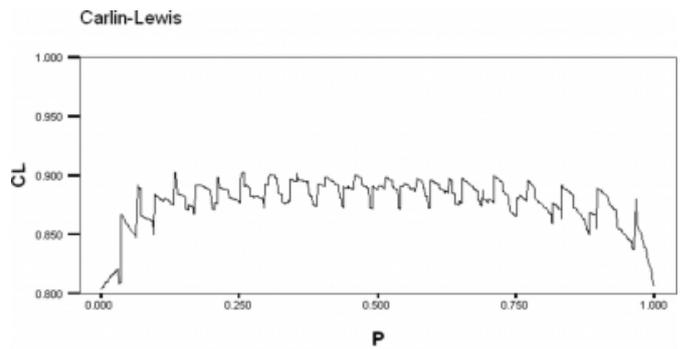


Figure 16

Table 3: Mean coverage probabilities for nominal 95% Wald-z and Wald-c as a function of p when p=0.3 with n=20, n=10.

	N	Minimum	Maximum	Mean	Std Deviation
Wald -Z	999	.806	.930	.90969	.017593
Wald-C	999	.932	.969	.96005	.005835
Newcombe-NS	999	.933	.978	.94973	.007131
Haldane	999	.812	.961	.91620	.045251
Jefferys-Perks	999	.832	.974	.93748	.042093
Agresti-Caffo	999	.928	.965	.95334	.005743
Carlin-Lewis	999	.804	.903	.87777	.019828
Fisher's-Z	999	.840	.955	.93529	.015990

DISCUSSION

The popularity of the Wald-z closed form confidence interval is based in part because of its overwhelming use in elementary statistics textbooks. The similarity of the computational formula to the confidence interval for the difference between two sample means is simplicity in action. Rarely do these elementary textbooks warn of the inadequacy of the Wald-z, Wald-c or even of a Wald-t. In the case of differences between two independent proportions the Wald-z confidence interval behaves so poorly with coverage probabilities below nominal values that it would seem that authors of these texts would at least consider other closed form methods which are better. Considering the coverage probability criterion, three alternative methods

demonstrate superior coverage properties and both are easily programmable. Our recommendation is in the equal  $n$  or nearly equal  $n$  case, use the Newcombe NS, Agresti-Caffo or Fisher's-Z methods. In the unequal  $n$  situation, use either the Newcombe NS or Agresti-Caffo methods.

### **References**

1. Agresti A and Coull BA. Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician* 1998;52:119-126.
2. Agresti A and Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician* 2000;54:280-288.
3. Beal SL. Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics* 1987;43:941-950.
4. Carlin BP and Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London, 1996.
5. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998;17:873-890.
6. Radhakrishna S, Murthy BN, Nair NGK, Jayabal P, and Jayasri R. Confidence intervals in medical research. *Indian J Med Res* 1992;96:199-205.
7. **Zou G and Donner A. A simple alternative confidence interval for the difference between two proportions. *Controlled Clinical Trials* 2004;25:3-12.**

**Author Information**

**James F. Reed III, Ph.D.**

Interim Chief of Health Studies and Director of Research, Lehigh Valley Hospital and Health Network