A NON-ITERATIVE ACCURATE ASYMPTOTIC CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

SYLVAN WALLENSTEIN
Biomathematical Sciences Department, Box 1023, Mount Sinai Medical Center, 1 Gustave Levy Place, New York, NY 10029, U.S.A.

SUMMARY
I propose a new confidence interval for the difference between two binomial probabilities that requires only the solution of a quadratic equation. The procedure is based on estimating the variance of the observed difference at the boundaries of the confidence interval, and uses least squares estimation rather than maximum likelihood as previously suggested. The proposed procedure is non-iterative, agrees with the conventional test of equality of two binomial probabilities, and, even for fairly small sample sizes, appears to yield actual 95 per cent confidence intervals with mean or median probabilities of coverage very close to 0·95. The Yates continuity correction appears to generate confidence intervals with the conditional probability of coverage at least equal to nominal levels. © 1997 by John Wiley & Sons, Ltd.

1. INTRODUCTION
It has become increasingly common to report confidence intervals, in addition to, or in replacement of, tests of statistical significance. Frequently, results are reported for the comparison of proportions in two groups that utilize as a measure of the disparateness of the two groups, either the odds ratio, relative risk or difference in proportions. Confidence intervals based on application of the central limit theorem involve calculation of the standard errors of these indices, or perhaps transformations of the indices, that depend on the true probabilities in the two groups. What we call (elementary) textbook confidence intervals simply replace the parameters by their observed values. More accurate confidence intervals replace the unknown parameters by their value at the boundary of the confidence intervals. Unlike the case for odds ratios or relative risks, asymptotic confidence intervals for differences in proportions, other than the ‘textbook’ variety, are to the best of my knowledge not implemented in computer programs, nor used by applied biostatisticians. Possible reasons are that the confidence intervals in the literature are either just slightly too difficult for simple implementation, or they are too ad hoc in nature. Many suggested procedures give confidence intervals that do not necessarily agree with results of tests of
significance for the equality of the two probabilities based on asymptotic theory. The continuity correction has often been explored in terms of its increase in the probability of coverage of the textbook solution rather than as a method to control probability of coverage of the confidence interval conditional on the total number of successes. I give in this paper a non-iterative procedure to compute confidence intervals, which shows better agreement with nominal probability of coverage than apparently does any other non-iterative procedure, and I explore the use of the continuity correction in achieving nominal coverage conditional on the total number of successes.

Formally, let \(X_1\) and \(X_2\) be two statistically independent binomial random variables with parameters \(n_1, \pi_1\), and \(n_2, \pi_2\), respectively. For a specific realization of the trial, let \(x_i\) be the observed number of successes in the \(i\)th group, and denote the observed proportion of success by \(\hat{p}_i = x_i/n_i\). Letting

\[
\text{var}(\hat{p}_1 - \hat{p}_2) = V(\pi_1, \pi_2) = \pi_1(1 - \pi_1)/n_1 + \pi_2(1 - \pi_2)/n_2
\]

the \((1 - \alpha)\) asymptotic confidence interval for \(\delta = \pi_1 - \pi_2\) is

\[
\hat{p}_1 - \hat{p}_2 \pm Z_{1-\alpha/2} \sqrt{V(\pi_1, \pi_2)}
\]

where \(Z_{\gamma}\) is the \(\gamma\)th percentile of the standardized normal distribution, and \(\hat{V}(\pi_1, \pi_2)\) is some estimate of \(V(\pi_1, \pi_2)\) based on the observed data. The simplest ‘textbook’ confidence interval that replaces \(\hat{V}(\pi_1, \pi_2)\) by \(V(\hat{p}_1, \hat{p}_2)\) has been shown\(^{2,3}\) often to be too short, and there have been various procedures suggested that give more reasonable confidence intervals.

Denote the lower and upper bounds of the confidence interval on \(\delta\) by \(d(x_1, x_2)\) and \(d(x_1, x_2)\), respectively, or simply ignoring all the arguments, as \((d', d)\). Let \(\hat{p}_{1|d}\) be an estimate of \(\pi_1\) based on assuming \(\delta = d\). Several authors\(^{3-6}\) note that \(d(x_1, x_2)\) is the solution of the implicit equation

\[
d = \hat{p}_1 - \hat{p}_2 + Z_{1-\alpha/2} \sqrt{V(\hat{p}_{1|d}, \hat{p}_{2|d})}.
\]

Beal\(^3\) sets \(\hat{p}_{1|d} + \hat{p}_{2|d} = \hat{p}_1 + \hat{p}_2\), allowing solution of \((3)\) explicitly. However, he suggests a different shrinkage-type procedure to remedy the observation that actual coverage of the resulting interval is, at times, appreciably less than the nominal 95 per cent coverage. Peskun\(^4\) does not estimate \(V(\pi_1, \pi_2)\), but rather finds the maximum value of \(d\), that satisfies \((3)\) (with the addition of a continuity correction) subject to the constraint \(\hat{p}_{1|d} - \hat{p}_{2|d} = d\), which tends to result in values of \(\hat{p}_{1|d}\) and \(\hat{p}_{2|d}\) centred around 0·5. His computation of exact probabilities of coverage suggests that the minimum probability of coverage of the resulting interval, with \(\pi_1\) and \(\pi_2\) taking any value on \([0, 1]\), is close to the nominal value. The actual probability of coverage for this interval has, however, apparently not had previous investigation. In the next section, I evaluate two procedures that solve \((3)\) implicitly based on maximum likelihood estimation, and I suggest a simpler approach that yields an explicit solution.

2. SOLVING THE IMPLICIT EQUATION FOR THE DIFFERENCE

Mee\(^5\) proposed to solve \((3)\) using a doubly iterative procedure. The first step fixes \(\delta = d\), and finds \(\hat{p}_{1|d}\) and \(\hat{p}_{2|d}\) that maximize the likelihood, while the second step updates \(d\) to satisfy \((3)\) more closely. Miettinen and Nurminen\(^6\) point out that one need not perform iteration to find the maximum likelihood estimates since they are solutions of a cubic, and thus one needs only a singly iterative procedure.

I propose a slight modification of Mee’s procedure in that, (i) I simplify the computational procedure by using least squares instead of maximum likelihood estimates for \(\hat{p}_{1|d}\) and \(\hat{p}_{2|d}\), and (ii) I generalize the problem slightly to allow a continuity correction, \(\varepsilon \geq 0\), by replacing \(\hat{p}_1 - \hat{p}_2\)
in (3) by \((\hat{p}_1 - \hat{p}_2 + \varepsilon)\). The least squares estimates of \(\hat{p}_{1d}\) and \(\hat{p}_{2d}\) in (3), subject to the constraint \(\hat{p}_{1d} - \hat{p}_{2d} = d\) are

\[
\hat{p}_{1d} = \bar{p} + d \frac{n_2}{N} \\
\hat{p}_{2d} = \bar{p} - d \frac{n_1}{N}
\]

(4)

where \(\bar{p} = (n_1 \hat{p}_1 + n_2 \hat{p}_2)/N\), and \(N = n_1 + n_2\). Substituting (4) into (3) and dropping the subscript on \(Z\), the preliminary value of \(d(x_1, x_2)\) is the larger solution of

\[
n_1n_2[(d - (\hat{p}_1 - \hat{p}_2 + \varepsilon))Z]^2 = n_2(\bar{p} + dn_2/N)(1 - \bar{p} - dn_2/N) + n_1(\bar{p} - dn_1/N)(1 - \bar{p} + dn_1/N).
\]

Alternatively, \(d\) is the larger solution of the quadratic \(ad^2 + bd + c = 0\), where

\[
a = 1 + N^{-1}Z^2(1 + (n_1 - n_2)^2/(n_1n_2))
\]

\[
b = -2[\hat{p}_1 - \hat{p}_2 + \varepsilon] + Z^2(1 - 2\bar{p})(n_1 - n_2)/n_1n_2
\]

\[
c = [\hat{p}_1 - \hat{p}_2 + \varepsilon]^2 - Z^2N\bar{p}(1 - \bar{p})/n_1n_2.
\]

(5)

When using the continuity correction, one should replace the expression in brackets by 1.0 if it exceeds 1.0.

To compute \(d'\), the lower bound of the confidence interval, replace \(\varepsilon\) by \(-\varepsilon\) in (5), with use of the same care concerning the term in brackets, and take the smaller of the solutions.

If \(\varepsilon = 0\), the upper and lower bounds on the confidence interval are the solutions of a single quadratic equation. For the special case of \(n_1 = n_2 = m\), the preliminary confidence interval on \(\delta\), \((d', d)\) is

\[
\frac{(\hat{p}_1 - \hat{p}_2) \pm (Z/\sqrt{2})\{(\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2) + Z^2\bar{p}(1 - \bar{p})/m\}^{1/2}}{1 + Z^2/2m}
\]

in agreement with Beal.

Note that for any solution to (3),

\[
V(\hat{p}_{1d}, \hat{p}_{2d}) = \text{var}(\hat{p}_1 - \hat{p}_2|\pi_1 = \pi_2) = V(\bar{p}, \bar{p}) = N\bar{p}(1 - \bar{p})/n_1n_2
\]

so that the \((1 - \alpha)\) confidence interval includes zero if the test based on the standard test statistic \((\hat{p}_1 - \hat{p}_2)/\sqrt{V(\bar{p}, \bar{p})}\) does not reject \(H_0: \delta = 0\), at level \(\alpha\). A similar comment holds for the continuity-corrected test statistic.

The procedure described above is valid only if substituting the solution of the quadratic in equation (5) into (4), satisfies \(0 \leq \hat{p}_{1d}, \hat{p}_{2d}, \hat{p}_{1id}, \hat{p}_{2id} \leq 1\). When computing 95 per cent confidence intervals, this constraint is often violated when the minimum expected sample size is less than two. In such cases, I propose a somewhat ad hoc solution, in which one replaces the estimate outside \([0, 1]\) with 0 or 1. For example, if \(\hat{p}_1 \geq \hat{p}_2\) and \(\hat{p}_{2id} < 0\), force \(\hat{p}_{2id} = 0\), and write the generalization of equation (3) as

\[
d - (\hat{p}_1 - \hat{p}_2 + \varepsilon) = Z\sqrt{(d(1 - d)/n_1)]}
\]

so that

\[
d = \frac{\hat{p}_1 - \hat{p}_2 + \varepsilon + Z^2/2n_1 + (Z/\sqrt{n_1})\{[\hat{p}_1 - \hat{p}_2 + \varepsilon][1 - (\hat{p}_1 - \hat{p}_2 + \varepsilon)] + Z^2/4n_1\}^{0.5}}{1 + Z^2/n_1}
\]

(6)

If \(\hat{p}_1 \geq \hat{p}_2\) and \(\hat{p}_{1id} > 1\), set \(\hat{p}_{1id} = 1, \hat{p}_{2id} = 1 - d\), so that a solution of the new quadratic is the same as (6) with \(n_1\) replaced by \(n_2\). If \(\hat{p}_1 \leq \hat{p}_2\), and \(\hat{p}_{1id} < 0\), which implies \(d' < 0\), solve (3) by setting \(\hat{p}_{1id} = 0, \hat{p}_{2id} = d'\). If \(\hat{p}_2 > \hat{p}_1\), the above logic is somewhat reversed, and it may be easier simply to reverse indices.
3. COMPARISON WITH OTHER ASYMPTOTIC PROCEDURES

Following the general approach suggested by Hauck and Anderson, and Peskun, I compare the coverage of the various 95 per cent confidence intervals by fixing \( n_1, n_2, \pi_1 \) and \( \pi_2 \), computing the 95 per cent confidence interval, \((d'(x_1, x_2), d(x_1, x_2))\), for each \( x_1 = 0, \ldots, n_i \), \( i = 1, 2 \), and then finding the exact probability of coverage

\[
PC(\pi_1, \pi_2) = \Sigma I\{d'(x_1, x_2) \leq \pi_1 - \pi_2 \leq d(x_1, x_2)\} \Pr(X_1 = x_1, X_2 = x_2|n_1, n_2, \pi_1, \pi_2)
\]  

(7)

where the summation is over all \((x_1, x_2)\), and \( I[s] = 1 \) if \( s \) is true, and is 0 otherwise.

I compared the probability of coverage of (a) the non-iterative 95 per cent confidence interval proposed here, (b) the standard ‘textbook’ 95 per cent confidence interval, (c) the iterative procedure suggested by Miettinen and Nurminen, (d) Peskun’s procedure, and (e) a recommended procedure of Hauck and Anderson described below. Comparisons were performed for \( E = \min\{n_1\pi_1, n_1(1 - \pi_1), n_2\pi_2, n_2(1 - \pi_2)\} \geq 2 \). The values of \( n_1, n_2, \pi_1, \) and \( \pi_2 \) have previously been used in the literature, specifically:

(i) A series of 34 cases from Tables 1 and 2 of Beal with \( n_1 = n_2 = (5, 15, 25, 35, 45) \), and \((\pi_1, \pi_2) = (0.05 \times (0.01, 0.05), 0.1 \times (0.1, 0.3, 0.5, 0.7, 0.9), 0.3 \times (0.3, 0.5, 0.7), 0.5 \times (0.01, 0.05, 0.5);

(ii) A series of 379 cases, mostly with \( n_1 \neq n_2 \), studied by Hauck and Anderson generally obtained by cross-tabulating values of \( E \) ranging from 2 to 15, values of \( n_1 + n_2 \) ranging from 20 to 100, \( \pi_1 \) from 0.05 to 0.5, and \( \pi_2 \) from 0.05 to 0.95. (To be precise, I studied all pairs enumerated by Hauck and Anderson, but deleted the cases with \( n_1 + n_2 = 200 \), and added back cases with \( E = 2 \).)

Table I lists the minimum, 10th, 25th, 50th, 75th and 90th percentile, and the maximum probability of coverage of various nominal 95 per cent confidence intervals, for both series. I summarize the information by AVE\{PC(\pi_1, \pi_2)\} and by \( \sqrt{\text{AVE}\{PC(\pi_1, \pi_2) - 0.95\}^2} \), though one should note that due to the skewness, the indices penalize intervals that underestimate, rather than overestimate, coverage.

As previously noted,\(^2\,\,^3\) the standard textbook procedure considerably overestimates the probability of coverage. The interval of Peskun, while having a minimum probability of coverage fairly close to the nominal value as claimed, yields intervals with probability of coverage considerably above 95 per cent. There was no evidence of a pronounced deterioration in coverage by using the non-iterative procedure proposed here as opposed to the iterative procedure of Miettinen and Nurminen.

A similar evaluation, not shown here but available upon request, demonstrates that my procedure is superior to the ‘Haldane’ procedure suggested by Beal whenever either \( n_1 \neq n_2 \) or when \( E < 2 \). Specifically, for both the textbook method and the method suggested by Beal, the minimum probability of coverage for the 31 pairs of \((n_1, n_2, \pi_1, \pi_2)\) that were removed from data set (i) because \( E < 2 \) was 0.26, while the tenth percentile was less than 0.74. In contrast, the minimum and tenth percentile for the method suggested here were 0.895 and 0.926, respectively. In a series of 27 cases generated for \( E > 2 \) but \( n_1 \neq n_2 \), the tenth percentiles of the probability of coverage of nominal 95 per cent confidence intervals was 0.91 for the method of Beal, as contrasted with 0.94 for the method suggested here.

As \( n_1 + n_2 \) becomes large, and \( \pi_1 \) and \( \pi_2 \) remain bounded away from 0 and 1, \( V(\hat{\theta}_1, \hat{\theta}_2) \) becomes closer to \( V(\hat{\theta}_1, \hat{\theta}_2) \), so the method proposed here becomes closer to the textbook method. However, the method due to Peskun apparently converges much more slowly. For example, for \( n_1 = n_2 = 150, \pi_1 = 0.25, \pi_2 = 0.15 \), the probability of coverage was 0.950 for the method
Table I. Exact probabilities of coverage of various nominal 95 per cent intervals

<table>
<thead>
<tr>
<th>Series</th>
<th>Method</th>
<th>Min</th>
<th>10th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>90th</th>
<th>Max</th>
<th>Mean</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>W</td>
<td>0.930</td>
<td>0.936</td>
<td>0.943</td>
<td>0.947</td>
<td>0.951</td>
<td>0.966</td>
<td>0.947</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>0.932</td>
<td>0.942</td>
<td>0.945</td>
<td>0.950</td>
<td>0.954</td>
<td>0.958</td>
<td>0.960</td>
<td>0.949</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>0.930</td>
<td>0.941</td>
<td>0.956</td>
<td>0.966</td>
<td>0.979</td>
<td>0.998</td>
<td>1.0</td>
<td>0.967</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.879</td>
<td>0.912</td>
<td>0.935</td>
<td>0.942</td>
<td>0.944</td>
<td>0.949</td>
<td>0.952</td>
<td>0.936</td>
<td>0.021</td>
</tr>
<tr>
<td>H</td>
<td>W</td>
<td>0.934</td>
<td>0.944</td>
<td>0.947</td>
<td>0.950</td>
<td>0.954</td>
<td>0.963</td>
<td>0.974</td>
<td>0.952</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>HA</td>
<td>0.877</td>
<td>0.914</td>
<td>0.954</td>
<td>0.961</td>
<td>0.964</td>
<td>0.977</td>
<td>0.980</td>
<td>0.953</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>0.936</td>
<td>0.949</td>
<td>0.955</td>
<td>0.967</td>
<td>0.979</td>
<td>0.991</td>
<td>1.0</td>
<td>0.968</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.784</td>
<td>0.874</td>
<td>0.892</td>
<td>0.922</td>
<td>0.935</td>
<td>0.942</td>
<td>0.951</td>
<td>0.912</td>
<td>0.048</td>
</tr>
</tbody>
</table>

$s = \sqrt{\text{AVE}}$ (probability of coverage $- 0.95)^2$

**Series**
- B: subset of 34 cases from a series of Beal
- H: subset of 379 cases from series of Hauck and Anderson

**Methods**
- W: series described in this paper
- M: Method of Miettinen and Nurminen as cited by Beal
- HA: Method 3 of Hauck and Anderson
- P: Method of Peskun
- T: Textbook method

suggested here, 0.949 for the textbook method and 0.932 for the method of Peskun. A very limited comparison indicates that the textbook method appears adequate for $n_1 + n_2 > 100$ and $E > 10$, and that perhaps this criteria could be loosened somewhat.

### 3.1. Coverage Based On The Continuity Correction

One can justify the use of a continuity correction with two different approaches. Numerical justifications might be based simply on the fact that, as noted in Table I, the typical textbook confidence interval is too small, and a continuity correction of order $1/n_i$ increases the length of the confidence interval while insuring that it does not become ‘too large’. If one adopted this approach, slightly different ‘continuity corrections’ are possible. Hauck and Anderson examine various such confidence intervals and recommend the interval obtained by setting 

$$
\varepsilon = 0.5\min(1/n_1, 1/n_2), \text{ and by replacing } \pi_i \text{ by } \hat{p}_i, \text{ and } n_i \text{ by } n_i - 1 \text{ in (1).}
$$

Table I suggests that the continuity correction is not needed to improve the probability of coverage for the method suggested here. Rather, one should use the continuity correction if one accepts the argument\(^7,8\) that for $2 \times 2$ tables, conditional tests are preferred, and that tests based on discrete distributions are inherently conservative. The conditional probability of coverage for a marginal $x_1 + x_2 = M$, computed in a manner analogous to (7) is

$$
\text{CPC}(M; \pi_1, \pi_2) = \sum \text{I} \left( x_1, x_2 \right) \leq \pi_1 - \pi_2
$$

$$
\leq \text{d}(x_1, x_2) \Pr (X_1 = x_1, X_2 = x_2 | X_1 + X_2 = M, n_1, n_2, \pi_1, \pi_2)
$$

where the sum is over $x_1 + x_2 = M$.  

\(^7\text{A. Beal}, \text{Journal of the Royal Statistical Society, Series B 56 (1994), 1-20.}\)  
\(^8\text{P. Peskun}, \text{Biostatistics 2 (1990), 533-539.}\)
I illustrate the conditional probability of coverage for the case \((\pi_1, \pi_2) = (0.3, 0.5)\) with \((n_1, n_2) = (25, 25)\) or \((45, 20)\). Attention is restricted in the first case to the marginals from \(M = 9\) to \(M = 41\), and in the second case to \(M\) from 11 to 47. For these marginals, \(\Sigma \Pr(X_1 + X_2 = M) \geq 0.9998\), and for each excluded marginal \(M'\), \(\Pr(X_1 + X_2 = M')/\max \Pr(X_1 + X_2 = j) \leq 0.0011\). I used the Yates continuity correction. The minimum conditional probability of coverage, over the range of \(M\) evaluated, is 0.944 for the first case, and 0.947 for the second case. Since the unconditional probability of coverage is

\[
PC(\pi_1, \pi_2) = \frac{\Pr(X_1 + X_2 = M) \cdot CPC(M; \pi_1, \pi_2)}{\max \Pr(X_1 + X_2 = j)}
\]

the restriction CPC(\(M\)) \(\geq 1 - \alpha\) will, especially for small or moderate sample sizes, result in PC(\(\pi_1, \pi_2\)) being appreciably more than the nominal value. In the examples above, the unconditional probabilities of coverage for the confidence intervals with the continuity correction were 0.973 and 0.975, as contrasted to 0.944 and 0.949 for the intervals without the continuity correction.

3.2. Exact Confidence Intervals

Recently, there has been considerable activity on finding exact confidence intervals (based on the binomial distribution rather than the normal approximation), with Santner and Snell being a good starting point for reference. As noted by Coe and Tamhane, exact \((1 - \alpha)\) confidence intervals on \(\delta\) must satisfy PC(\(\pi_1, \pi_2\)) \(\geq 1 - \alpha\) for all \(\pi_1, \pi_2\). They also list several other desirable properties: that the interval be invariant with respect to labelling of populations and of success; that it be monotone in a certain sense, and that it should be short in some overall sense. They note that the last property is difficult to insure formally because it involves solving a highly complicated discrete optimization problem, and apparently for this reason, different exact procedures can give different results.

Coe and Tamhane and Santner and Yamagami have recently developed algorithms satisfying the properties listed above, with the former algorithm apparently giving shorter intervals over most of the sample space, but the latter being appreciably faster. Both these algorithms, as well as one by Soms, are available as FORTRAN programs, but appear to be somewhat difficult to implement for \(n_i > 20\). Recently a commercially available program, StatXact3, was introduced that applies a technique of Berger and Boos to confidence intervals, so that part of the \(\alpha\) error can be used to exclude unlikely values of the nuisance parameter, \(\pi_1\).

All these exact procedures are unconditional in nature, and the documentation of StatXact notes that there exists no conditional exact confidence interval, since there is no sufficient statistic. These exact confidence intervals thus agree with neither Fisher’s exact test, nor the asymptotic test statistic, \((\hat{p}_1 - \hat{p}_2)/\sqrt{V(\hat{p}, \hat{\pi})}\). Soms’ procedure appears to agree with an exact test based on \((\hat{p}_1 - \hat{p}_2)/\sqrt{V(\hat{p}_1, \hat{p}_2)}\) described by Suissa and Shuster, while StatXact defines a new test statistic that is consistent with the confidence interval. Based on definition and construction (in particular by the requirement PC(\(\pi_1, \pi_2\)) \(\geq 1 - \alpha\) for all or most \(\pi_1\) and \(\pi_2\)), these exact unconditional intervals are best approximated by the asymptotic confidence interval proposed here utilizing the Yates continuity correction.

4. EXAMPLES

Campbell’s Letter to the Editor of a medical journal stresses the importance of constructing confidence intervals to interpret results, and appears to advocate the procedure of Peskun,
although Campbell apparently (but erroneously) finds it too complex. The data in the letter are based on a trial\textsuperscript{18} that reported a response rate of 4/16 for the low dose of an anti-cancer agent, and 0/15 for the high dose. Ignoring the continuity correction, the coefficients of the quadratic equation in (5) are $a = 1.1244$, $b = -0.48812$ and $c = 0.00673$. The preliminary bounds for the confidence interval are $d' = 0.014$ and $d = 0.420$. Since $\hat{p}_{2(0.42)} = 0.129 - 0.42(16/31) = -0.088$, one needs to reset $\hat{p}_{2id} = 0$, and replace the preliminary upper bound with the value in (6), given by 0.495. Thus, the proposed 95 per cent confidence interval is (0.014, 0.495), nearly identical to the interval (0.016, 0.499) based on Miettinen and Nurminen’s procedure. (Their iterative procedure did not require a second step to remedy boundary conditions, but did also result in $\hat{p}_{2(0.499)} = 0$.)

The ‘textbook’ 95 per cent confidence interval for the difference is $0.25 \pm 1.96 \sqrt{0.0117} = (0.038, 0.462)$, which is shorter than my interval. The interval given by Peskun, without use of the continuity correction, is $(-0.10, 0.545)$ which is too large and also does not agree with the test of statistical significance. (The corrected interval might be only very slightly larger due to a novel approach for the continuity correction when $n_1 \neq n_2$.) Hauck and Anderson’s interval ($-0.002, 0.502$) is numerically close to the one proposed here, but does not agree with the conclusion concerning statistical significance.

The lower bound for continuity corrected confidence interval is constructed by replacing $\varepsilon$ in (5) by $-0.5(1/15 + 1/16) = -0.0646$, so that $d' = -0.0051$. Computation of the upper bound will again require boundary conditions and would be given by (6), with $\varepsilon = 0.0646$ and $d = 0.558$. This part of the example is for illustrative purposes, as one would hesitate to apply asymptotic results based on $M = 4$.

An example prominently cited in the promotion for StatXact concerns a legal case dealing with racial discrimination where $n_1 = 379$, $n_2 = 6$, $x_1 = 379$, $x_2 = 1$. The textbook confidence interval on $\delta$ cited in the promotion is (0.535, 1.13) (the upper bound truncated to 1.0), while the exact confidence interval cited is (0.359, 0.996). Applying the methods here with the Yates continuity correction, $\varepsilon = 0.0847$, yields an asymptotic confidence interval of (0.362, 0.992). In this example, the choice of asymptotic intervals is much more important than whether an asymptotic or exact procedure is used.

5. CONCLUSIONS

With an objective, as apparently is the case in many of the papers cited,\textsuperscript{1, 3, 5, 6} to construct a confidence interval that yields unconditional probability of coverage close to nominal levels, the method here without the continuity correction appears clearly preferable. It is non-iterative, it agrees with the conventional test of equality of proportions, and it yields actual 95 per cent confidence intervals with mean or median probabilities of coverage very close to 0.95. The limited work presented here suggests that the Yates continuity correction usually generates confidence intervals with the conditional probability of coverage at least equal to the nominal levels. Asymptotic confidence intervals with the Yates continuity correction are appropriate if one is particularly sensitive about minimum probability of coverage of unconditional intervals, or wants to mimic exact confidence intervals.

It is important to check that estimates termed $p_{lid}$ and $p_{lid'}$, $i = 1, 2$ fall in the interval $[0, 1]$. I give an alternative, somewhat ad hoc, confidence interval in case this condition is not met, and my numerical results suggest reasonably good agreement with nominal coverage, even in these cases. However, especially if the estimates are well outside $[0, 1]$, it would be reasonable in these cases to question the appropriateness of asymptotic theory and utilize exact methods. If, as in the legal discrimination example, only the lower bound is of interest, it can be used even if the upper bound is associated with values of $\hat{p}_{1id}$, $\hat{p}_{2id}$ outside of $[0, 1]$.
I thank Nathan Mantel for his comments, several years ago, concerning the need for a continuity correction, and the reviewers for their helpful comments. Research was supported by NIH grants funding Clinical Research Centers and Environmental Health Sciences Centers.

REFERENCES