

On Finding the Upper Confidence Limit for a Binomial Proportion when Zero Successes are Observed

Courtney E McCracken¹ and Stephen W Looney^{2*}

¹Department of Pediatrics, Emory University, Atlanta, USA

²Department of Biostatistics and Epidemiology, Augusta University, Medical College of Georgia, USA

Abstract

We consider confidence interval estimation for a binomial proportion when the data have already been observed and x , the observed number of successes in a sample of size n , is zero. In this case, the main objective of the investigator is usually to obtain a reasonable upper bound for the true probability of success, i.e., the upper limit of a one-sided confidence interval. In this article, we use observed interval length and p -confidence to evaluate eight methods for finding the upper limit of a confidence interval for a binomial proportion when x is known to be zero. Long-run properties such as expected interval length and coverage probability are not applicable because the sample data have already been observed. We show that many popular approximate methods that are known to have good long-run properties in the general setting perform poorly when $x=0$ and recommend that the Clopper-Pearson exact method be used instead.

Keywords: p -Confidence; Exact methods; Interval length; One-sided interval; Sample size

Introduction

This study was motivated when one of the authors was approached by a clinical investigator in the dental school who was seeking advice on how to design a research study consisting of a sequence of independent trials, each with only two possible outcomes (“success” and “failure”). Based on a previously published study he had conducted [1], as well as his clinical judgment and experience, the investigator had strong reason to believe that the observed number of “failures” in the study he was planning would likely be zero. His question for us was, “Assuming that there will be no failures in my study, how many trials do I need to conduct so that I can be reasonably sure that the true probability of failure is no greater than .05”? Since the trials can be assumed to be independent, it is reasonable to assume that X =the number of “failures” in the clinical study follows a binomial distribution, with n =the number of independent trials and π =the probability of “failure” on any one trial.

Many research studies in clinical areas and other applied fields meet the assumptions of the binomial distribution (Without loss of generality, we will refer to the two outcomes as “success” and “failure” and the outcome of primary interest as “success”). Furthermore, it is not uncommon in these studies for x , the observed number of successes in a sample of size n , to be zero. Examples can be found [2-6].

Only a few publications in the statistical literature have examined or compared methods for analyzing binomial data in which the observed number of successes is zero; for example, see [7-12]. Observing $X=0$ in a binomial sample lends itself to the Bayesian approach since one can condition on the observed data, and several authors have considered this approach [13-16]. In the present article, we approach the problem from a frequentist point of view; however, we also include a Bayesian approach for comparison purposes.

We restrict our attention to the situation in which one is interested only in finding the one-sided (upper) confidence limit for the true value of the binomial proportion when the number of successes has already been observed to be zero. The use of one-sided confidence limits in this situation is controversial [8,10], but some authors have recommended their use [17,18]. Following our discussions with the clinical investigator, we decided that a one-sided upper confidence limit

would be appropriate. Our review of the relevant literature indicated that there was no clear consensus on the best upper confidence limit to use when the observed number of successes is zero, especially if one wished to approach the problem from a frequentist perspective. Hence, in order to be able to provide a well-informed response to the dental researcher's question, we concluded that it would be worthwhile to systematically compare the most widely-used binomial confidence interval (C.I.) methods under the assumption that $x=0$.

In Section 2, we describe each of the methods for finding confidence limits for a binomial proportion that we compare; in Section 3, we describe our methodology for comparing the performance of these methods; in Section 4, we provide a summary of the results of our comparisons; and in Section 5, we discuss our results and make recommendations concerning the best methods to use in practice.

Methods

In this section, we describe the eight methods we compared for finding π_u , the upper 100 (1- α)% confidence limit for the true probability of success for a binomial random variable, under the assumption that the number of successes has already been observed to be 0. We selected these methods either because (1) they are commonly covered in introductory statistics or biostatistics textbooks, or (2) they have been recommended for general use when finding confidence limits for a binomial proportion.

One method we did not include is the Wald interval, which continues to be one of the most widely used methods for finding confidence limits for a binomial proportion. The Wald interval is based on the normal approximation to the binomial, and the approximate

***Corresponding author:** Looney SW, Department of Biostatistics and Epidemiology, Augusta University, Medical College of Georgia, Augusta, GA 30912, USA, Tel: (706) 721-4846; E-mail: slooney@augusta.edu

Received March 10, 2017; **Accepted** March 27, 2017; **Published** March 30, 2017

Citation: McCracken CE, Looney SW (2017) On Finding the Upper Confidence Limit for a Binomial Proportion when Zero Successes are Observed. J Biom Biostat 8: 338. doi:[10.4172/2155-6180.1000338](https://doi.org/10.4172/2155-6180.1000338)

Copyright: © 2017 McCracken CE, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

100 (1- α)% upper confidence limit for π for a one-sided interval is given by:

$$\pi_u = p + Z_{1-\alpha} \sqrt{p(1-p)/n},$$

where $p=x/n$ is the observed proportion of successes and $Z_{1-\alpha}$ is the standard normal deviate associated with an upper-tail area of α . However, when $p=0$, the Wald interval is (0, 0) for any confidence coefficient and is obviously uninformative. Nevertheless, such intervals are still reported in the scientific literature [2].

If one uses the normal approximation to the binomial with continuity correction, the upper confidence limit for π given by:

$$p + Z_{1-\alpha} \sqrt{p(1-p)/n} + [1/(2n)]$$

[19]. When $x=0$, this reduces to $\pi_u = 1/(2n)$, which is not a valid confidence limit because it does not depend on the confidence coefficient.

The Agresti-Coull method

Agresti and Coull [20] proposed a modification to the standard 95% Wald interval that consisted of adding four pseudo observations (two successes and two failures) to the sample. When $x=0$, their “adjusted Wald” approximate 95% upper confidence limit is given by:

$$\pi_u = [2/(n+4)] + 1.645 \sqrt{[2/(n+4)]\{1-[2/(n+4)]\}/(n+4)}.$$

Wilson’s “Score” interval

Wilson [21] proposed a method based on inverting the score test of the null hypothesis $H_0: \pi=\pi_0$. It is similar to the Wald interval except that the standard error is based on the hypothesized null value π_0 instead of the observed proportion of success p . The approximate 100 (1- α)% upper confidence limit is given by

$$\pi_u = Z_{1-\alpha}^2 / (n + Z_{1-\alpha}^2)$$

when zero successes are observed.

Continuity-corrected Wilson interval

Under some conditions, Wilson’s interval fails to maintain its pre-specified 1- α nominal coverage probability [22]. Therefore, Casella [22] recommended using the continuity-corrected normal approximation to the binomial when inverting the score test. When $x=0$, the approximate upper limit for a 100 (1- α)% C.I. is:

$$\pi_u = [0.5 + (Z_{1-\alpha}^2/2) + Z_{1-\alpha} \sqrt{0.5 - (0.25/n) + (Z_{1-\alpha}^2/4)}] / (n + Z_{1-\alpha}^2).$$

Clopper-Pearson “exact” (C-P) method

Clopper and Pearson [23] proposed what is usually called the “exact” method for constructing binomial confidence limits. This method is based on the inversion of the exact test of $H_0: \pi=\pi_0$ using the binomial distribution. Thus, the upper limit of an exact 100 (1- α)% one-sided interval is found by solving for π_U in the following equation:

$$\sum_{k=0}^x \{n!/[k!(n-k)!]\} \pi_U^k (1-\pi_U)^{n-k} = \alpha$$

When $x=0$ successes are observed, the exact method provides the following 100 (1- α)% upper confidence limit:

$$\pi_u = 1 - \alpha^{1/n}.$$

Louis [7] noted that $1 - \alpha^{1/n}$ is “exactly the 95 percent Bayesian prediction interval based on a uniform prior distribution for π .” In addition, the C-P upper limit is equivalent to the upper limit of the

modified Jeffreys interval proposed by Brown, Cai, and DasGupta [24] when $x=0$.

Mid-P adjusted Clopper-Pearson method

Some authors have recommended using the mid- p adjustment [25] to help overcome the possibly extreme conservatism of the Clopper-Pearson method, especially when n is small [26]. This approach involves inverting the mid- p adjusted exact test of $H_0: \pi=\pi_0$. The mid- p adjusted one-tailed p -value is obtained by subtracting half the point probability of the observed value of X from the appropriate one-tailed exact p -value. When $x=0$, the mid- p adjustment yields an upper 100 (1- α)% confidence limit of

$$\pi_u = 1 - (2\alpha)^{1/n}.$$

Poisson approximation

The binomial distribution with large n and small π can be approximated by the Poisson distribution with parameter $\lambda=n\pi$ and the resulting approximation can be used to derive approximate confidence limits for π [27]. For $x=0$, the Poisson approximation to the binomial yields an approximate upper 100 (1- α)% confidence limit of

$$\pi_u = -\ln(\alpha) / n.$$

The Poisson approximation for a 95% upper limit is equivalent to the Rule of 3, in which $\pi_u = 3/n$ [28].

The single augmentation with an imaginary failure or success (SAIFS) method

Borkowf [29] suggested an approach similar to the Agresti-Coull method. His method adds an imaginary failure to the sample when finding the lower confidence limit and an imaginary success when finding the upper confidence limit. When zero success occur in a sample of size n , the SAIFS method gives a one-sided upper confidence limit of

$$\pi_u = (1 + Z_{1-\alpha}) / (n + 1).$$

Borkowf showed that replacing the $Z_{1-\alpha}$ quantile with a $t_{1-\alpha}$ quantile with $n-1$ degrees of freedom could improve coverage probabilities, so that is the method we will consider here.

Bayes-Laplace HPD interval

The Bayes-Laplace (B-L) highest posterior density (HPD) credible interval for a binomial proportion is derived using the Bayesian approach with a non-informative Bayes-Laplace prior of beta (1, 1). Tuyl, Gerlach, and Mengerson [14] recommended the B-L prior as a consensus prior when $x=0$ is observed, and this approach yields the following 100 (1- α)% upper confidence limit:

$$\pi_u = 1 - \alpha^{1/(n+1)}.$$

Example:

We illustrate the methods described above using data from the clinical study that motivated this investigation [30]. Because dental devices emit electromagnetic energy, dental patients with implanted biotechnological devices can be at risk for adverse events relating to the functioning of their implants. The implant of interest in [30] was a neurostimulator (NS) device commonly used to relieve various types of pain, including spinal axial pain, ischemic limb pain, certain angular pain syndromes, and chronic diabetic neuropathic pain. The purpose of the study was to investigate the effects of electromagnetic interference on an NS during the operation of three dental devices: an

electric pulp tester, an apex locator, and an electrocautery unit. The investigator wished to estimate the proportion of experimental trials in which the dental device would damage an NS device located in different tissues. Based on his previously published study [1], in which he found that various dental devices had no impact on cochlear implants, the investigator had strong reason to believe that the observed number of “failures” in the NS study he was planning would also be zero. He wished to know the number of trials he would need to conduct to be 95% confident that the true probability of failure is no greater than $\pi_{\max} = 0.05$. A sample size calculation for the C-P method assuming that a one-sided confidence limit would be used in the data analysis yielded

$$n = \ln(\alpha) / \ln(1 - \pi_{\max}) = \ln(.05) / \ln(1 - .05) \approx 59.$$

The investigator chose to perform 70 independent trials in order to ensure that data from an adequate number of valid trials would be available. Once the study data were collected, no damage to the NS implant was observed in any of the $n=70$ trials. For purposes of illustration, we calculated the upper 95% confidence limit using these data for each of the eight methods described above (Table 1). In Table 1, we have also indicated the confidence coefficient of the C-P interval based on each of these upper limits. These coefficients range from 90.0% for the mid- p adjusted C-P upper limit to 98.5% for the Agresti-Coull upper limit.

Criteria for comparing the confidence interval methods

In this article, we use two criteria to evaluate the different methods for constructing upper one-sided binomial confidence limits when x is known to be zero: observed interval length and p -confidence [31]. Properties such as coverage probability and expected interval length that are typically used to evaluate long-run performance of C.I. procedures are not applicable for our study because we consider only the situation when the data have already been observed and the value of x is known to be zero. Therefore, the expected values required for coverage probability and expected interval length cannot be calculated.

Observed interval length

Observed interval length is calculated by subtracting the lower confidence limit of a C.I. from the upper confidence limit. Since we are only interested in calculating the upper confidence limit for a one-sided interval, the lower limit is assumed to be zero, and therefore the observed interval length is simply the value of the upper confidence limit.

P-Confidence

In the general case, p -confidence is defined for the binomial in terms of the equal-tail p -value function using the binomial distribution and the observed number of successes [31]. In the upper-tailed one-sided case, when $x=0$, the p -value function reduces to a function of π alone:

$$P(\pi, 0) = \Pr(X = 0 | \pi).$$

Let π_u = the upper confidence limit of the one-sided $100(1-\alpha)\%$ C.I. computed using any particular method (continuity-corrected Wilson, Agresti-Coull, etc.). For a C.I. $I(0)$ based on zero successes,

$$p\text{-confidence} = \left(1 - \sup_{\pi \in I(0)} P(\pi, 0)\right) \times 100\%$$

$$= [1 - \sup_{\pi \in I(0)} \Pr(X = 0 | \pi)] \times 100\%.$$

$$= [1 - \Pr(X = 0 | \pi_u)] \times 100\%$$

$$= [1 - (1 - \pi_u)^n] \times 100\%.$$

Vos and Hudson [31] give the following interpretation of p -confidence: “The p -confidence is a measure of how large the associated p values are for parameters outside an interval. If we are interpreting a C.I. as the inversion of a hypothesis test, then values of the parameter outside a good approximate C.I. should not have associated p values that are appreciably larger than α where $(1-\alpha) \times 100\%$ is the confidence level.” Thus, it is desirable for p -confidence to be close to the nominal $100(1-\alpha)\%$ confidence coefficient for approximate confidence intervals.

The relationship between coverage probability, which is a measure of the long-run performance of a C.I. procedure, and p -confidence, which is a “post-data” measure of performance, has a connection with the relationship between Neyman’s Type I error rate and Fisher’s p -value in hypothesis testing [31]. The concept of p -confidence is closely related to the concept of consonance [32], and therefore is especially helpful when thinking of C.I.’s as inversions of hypothesis testing procedures, as in the present article. It is useful when interpreting the interval $(0, \pi_u)$ to ask the question: “How large could the probability of zero successes in n trials be for values outside this interval?” The answer is “ $1-p$ -confidence.”

For example, consider the 95% Agresti-Coull interval when $x=0$.

Method	Formula for Upper Limit	Upper Limit	Clopper-Pearson Confidence Coefficient†
Agresti-Coull	$[2 / (n + 4)] + 1.645 \sqrt{[2 / (n + 4)][1 - 2 / (n + 4)] / (n + 4)}$	0.0580	98.5
Wilson	$Z^2_{1-\alpha} / (n + Z^2_{1-\alpha})$	0.0372	93.0
Continuity-Corrected Wilson	$\left\{ [0.5 + (Z^2_{1-\alpha} / 2) + Z_{1-\alpha}] \sqrt{0.5 - (0.25 / n) + (Z^2_{1-\alpha} / 4)} \right\} / (n + Z^2_{1-\alpha})$	0.0500	97.2
Clopper-Pearson	$1 - (\alpha)^{1/n}$	0.0419	95.0
Mid-P Adjusted Clopper-Pearson	$1 - (2\alpha)^{1/n}$	0.0324	90.0
Poisson Approximation	$-\ln(\alpha) / n$	0.0428	95.3
SAIFS	$(1 + t_{n-1, 1-\alpha}) / (n + 1)$	0.0376	93.2
Bayes-Laplace Prior	$1 - (\alpha)^{1/(n+1)}$	0.0413	94.8

SAIFS: Single Augmentation with an Imaginary Failure or Success

†The Clopper-Pearson confidence coefficient based on the specified upper limit.

Table 1: Upper Confidence Limits for One-sided Intervals when $x=0$, $n=70$, $\alpha=.05$.

The values of p -confidence for these intervals range from 95% for $n=5$ to 98.5% for $n=100$. In other words, the probability of zero successes in n trials for values outside these intervals ranges from 0.015 for $n=100$ to 0.05 for $n=5$. In contrast, the probability of zero successes in n trials for values outside the 95% C-P interval is 0.05 for all n .

Another criterion that can be used to evaluate C.I.'s after the data have been observed is p -bias [31]. However, p -bias is equal to zero for the one-sided intervals examined in this article, so it will not be considered further.

Acceptability limits

To assist us in determining if the p -confidence for an approximate C.I. procedure differs in a meaningful way from the nominal $100(1-\alpha)\%$ confidence level, we adapted the following "liberal" guideline proposed by Bradley [33] for evaluating the robustness of a statistical test: if the true significance level differs from the nominal level by no more than $\alpha/2$, one can conclude that the test is robust. If the true significance level differs by more than $\alpha/2$ from the nominal level (either above or below), one can conclude that the test is not robust. In the present study, we applied the Bradley criterion as follows: if the p -confidence differed from the nominal confidence level by no more than $\alpha/2$, we concluded that the p -confidence for that procedure was within acceptable limits. If the p -confidence differed by more than $\alpha/2$ from the nominal confidence level (either above or below), we concluded that the p -confidence for that method was not acceptable. Thus, for a 90% C.I., the p -confidence for a C.I. procedure must be between 85% and 95% to be acceptable; for 95% and 99% C.I.'s, these limits are 92.5% and 97.5% and 98.5% to 99.5%, respectively.

Results

Observed interval length

Figure 1 displays the ratio of observed C.I. length relative to the length of the C-P interval for $n=5$ (1) 10 (5) 50 (10) 100 and a nominal confidence level of 95%. (Similar figures for 90% and 99% intervals are available from the second author. The conclusions based on these figures are very similar to those based on Figure 1). Of the seven methods that were evaluated relative to the C-P method, the Wilson and mid- p methods produce the shortest intervals for values of n between 5 and 100 (Figure 1). The B-L HPD intervals and the SAIFS

intervals for $n>10$ are also consistently shorter than the C-P intervals. The 95% C.I.'s based on the Poisson approximation (the Rule of 3) and the continuity-corrected Wilson intervals are always longer than the C-P intervals. The Agresti-Coull method consistently produces the longest 95% intervals for $n>10$.

P-Confidence

Figures 2 through 4 give the range of p -confidence values for $n=5$ (1) 10 (5) 50 (10) 100 and confidence coefficients of 90%, 95%, and 99%, respectively. The "acceptability" limits are indicated on the figures by a dotted, black, boldface line.

For 90% C.I.'s, the C-P method, the continuity-corrected Wilson method, the SAIFS method, and the B-L HPD method maintain p -confidence within the acceptable limits of 85% to 95% for all n (Figure 2). Intervals based on the Poisson approximation have acceptable p -confidence except when $n=5$. The mid- p and Wilson intervals have p -confidence well below the lower acceptability limit of 85% for all values of n .

Similar to the results for 90% intervals, the p -confidence results for 95% intervals for the continuity-corrected Wilson method and the SAIFS method fall within the acceptability limits for all n (Figure 3). For the B-L HPD intervals, p -confidence falls within the acceptability limits as long as $n \geq 7$.

In contrast, the 95% Agresti-Coull intervals maintain p -confidence above 98% for $n \geq 20$. The Poisson (Rule of 3) intervals have high values of p -confidence for values of $n < 9$ but as n increases, the p -confidence falls between the acceptability limits. The mid- p method maintains a p -confidence of 90%, well outside the lower acceptability limit. The Wilson intervals eventually reach an acceptable level of p -confidence when $n \geq 30$, but p -confidence can be as low as 88.5% for smaller values of n .

For 99% C.I.'s, p -confidence for the mid- p intervals is 98%, below the lower acceptability limit of 98.5% (Figure 4). Except for small values of n , the SAIFS intervals have p -confidence below 98.5% and can be as low as 96.6%. For $n < 20$, the Poisson approximation method produces p -confidence as high as 99.9%. The continuity-corrected Wilson intervals have acceptable p -confidence only for $n < 10$; otherwise, p -confidence is greater than the upper acceptability limit.

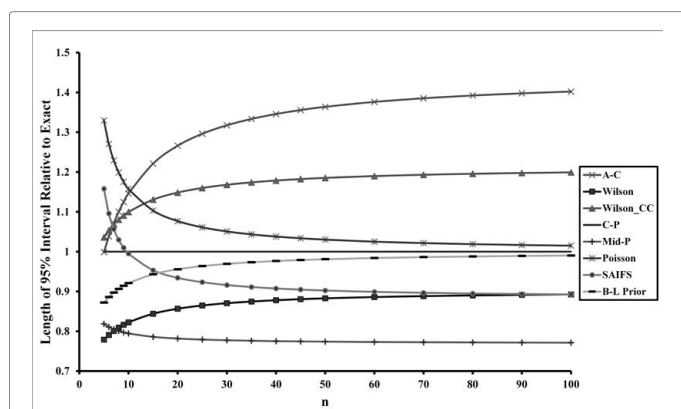


Figure 1: Ratio of interval length relative to the Clopper-Pearson (C-P) interval for the seven 95% confidence intervals when $n=5$ (1) 10 (5) 50 (10) 100. A-C: Agresti-Coull Interval; Wilson_CC: Continuity-Corrected Wilson Interval; SAIFS: Single Augmentation with an Imaginary Failure or Success; B-L: Bayes-Laplace Highest Posterior Density Interval.

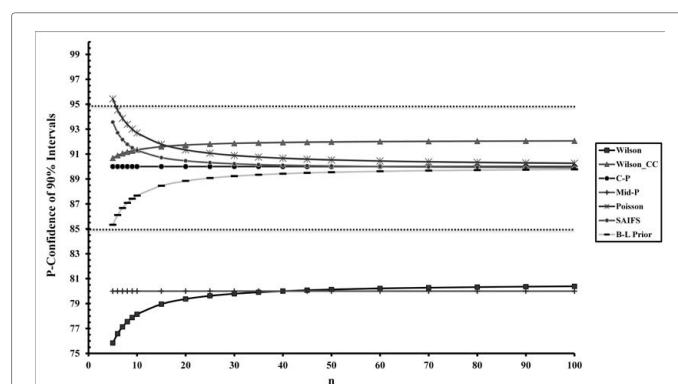


Figure 2: P -confidence for the seven 90% confidence intervals when $n=5$ (1) 10 (5) 50 (10) 100. The bold face dotted lines indicate the arbitrary "acceptability limits" defined in the text. Wilson_CC: Continuity-Corrected Wilson Interval; C-P: Clopper-Pearson Interval; SAIFS: Single Augmentation with an Imaginary Failure or Success; B-L: Bayes-Laplace Highest Posterior Density Interval.

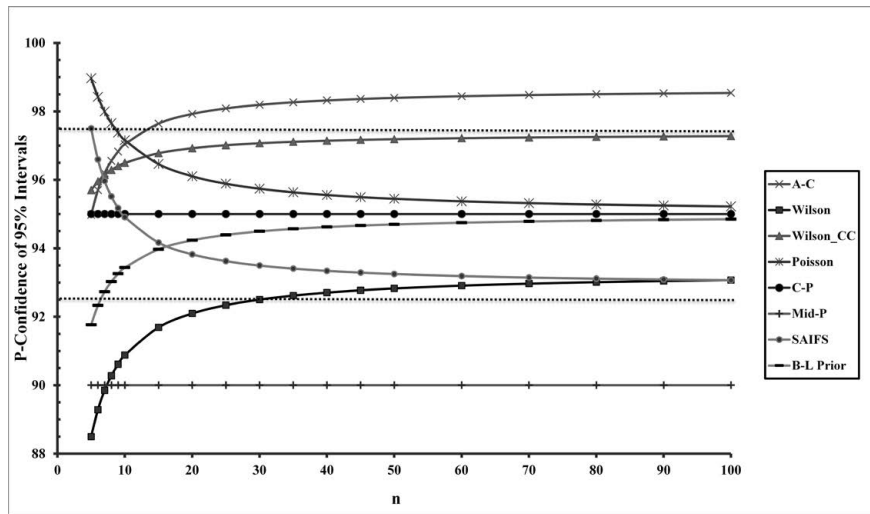


Figure 3: P-confidence for the eight 95% confidence intervals when $n=5$ (1) 10 (5) 50 (10) 100. The bold face dotted lines indicate the arbitrary “acceptability limits” as defined in the text. A-C: Agresti-Coull Interval; Wilson_CC: Continuity-Corrected Wilson Interval; C-P: Clopper-Pearson Interval; SAIFS: Single Augmentation with an Imaginary Failure or Success; B-L: Bayes-Laplace Highest Posterior Interval. Note that for 95% intervals, the Poisson approximation is equivalent to the Rule of Three.

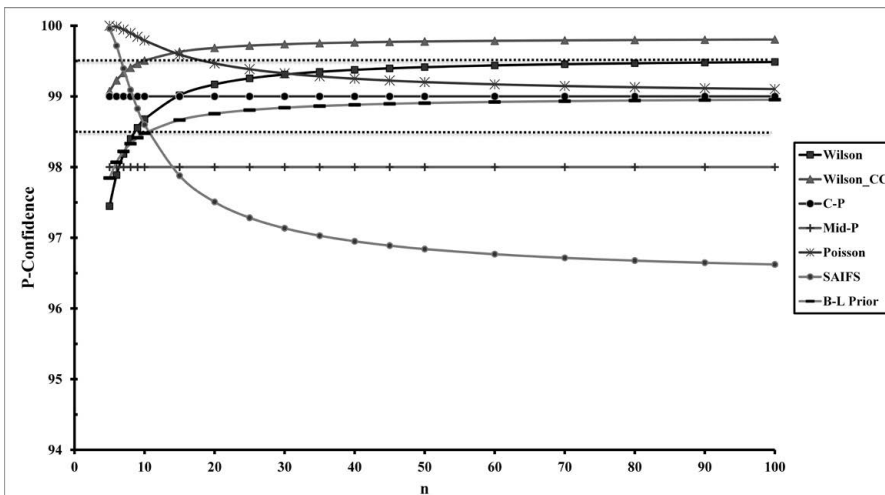


Figure 4: P-confidence for the seven 99% confidence intervals when $n=5$ (1) 10 (5) 50 (10) 100. The bold face dotted lines indicate the arbitrary “acceptability limits” defined in the text. Wilson_CC: Continuity-Corrected Wilson Interval; C-P: Clopper-Pearson Interval; SAIFS: Single Augmentation with an Imaginary Failure or Success; B-L: Bayes-Laplace Highest Posterior Density Interval.

In contrast, the Wilson intervals produce p -confidence less than the lower acceptability limit for $n < 9$ and acceptable values otherwise. The B-L HPD intervals have unacceptable p -confidence for $n \leq 10$, but acceptable values otherwise.

Discussion and Conclusion

Our results indicate that many popular approximate methods that are known to have good long-run properties in terms of coverage probability and expected interval length in the general binomial setting perform poorly when $x=0$. For example, the Agresti-Coull method consistently produces unnecessarily long 95% C.I.’s, resulting in values of p -confidence much larger than 95%. Thus, we recommend that the Agresti-Coull method not be used when $x=0$. Although the mid- p interval has been recommended for general use when finding a C.I. for a binomial proportion, it should not be used when zero successes are

observed because it produces intervals that is much too short, resulting in unacceptably low p -confidence.

If one wishes to use an approximate method, the Wilson, continuity-corrected Wilson, Poisson, SAIFS, and Bayes-Laplace HPD intervals are all acceptable for various combinations of sample size and confidence coefficient. However, none of these methods had acceptable values of p -confidence for all combinations of $n=5$ (1) 10 (5) 50 (10) 100 and nominal confidence coefficients of 90%, 95%, and 99%.

Despite criticism that the C-P method tends to produce confidence intervals that are too long in the general estimation setting [18, 20], this method performed quite well when $x=0$. The moderate length of the C-P intervals, relative to the other methods, together with the fact that C-P intervals will always have p -confidence equal to the nominal confidence level, regardless of sample size, lead us to recommend the C-P method for general use when $x=0$. An additional advantage

is that when $x=0$ the C-P interval is equivalent to a Bayesian credible interval based on the modified Jeffreys prior recommended by Brown, Cai, and DasGupta [24] for general use when estimating the binomial parameter. In terms of the clinical problem that prompted our investigation, the C-P interval has another advantage in that it has a very simple interpretation for the client: "We can be quite certain that the probability of damaging the neurotransmitter is less than 4.2% because, if it is greater than that, observing no damaged neurotransmitters in 70 patients is very unlikely, occurring with probability less than 0.05."

One could argue that we "stacked the deck" in favor of the C-P method in our comparison of the various methods since any exact C.I. method will always have p -confidence equal to the nominal confidence level, regardless of sample size [31]. However, several authors have argued against use of confidence intervals based on the C-P method under any circumstances and instead recommended that an approximate method be used; for example [20,29,34,35]. Agresti and Coull [20] even titled their article "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions."

Because of this "exact vs. approximate" controversy, we undertook our study to examine the behavior of commonly used and commonly recommended approximate methods, in addition to the exact C-P method, in the special situation when $x=0$. We defined "acceptability limits" based on a modification of Bradley's robustness criterion to assist us in evaluating the p -confidence of the approximate C.I. methods. If one of the approximate methods had achieved acceptable p -confidence for all combinations of n and confidence coefficient that we considered, we would have also recommended it for routine use when $x=0$, especially if it was competitive with the C-P method in terms of observed interval length. However, none of the approximate methods satisfied these conditions. For this reason, we recommend that the Clopper-Pearson exact method be used to estimate the upper confidence limit of a one-sided interval whenever $x=0$ is observed.

Acknowledgement

The authors wish to thank Steven Roberts for providing us with the clinical example that motivated this study, Carl Russell for suggesting the use of exact methods for finding the upper confidence limit, and Alan Agresti, Rickey Carter, Daniel Linder, Warren May, and Paul Vos for their insightful comments on the manuscript.

References

1. Roberts S, West LA, Liewehr FR, Rueggeberg FA, Sharpe DE, et al. (2002) Impact of dental devices on cochlear implants. *J Endodontics* 28: 40-44.
2. Lehman CD, Gatsonis C, Kuhl CK, Hendrick RE, Pisano ED, et al. (2007) MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Eng J Med* 356: 1295-1303.
3. Gudipudi DK, Hartzfeld K, Montemarano N, Del Priore G (2008) Decreased breast cancer rates among women with choriocarcinoma. *Gynecol Oncol* 110: 202-205.
4. Ludbrook J, Lew MJ (2009) Estimating the risk of rare complications: Is the 'Rule of Three' good enough? *ANZ J Surg* 79: 565-570.
5. Romanowski B, Schwarz TF, Ferguson LM, Ferguson M, Peters K, et al. (2014) Immune response to the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose or 3-dose schedule up to 4 years after vaccination: Results from a randomized study. *Human Vacc Immunotherap* 10: 1155-1165.
6. Cui F, Liang X, Wang F, Zheng H, Hutin YJ, et al. (2014) Development, production, and postmarketing surveillance of hepatitis A vaccines in China. *J Epidemiol* 24: 169-177.
7. Louis TA (1981) Confidence intervals for a binomial parameter after observing no successes. *Am Stat* 35: 154.
8. Jovanovic BD, Levy PS (1997) A look at the Rule of Three. *Am Stat* 51: 137-139.
9. Razzaghi M (2002) On the estimation of binomial success probability with zero occurrence in sample. *J Mod Appl Stat Meth* 1: 326-332.
10. Olivier J, May W (2006) Weighted confidence interval construction for a binomial parameter. *Stat Meth Med Res* 15: 1-10.
11. Carter RE, Woolson RF (2007) Safety assessment in pilot studies when zero events are observed. *Adv Stat Meth Health Sci* 349-358.
12. Liu J, Chen F, Zhu F-C, Bai J-L, Li J-X, et al. (2015) Evaluation of vaccine seroresponse rates and adverse event rates through Bayesian and frequentist methods. *Human Vacc Immunotherap* 11: 1557-1563.
13. Winkler RL, Smith JE, Fryback DG (2002) The role of informative priors in zero-numerator problems: Being conservative versus being candid. *Am Stat* 56: 1-4.
14. Tuyl F, Gerlach R, Mengerson K (2008) A comparison of Bayes-Laplace, Jeffreys, and other priors: The case of zero events. *Am Stat* 62: 40-44.
15. Tuyl F, Gerlach R, Mengerson K (2009) The Rule of Three, its variants and extensions. *Int Stat Rev* 77: 266-275.
16. Longford NT (2009) Analysis of all-zero binomial outcomes with borderline and equilibrium priors. *J Appl Stat* 36: 1259-1265.
17. Pradhan V, Evans JC, Banerjee T (2013) Binomial confidence intervals for testing non-inferiority or superiority: A practitioner's dilemma. *Statistical Methods in Medical Research*, published online 2 August 2013.
18. Hernandez AM, Andres M, Tejedor H (2016) One-sided asymptotic inferences for a proportion. *J Appl Stat* 43: 1738-1752.
19. Snedecor GW, Cochran WG (1980) *Statistical Methods*: (7th edn). Iowa State University Press, Ames, Iowa.
20. Agresti A, Coull B (1998) Approximate is better than 'exact' for interval estimation of binomial proportions. *Am Stat* 52: 119-126.
21. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 22: 209-212.
22. Casella G (2001) Comment on interval estimation for a binomial proportion. *Stat Sci* 16: 120-122.
23. Clopper CJ, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26: 404-413.
24. Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. *Stat Sci* 16: 101-133.
25. Lancaster HO (1961) Significance tests in discrete distributions. *J Am Stat Assoc* 56: 223-234.
26. Agresti A (2007) *An Introduction to Categorical Data Analysis*. (2nd edn). John Wiley & Sons, Hoboken, NJ.
27. Leemis LM, Trivedi KS (1996) A Comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 50: 63-68.
28. Hanley JA, Lippman-Hand A (1983) If nothing goes wrong, is everything alright? Interpreting zero numerators. *J Am Med Assoc* 249: 1743-1745.
29. Borkowf CB (2006) Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failure or success. *Stat Med* 25: 3679-3695.
30. Roberts S, Vender JR, Causey MS, Roberts JR, Loushine RJ, et al. (2009) Impact of dental devices on neurostimulators. *J Endodontics* 35: 422-425.
31. Vos P, Hudson S (2009) Evaluation criteria for discrete confidence intervals: Beyond length and coverage. *Am Stat* 59: 137-142.
32. Kempthorne O, Folks L (1971) *Probability, statistics, and data analysis*. Iowa State University Press, Ames, Iowa.
33. Bradley JV (1978) Robustness? *Br J Math Stat Psychol* 31: 144-152.
34. Pires AM, Amado C (2008) Interval estimators for a binomial proportion: Comparison of twenty methods. *REVSTAT-Stat J* 6: 165-197.
35. Yu W, Guo X, Xu W (2014) An improved score interval with a modified midpoint for a binomial proportion. *J Stat Comput Simul* 84: 1022-1038.