PharmaSUG 2016 - Paper SP08

Everything or Nothing - A Better Confidence Intervals for Binomial Proportion in Clinical Trial Data Analysis.

Sakthivel Sivam, Quartesian LLC, Princeton, New Jersey
Subbiah Meenakshisundaram, L.N Government College, Ponneri, India

ABSTRACT

Often in Pharmaceutical Research, Confidence intervals have become an important aspect of reporting statistical results. In particular, extensive literature is available for the interval estimators for a binomial proportion (p) that usually arises from the inferential problem for number of successes X ~ Binomial (n, p). However, the occurrences of the event exactly at boundaries (x = 0 or n) have drawn more research interest in spite of few recommendations such as continuity corrections and truncating the limits with 0 or 1. In this paper, one of the widely applied methods, score interval has been considered to improve its performance exactly when x = 0 or n. The proposed approach is based on boundary corrections due to exact method and the Score Interval in its original form; this alleviates the issues related to continuity corrections and adding any pseudo successes or failures. Performance and comparative analyses have been carried out to study the robustness of the present approach using coverage probability, expected length and Root Mean Square Error as evaluation criteria under different n, x and p with special focus on proportion exact boundary. Results have revealed that the proposed interval uniformly achieves nominal coverage and has uniformly minimum expected length as well. We illustrate the implementation of the proposed intervals in a computing environment with real time clinical study data.

KEY WORDS

Binomial proportion, Confidence interval, Score Interval, Approach, Boundary.

INTRODUCTION

In pharmaceutical research there are numerous binary events for which the expected proportion is low. For example, the presence of life threatening adverse effect during post-marketing testing when none have been observed during pre-marketing clinical trials. In the realm of scientific study, a binary outcome with x = 0 observed successes are an often ignored issue, and sometimes incorrectly interpreted as indicating p = 0. Observing zero successes is more likely to happen in clinical research, this issue is exacerbated by small sample sizes. The polar opposite of this case, x = n observed successes, is also of great importance. Yet, a binomial distribution can be defined by either the number of successes as well as by the number of failures. Interpretation made from confidence intervals estimated from x = 0 successes and x = 0 failures are equivalent.

In 1983, Hanley, a Canadian statistician, published the paper "If nothing goes wrong is everything alright". This paper describes in detail the statistical implications if an event of interest fails to occur in a finite number of trials/experiments. Instead of assuming that a trial is safe because of zero numerators, we should look at confidence intervals between zero and a certain upper limit.

According to Rothman (1986) 'confidence intervals convey information about magnitude and precision of effect simultaneously, keeping these two aspects of measurements closely linked'. It is very important to use a confidence procedure that presents the full spectrum of value for an outcome of interest, consistent with the observed data. On the community level, it would be unethical and unfortunate to either underestimate or overestimate the effect being studied. It is more appropriate to produce narrower, less conservative intervals that align the mean coverage with 1 – α.

There have been several papers written on the subject of interval estimation for the binomial in this special case, each offering possible solutions (Wilson, 1927; Louis, 1981; Hanley and Lippman-Hand, 1983; Newcombe, 1998, 2011; Agresti and Coull, 1998; Brown et al, 2001; Fleiss et al, 2003; Borkowf, 2006; Oliver et al, 2007; Reed, 2007; Pires et al., 2008; Xiaomin He et al, 2009; Martin et al., 2014; Thulin, 2014).

Interval estimation of binomial proportion is one of the most basic problems in statistics theory. Let X be a binomial random variable with n trials and a binomial proportion p. The confidence interval for the parameter p is of interest in this paper. Many solutions have been proposed, however, no one is even close to being accepted by the majority of statisticians or applied scientists. The commonly applied confidence interval are basically derived by using either normal approximations or exact probability calculations on binomial distributions. For convenience of discussion, we call these approximate intervals and exact intervals, respectively. Thulin et al, (2014) mentioned that "Constructing a confidence interval for a proportion p based on a binomial sample is a basic but important problem in statistics. Due to the discreteness of the binomial distribution, it is not possible to construct confidence intervals with exact coverage".

In this paper, one of the widely applied methods, Wilson (Score) interval has been considered to improve its performance exactly when x = 0 or n. The proposed approach is based on boundary corrections due to exact method and the Wilson Interval in its original form; this alleviates the issues related to continuity corrections and adding any pseudo successes or failures.

Five exact and approximate methods have been considered for comparative analysis; they are identified as Clopper-Pearson (C-P), Wilson, Wilson with continuity correction and boundary truncation (WCCBT), Agresti-Coull (A-C) and Wilson with exact method boundary correction for extreme cases (WEMBC).

Performance and comparative analyses have been carried out to study the robustness of the present approach using coverage probability, expected length and Root Mean Square Error as evaluation criteria under different n, x and p with special focus on proportion exact boundary.

Illustrative data sets are extracted from many published application studies to resemble the parametric values of the binomial experiment. Results have revealed that the proposed interval performs better when compared to other methods and uniformly achieves nominal coverage and has uniformly minimum expected length as well.

METHODS AND NOTATIONS

In this section, we discuss different approaches to compute confidence interval for single binomial proportion using well established approximate and exact intervals. The intervals are defined by (1-α) 100 % confidence level. Intervals for all the discussed methods were calculated using SAS 9.3 ®.

ADJUSTMENTS

CONTINUITY CORRECTION

Continuity correction was developed in order to reduce the anticonservatism of the Wald interval. Incorporating a continuity correction (CC) 1/ (2n) improves coverage and avoids degenerate intervals, but increases the propensity to overshoot the boundaries at 0 and 1.

PSEUDO FREQUENCY

Many authors (Agresti-Coull, 1998; Borkowf, 2006; Martin et al., 2011) have tried to improve the performance and quality of the confidence interval of existing methods, by adding a pseudo frequency h to the original data, that is according to the data (x+h, (n-x) +h, n+2h). One such adjustments is from well-known Agresti-Coull method for the improvisation of the Wald interval. This approach was introduced to improve the coverage probability and expected width of the interval to avoid anti-conservatism.

ABERRATIONS OF TWO KINDS – OVERSHOOT, DEGENERACY

Overshoot - For low proportions such as prevalence's, when the numerator is small the calculated lower limit can be below zero. Conversely, for proportions approaching one, such as the sensitivity and specificity of diagnostic or screening tests, the upper limit may exceed one. The glaring absurdity of overshoot is readily avoided by truncating the interval to lie within [0, 1], of course, but even this is not always done.

Degenerate, zero width interval [ZWI] occurs when p=0 or 1 for any 1- α < 1.

METHODS

C-P METHOD

The Clopper-Pearson (C-P) (1934) introduced this interval and this confidence interval is an early and considerably common method for calculating binomial confidence intervals. The C-P confidence interval is commonly called an exact confidence interval because it is based on the cumulative probabilities of the binomial distribution. C-P interval eliminates both aberrations and guarantees strict conservatism in that the coverage probability is at least 1- α for all p with 0 < p < 1. Expressions for the lower and upper limits can be written using quantiles of Beta distribution.

$$(LCL, UCL) = \left(Beta\left(\frac{\propto}{2},\ X, n-X+1\right),\ Beta\left(1-\frac{\propto}{2}, X+1, n-X\right)\right), when\ 0 < X < n$$

$$(LCL, UCL) = \left(0,\ 1-\left(\frac{\propto}{2}\right)^{\frac{1}{n}}\right), when\ X = 0$$

$$(LCL, UCL) = \left(\left(\frac{\propto}{2}\right)^{\frac{1}{n}}, 1 \right), when\ X = n$$

By design, the C-P method gives CIs with at least nominal (1-α) 100 percent coverage for all values of $p \in [0,1]$ and for all sample sizes (n). Coverage can be quite supranominal in some cases, however, especially of values of p near 0 and 1 and for small sample sizes (n).

WILSON (SCORE)

Wilson (1927) introduced this interval and we call this interval the Wilson interval. Wilson interval are based on inverting the binomial test statistic and are thus theoretically justified. Lower and upper limits are defined as

$$(LCL, UCL) = \left(\frac{2np + z^2 - z\sqrt{\{z^2 + 4npq\}}}{2(n + z^2)}, \frac{2np + z^2 + z\sqrt{\{z^2 + 4npq\}}}{2(n + z^2)} \right);$$

$$where\ z\ is\ the\ z_{1-\frac{\propto}{2}}$$

WILSON METHOD WITH CONTINUITY CORRECTION AND BOUNDARY TRUNCATION (WCCBT)

Wilson method incorporating continuity correction and boundary truncation. Expressions for the lower and upper limits can be written as

$$LCL = \frac{2np + z^2 - 1 - z\sqrt{\left\{z^2 - 2 - \frac{1}{n} + 4p(nq + 1)\right\}}}{2(n + z^2)}$$

$$UCL = \frac{2np + z^2 + 1 + z\sqrt{\left\{z^2 + 2 - \frac{1}{n} + 4p(nq - 1)\right\}}}{2(n + z^2)}\ ;\ where\ z\ is\ the\ z_{1-\frac{\propto}{2}}$$

However, if x = 0, LCL must be taken as 0; if x =n, UCL is then 1.

AGRESTI- COULL (A-C)

Agresti and Coull (1998) have proposed an alternative to the Wald interval known as the "Add two successes and two failures adjusted Wald interval." Augmenting the original data by adding Pseudo frequency to compute the confidence bounds. This interval is constructed using a modified point estimate with the Wald form and Confidence bounds are truncated, if they fall outside the unit interval.

95 % interval estimate is

$$(LCL, UCL) = \left(\frac{x + 2}{n + 4} - Z_{\frac{\alpha}{2}}\sqrt{\frac{(x + 2)(n + 2)}{(n + 4)^3}}, \quad \frac{x + 2}{n + 4} + Z_{\frac{\alpha}{2}}\sqrt{\frac{(x + 2)(n + 2)}{(n + 4)^3}} \right)$$

WILSON INTERVAL WITH EXACT METHOD BOUNDARY CORRECTION FOR EXTREME CASES (WEMBC)

$$(LCL, UCL) = \left(\frac{2np + z^2 - z\sqrt{\{z^2 + 4npq\}}}{2(n + z^2)}, \frac{2np + z^2 + z\sqrt{\{z^2 + 4npq\}}}{2(n + z^2)} \right);$$

$$when\ 0 < X < n$$

$$where\ z\ is\ the\ z_{1-\frac{\propto}{2}}$$

$$(LCL, UCL) = \left(0,\ \ 1 - \left(\frac{\propto}{2}\right)^{\frac{1}{n}}\right), when\ X = 0$$

$$(LCL, UCL) = \left(\left(\frac{\propto}{2}\right)^{\frac{1}{n}}, 1\right), when\ X = n$$

CRITERIA FOR EVALUATION

It must first be decided what are the appropriate criteria by which to evaluate competing methods. The issue of coverage probability, conservatism and interval width is crucial.

COVERAGE PROBABILITY

For any confidence interval procedure for estimating p, the actual coverage probability at a fixed value of p is

$$C_n(p) = \sum_{x=0}^{n} Bin(x; n, p)\ I_{[p \in L(x)]}\ ; where\ \ I_{[p \in L(x)]} = 1, if\ p\ \in L(x),$$

$$Else\ 0.$$

$$L(x)\ is\ the\ limit\ of\ x.$$

EXPECTED WIDTH

$$W_n(p) = \sum_{x=0}^{n} [UCL(x) - LCL(x)]\ Bin(x; n, p)$$

ROOT MEAN SQUARE ERROR (RMSE)

Root Mean Square Error is meant to describe how far actual coverage probabilities typically fall from the nominal confidence level.

$$RMSE = \sqrt{\int_0^1 (C_n(p) - (1-\propto))^2\ dp}$$

RESULTS AND DISCUSSIONS

EXAMPLE

We illustrate the discussed methods in real time clinical study data extracted from Newcombe (2011) and other published application studies. Simulation study has been performed for comparative analysis to evaluate the performance and robustness of the methods in exactly boundary cases using the sample sizes n of 9, 63 and 257 selected from the literature to cover the small, moderate and large small size scenarios and presented the intervals for 90 & 95 %.

Age group	Prop. Surviving to discharge	95 % confidence interval from different methods				
		C-P	Wilson	WCCBT	A-C	WEMBC
75-84	3/265 (0.0113)	(0.0023, 0.0327)	(0.0039, 0.0327)	(0.0029, 0.0355)	(0.0024, 0.0347)	(0.0039, 0.0327)
85 and over	0/115 (0)	(0.0000, 0.0316)	(0.0000, 0.0323)	(0.0000, 0.0403)	(-0.0063, 0.0399)	(0.0000, 0.0316)

Table 1. Survival rates for elderly patients suffering cardiac arrest at home in whom resuscitation was attempted using automatic external defibrillators (Colquhoun et al., 2008)

Table 1 indicates the following observations; for the non-zero numerator with moderate size n, A-C and WCCBT provides a wider confidence interval, while Wilson & WEMBC gives the shortest.

When x=0, It is worth noting that the A-C interval gives confidence bounds that are beyond the boundary values (i.e. less than zero), the problem of overshoot. Also, WCCBT provides a wider interval followed by A-C and Wilson, whereas WEMBC provides a narrower interval.

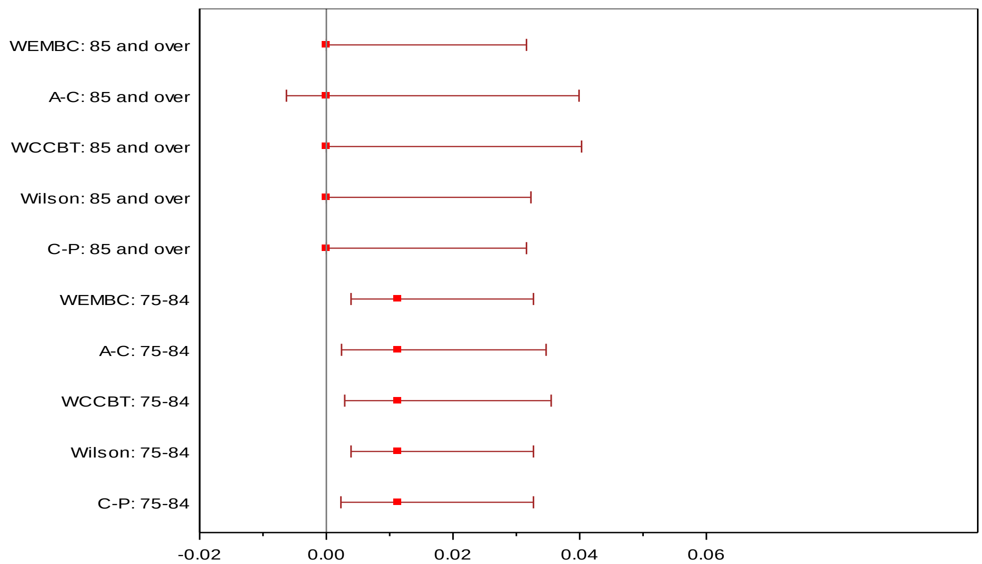

Figure 1. Confidence intervals of different methods by age groups

Figure 1 displays the confidence intervals computed from different methods for each age group. From this figure, the problem of Overshoot and overestimation is very much evident.

COMPARISON OF COVERAGE PROBABILITY

n	Level	C-P	Wilson	WCCBT	A-C	WEMBC
265	90%	0.919	0.894	0.919	0.904	0.904
	95%	0.960	0.942	0.960	0.954	0.952
115	90%	0.928	0.904	0.928	0.907	0.904
	95%	0.965	0.944	0.963	0.956	0.954
9	90%	0.965	0.903	0.967	0.911	0.920
	95%	0.986	0.945	0.983	0.963	0.958
63	90%	0.931	0.895	0.936	0.906	0.905
	95%	0.969	0.945	0.968	0.957	0.954
257	90%	0.919	0.892	0.921	0.904	0.902
	95%	0.960	0.952	0.961	0.954	0.952

Table 2. Mean Coverage Probability

From Table 2, it is observed that C-P, WCCBT and A-C achieves coverage probability of at least (1-α) 100 percent consistently in small, moderate and large n, which obviously leads to conservative interval and results in overestimation of the effect being studied. However, Wilson is close to nominal coverage probability and WEMBC interval uniformly achieves nominal coverage for any n at different α levels.

COMPARISON OF EXPECTED WIDTH

n	Level	C-P	Wilson	WCCBT	A-C	WEMBC
265	90%	0.082	0.078	0.082	0.079	0.078
	95%	0.097	0.093	0.097	0.094	0.093
115	90%	0.127	0.118	0.127	0.120	0.119
	95%	0.149	0.141	0.149	0.143	0.141
9	90%	0.469	0.387	0.468	0.406	0.397
	95%	0.531	0.452	0.526	0.484	0.460
63	90%	0.174	0.159	0.174	0.162	0.159
	95%	0.203	0.189	0.204	0.194	0.189
257	90%	0.083	0.080	0.083	0.080	0.080
	95%	0.098	0.095	0.099	0.095	0.095

Table 3. Mean Expected Width

Table 3 reveals that C-P, WCCBT and A-C provides a wider expected width which indicates the overestimation problem, while Wilson & WEMBC gives the shortest. However, WEMBC provides consistently narrower interval for varying sample sizes and different α levels and achieves uniformly the desirable property of minimum expected length.

COMPARISON OF RMSE

n	Level	C-P	Wilson	WCCBT	A-C	WEMBC
265	90%	0.025	0.021	0.025	0.017	0.017
	95%	0.012	0.010	0.013	0.010	0.008
115	90%	0.032	0.018	0.033	0.018	0.018
	95%	0.018	0.011	0.015	0.011	0.010
9	90%	0.069	0.065	0.069	0.044	0.042
	95%	0.037	0.029	0.035	0.022	0.023
63	90%	0.036	0.032	0.040	0.020	0.025
	95%	0.022	0.016	0.020	0.012	0.012
257	90%	0.024	0.021	0.026	0.016	0.017
	95%	0.013	0.008	0.014	0.010	0.008

Table 4. Root Mean Square Error

Table 4 reports the uniform-weighted root mean squared error of those probabilities about that confidence level. These values indicate that the variability about the nominal level is much smaller for the WEMBC confidence interval under varying sample sizes and level α than for A-C, C-P and WCCBT intervals.

DISCUSSIONS

We examined C-P, A-C and three different versions of Wilson method confidence intervals in order to determine the most useful and reliable method to estimate confidence intervals when zero successes or failures. In this study, we calculated the Expected width, RMSE of the confidence intervals and compared the accuracy of the coverage probabilities.

Procedures for estimates near the boundary are too often ignored. The current approach of blending two or more of the intervals is motivated by the realization that some of the confidence intervals behave very differently for small samples values near the boundary. Hence, the proposed approach accounts for situations when x is 0 or n.

Although C-P interval is typically treated as gold standard in statistical practice. However, it provides wider interval which is conservative and always results in overestimation of the effect being studied when compared to other approximate intervals with reasonable adjustments.

The A-C interval is more conservative than the Wilson interval. It provides wider interval and results in overshoot problem, which leads to truncation of limits at boundary.

The WCCBT interval provides a wider interval when x = 0 or n which results in undesirable property of overestimation. And it is often as conservative as C-P interval and hence it is not reliable to use in the practical situation of exact or near boundary.

Though Wilson and WEMBC method has coverage close to the nominal coverage and comparatively shorter expected width on an average. But in view of accuracy of the interval in boundary cases, Wilson interval is not as good as WEMBC interval.

When estimating the binomial parameter on the boundary, from the anti-conservative and coverage consideration standpoint, it would be more reasonable to use the WEMBC confidence interval as it does not require CC or pseudo frequency adjustments, provides narrower interval and uniformly achieves nominal coverage and minimum expected length as well for different sample sizes and level α and shown to have better performance than the C-P and WCCBT confidence interval.

Based on the analysis, we recommend WEMBC interval for any n and level α. WEMBC interval is therefore can be considered as natural benchmark for the new confidence intervals.

CONCLUSIONS

Problem on Confidence interval estimation for binomial proportion looks very simple, but yet it is very interesting even today. The non-occurrence of any outcome in the clinical trial data does not mean that it cannot happen. It can, and the true rate of occurrence can be estimated from its confidence interval. It is a good estimate of the worst case that is compatible with the observed data.

When considering the computation of an interval estimate for a binomial proportion the first decision the applied statistician must take is related to the balance between degree of conservativeness and efficiency. As there are no standard recommendations for the use of CC and pseudo frequency to improve the efficiency of interval. Usage of CC and pseudo frequency impacts the quality of the interval in many cases.

If strict conservativeness is mandatory, then go for C-P. If conservativeness is not a major concern, then focus on shortest interval. WEMBC approach remains a valid choice.

WEMBC interval provides significantly closer match to nominal confidence, especially at small values of n. WEMBC interval is appropriate when a high degree of accuracy is required.

Finally, the diverse approaches to binomial CI construction that one finds in the statistical literature reflect the broad range of real-world applications for these methods. Statisticians and researchers have a wonderfully diverse collection of methods from which to choose, and they should carefully consider the consequences of their choices for the application on which they work.

REFERENCES

- Hanley JA and Lippman-Hand A. 1983. "If nothing goes wrong, is everything all right? Interpreting zero numerators". JAMA. 249(13),1743-1745.
- Rothman, K. 1986. Modern Epidemiology. Boston: Little Brown.
- Wilson, E.B. 1927. “Probable Inference, the Law of Succession, and Statistical Inference”. Journal of the American Statistical Association. 22, 209–212.
- Newcombe, R.G. 1998. “Two sided confidence intervals for the single proportion: Comparison of seven methods”. Statistics in medicine. 17, 857-872.
- Agresti, A and Coull, B.A. 1998. “Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions”. The American Statistician. 52,119–126.
- Brown, L.D., Cai, T.T. and DasGupta, A. 2001. “Confidence intervals for a binomial proportion (with discussion)”. Statistical Science. 16, 101–133.
- Fleiss, J.L., Levin, B., and Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. New York: John Wiley & Sons.

- Borkowf, C.G. 2006. “Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failures or success”. Statistics in Medicine. 25, 3679-3695.
- Oliver, J. and May, W.L. 2007. “A Discussion of Binomial Interval Estimators on the Boundary”. Journal of the Mississippi Academy of Sciences. 52, 178-183.
- Reed, J.F. 2007. “Better Binomial confidence intervals”. Journal of Modern Applied Statistical Methods. 6 (1), 153-161.
- Pires, A.M. and Amado, C. 2008. “Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods”. REVSTAT – Statistical Journal. 6, 165-197.
- Xiaomin, He. and Shwu-Jen, Wu. 2009. “Confidence Intervals for the Binomial Proportion with Zero Frequency”. Paper SP10-2009.
- Martin Andres, A. and Alvarez Hernandez, M. 2014.” Two-tailed asymptotic inference for a proportion”. Journal of Applied Statistics. 41 (7), 1516-1529.
- Mans, Thulin. 2014. “The cost of using exact confidence intervals for a binomial proportion”. Electronic Journal of statistics. 8 (1), 817-840.
- Mans, Thulin. 2014.” Coverage-adjusted confidence intervals for a binomial proportion.”. Scandinavian Journal of Statistics. 41(2), 291-300.
- Clopper, C.J. and Pearson, E.S. 1934. “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial”. Biometrika. 26, 404–413.
- Newcombe, R.G. 2011. “Measures of location for confidence intervals for proportions”. Communication in Statistics - Theory and Methods. 40, 1743 - 1767.
- Newcombe, R.G. and Nurminen, M, M. 2011. “In Defence of Score intervals for proportions and their difference”. Communication in Statistics - Theory and Methods. 40, 1271-1282.
- Colquhoun, M., Chamberlain, D.A., Newcombe, R.G., Harris, R., Harris, S., Peel, K., Davies, C.S. and Boyle, R. 2008. “A national scheme for public access defibrillation in England and Wales: early results. Resuscitation. 78, 275-280.

ACKNOWLEDGMENTS

We are grateful to the Senior management team of Quartesian LLC for their encouragement.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Name: Sakthivel Sivam, M.Sc., M.Phil., (Ph.D).
Enterprise: Quartesian LLC
Address: 3 Independence Way, Suite 106,
City, State ZIP: Princeton, NJ 08540
Work Phone: 609 454 3312
E-mail: Sakthivel.sivam@quartesian.com
Web: www.quartesian.com