Reasonable Confidence Limits For Binomial Proportions

SUMMARY

This article adds an entry to the long list of binomial confidence intervals introduced since 1934. The importance of the new entry is demonstrated by comparison to the two most commonly used intervals: the "Wald" and the "Exact". Those two intervals are so wide that their confidence limits can be statistically significantly different from the sample proportion on which basis they were calculated. In contrast, the interval introduced in this article is as wide as possible without its limits differing significantly from the sample proportion. Thus the decision to use this new interval may be easier to defend to regulatory bodies such as the FDA; use of any other interval may appear to be arbitrary or unreasonable. The proposed name for the new entry is: "Reasonable Confidence Interval for a Binomial Proportion", the extreme values of which would then be "Reasonable Confidence Limits". Also discussed is the simultaneous use of two intervals (the Exact and the Reasonable).

KEY WORDS: binomial confidence interval; binomial confidence limit; proportion; Wald; normal approximation; Exact; Reasonable; Score Method, zone of uncertainty.

INTRODUCTION

Confidence intervals are serious business! Recently, the FDA told a medical-device start-up company that, in regards to the company's proposed clinical trial, "The equivalence of the device to the predicate can be demonstrated if the confidence interval for the difference in the mean values for the tested parameter excludes a difference larger than 20% from the predicate." Unfortunately, the company could not meet that FDA mandate because in order to reduce the width of the confidence interval so that it would achieve that exclusion, a much larger number of patients was required than the company could afford.
The term "confidence interval" was coined and first published by J. Neyman in 1934, when he applied it to binomial as well as variables data. He described confidence intervals as ranges "in which we may assume are contained the values of the estimated characters of the population." As applied to a "proportion...of individuals in the sample" (which in this article will be represented by "Ps") that has been derived from an unknown "proportion...of individuals in the population" ("Pp", in this article) whose "distribution...is then a binomial", he defined the confidence interval for Pp as having the form Pt ≤ Ps ≤ Pu, where Pt is what we now call the "lower confidence limit", Pu is the "upper confidence limit", and Pp is assumed with a specified level of "confidence" to be somewhere in the interval Pt − Pu. A few months later, the first rigorous method for calculating such an interval was published, by Clopper and Pearson.

There is only one generally accepted method for calculating confidence intervals for variables data; that method involves "t" tables and the standard error of the mean, as described in any basic statistics text book. However, there are many methods in use for calculating confidence intervals for binomial data, each such method resulting in different confidence limits and different interval widths.

The reason that there are many binomial methods is partly historical and partly theoretical. The historical part is that, prior to the widespread availability of computers, binomial calculations were difficult. For that reason, simpler-to-calculate alternatives were developed; as was said several decades ago, "To calculate [large-sample binomial] probabilities would be an almost insurmountable task. Therefore, some method of approximation must be used." The theoretical part is the lack of agreement on criteria for judging an interval (we'll discuss this in more detail later in this article). That disagreement was present from the birth of the interval concept in 1934: Neyman's description for the interval was the equivalent of Pt ≤ Ps ≤ Pu, whereas Clopper & Pearson's description was Pt < Ps < Pu; notice the use of ≤ vs. <.

Some of those many binomial confidence interval methods involve formulæ calculations: for example, what is commonly referred to as the "Wald" formula uses a "binomial standard deviation" coupled with a Standard Normal Distribution Z-table to calculate an approximate confidence interval; Wald intervals are based upon the fact that when sample size is large or Ps is near 50%, the Normal distribution is a "reasonable" model of the Binomial, even if such an application is "not completely accurate." Other methods involve trial-and-error: for example, in order to calculate each confidence limit for what is commonly called the "Exact binomial confidence interval", repeated attempts must be made to determine the proportion that yields a cumulative binomial histogram probability of exactly half the chosen significance level (we'll discuss this in more detail, below).

It has been said recently that the Wald interval is "in virtually universal use." Similarly, the National Institute for Standards and Technology (NIST) "e-Handbook of Statistical Methods" website states that the Wald formula is the "confidence [interval] expression most frequently used." And Wald's was the only method included in a binomial-calculation-spreadsheet issued in 1998 by the CDRH division of the FDA, for general use...
in handling clinical trials data. On the other hand, the Exact interval is the sole method provided in some mainstream statistical software programs (e.g., Statgraphics®) and the Exact is the only non-Z-table method for calculating binomial proportion confidence limits that is mentioned on the NIST website; indeed, some statisticians (e.g., Agresti and Coull) refer to the Exact method as the "gold standard."

The implicit if not explicit focus of interest in a binomial confidence interval is its largest and smallest values, i.e., Pu and Pl. For example, a clinical trial might be considered successful only if the lower confidence limit on the outcome success rate is larger than a protocol-specified value. Because of such focus, this paper introduces a new criterion for comparing the validity of confidence interval methods (since 1934, many other criteria have been proposed). The new criterion is this: Are the confidence limits of the interval "reasonable"? If it is unreasonable to conclude that Pl and Pu could be Pr, then it is unreasonable to use the confidence interval method that generated them. And being "reasonable" is what even J. L. Fleiss has urged: "A confidence interval for a statistical parameter is a set of values that are...reasonable candidates for being the true underlying value" (underlining not in the original text).

**REASONABLE CONFIDENCE INTERVALS AND LIMITS**

Before we define what it means to be "reasonable", let's explain the basis of the definition. A test of reasonableness is equivalent to performing a binomial test of a statistically significant difference using what J. L. Fleiss has called the "traditional statistical approach" for an "inference for a single proportion". It involves "calculating the probability, assuming the null hypothesis holds, of obtaining the outcome that actually occurred, plus the probabilities of all other outcomes as extreme as, or more extreme than, the one that was observed; and rejecting the null hypothesis in favor of the alternative hypothesis if the sum of all these probabilities — the so-called p-value — is less than or equal to a predetermined level, denoted by α, called the significance level."

Based on that approach, let's call a confidence interval "reasonable" only if such a test results in a conclusion of "not statistically different" when the test compares the observed Sample proportion (Ps) to either of the two most extreme values in the interval, namely the upper and lower confidence limits (Pu and Pl, respectively). In terms a bit more mathematical, let's define "unreasonable" as follows (for sample size = N, observed number of successes in that sample = K, and K/N = Ps): In regards to the probability distribution histograms derived from limits Pl and Pu, if either of the distribution tails in which K is found represents a cumulative probability of occurrence of less than or equal to α/2, we conclude that Ps is statistically significantly different from the limit that generated that distribution, and that therefore the confidence interval Pl − Pu is unreasonable.

In that definition, we use α/2 rather than α because we are performing a 2-sided test twice. The first test determines whether a random sample proportion differs from Pl; the second test determines whether that random sample proportion differs from Pu; in each case, we
focus on tails that represent $\alpha/2$ of the distribution (see Figures 1A and 1B). That is the standard way to approach such tests.\textsuperscript{10, 17, 18, 19}

It is important to note that the test of significance just described is performed using the probability distribution histograms generated from the two confidence limits ($P_L$ and $P_U$) rather than being performed using the single probability distribution histogram generated from the observed Sample proportion ($P_S$). Such an approach is used because $P_L$ and $P_U$ are together considered to be the "null hypothesis"; based upon the "new criterion" described above, the question we are trying to answer is this: Is it reasonable to assume that the random sample proportion $P_S$ could have been obtained from a population that had a proportion equal to either $P_L$ or $P_U$ (i.e., could either of them be Fleiss's "true underlying value", $P_P$)?\textsuperscript{15, 17, 18}

Let's examine the "reasonableness" of Wald limits, after we first explain how to calculate them. As demonstrated on the NIST website\textsuperscript{20}, calculation of the upper and lower confidence limits of a Wald ("Normal Approximation") binomial confidence interval uses the following formula: $P_S \pm Z_{\alpha/2} \times SD_{PS}$, where $P_S$ is the observed Sample proportion, $Z_{\alpha/2}$ is the two-tailed value from a normal distribution Z-table at the chosen $\alpha$ significance level, and $SD_{PS}$ is the binomial standard deviation for the observed proportion calculated as the square root of $P_S(1-P_S)/N$. The limits of such an interval can be calculated as shown below, using Microsoft® Excel\textsuperscript{TM} (subscript "W" indicates the "Wald" method, "Normsinv" and "Sqrt" are MS Excel functions that output Z-table values and square roots respectively, " * " is the MS Excel symbol for "multiply", and the other terms are as defined above):

\begin{align*}
  P_{UW} &= P_S + \text{Normsinv}(1-\alpha/2) \times \text{Sqrt}(P_S \times (1-P_S) / N) \\
  P_{LW} &= P_S - \text{Normsinv}(1-\alpha/2) \times \text{Sqrt}(P_S \times (1-P_S) / N)
\end{align*}

Because the "normal approximation" assumption becomes less valid as the sample size becomes smaller and/or as $P_S$ departs farther from 0.500 (50%), the Wald calculation is typically restricted to situations where both of the following are true: $N(P_S) > 5$ and $N(1-P_S) > 5$. Even the FDA's spreadsheet\textsuperscript{11} includes the warning "minimum $[N(P_S), N(I-P_S)]...must be > 5 to use normal approximation".

Let's examine the reasonableness of Wald confidence limits for the following situation: $\alpha = 5\%$ (and therefore Confidence = $1-\alpha = 95\%$), Sample size = $N = 100$, Successes = $K = 10$, $P_S = K/N = 10/100 = 0.10$, and $N(P_S) = 10$. The resulting limits are: $P_{LW} = 0.041201116$ and $P_{UW} = 0.158798884$; how reasonable are they?

Let's focus just on $P_{LW}$, the lower limit. The probability distribution for a population whose proportion equals that $P_{LW}$ is shown in Figure 2. Notice that the probability of occurrence of the observed Sample result or a more extreme result (i.e., $K \geq 10$) is less than $\alpha/2 = 2.5\%$ (in fact, it is 0.8%); we therefore conclude that $P_S$ is statistically significantly different from $P_{LW}$. Based upon our definition of reasonableness, it is therefore unreasonable to conclude that $P_{LW} = P_P$. If it is unreasonable to conclude that
PLW = PP (at α = 5%), then it is unreasonable to consider PLW − PUW to be the 1 − α = 95% confidence interval. As evidenced by this example, Wald intervals and confidence limits can be unreasonable.

Let's now examine the reasonableness of Exact limits, after we first explain how to calculate them. What are sought (by trial and error) are two proportions, one larger and one smaller than the observed sample proportion (Ps); each must have a cumulative binomial probability of "exactly" α/2 for obtaining the observed Sample result or a more extreme value (i.e., 0 to K, or K to N). An MS Excel spreadsheet can be used to calculate the limits as accurately as (for example) Statgraphics, to at least a billionth of a probability unit (= 9 places to the right of the decimal point); how to do so is described on the NIST website. The following is a generic example of applying the NIST/MSExcel method (α, N, K, Pr, Ps, Pu, and PL are as defined above; subscript "E" indicates the "Exact" method; and P is a proportion we seek):

PL = the value of P (P < PS) needed to ensure that the MS Excel function Binomdist(K−1, N, P, True) outputs a probability value of α2 precisely (to the desired number of significant digits).

PU = the value of P (P > PS) needed to ensure that the MS Excel function Binomdist(K, N, P, True) outputs a probability value of α2 precisely (to the desired number of significant digits).

Those formulas, applied to the situation we evaluated previously (N = 100, K = 10, PS = 0.10, and α = 5%), result in the following Exact limits: PLE = 0.049004689 and PUE = 0.176222598. The probability distribution for a population whose proportion equals that PLE is shown in Figure 3. Notice that the probability of occurrence of the observed Sample result or a more extreme result (i.e., K ≥ 10) is precisely equal to α/2 = 2.5%; the entire histogram bar representing K=10 (the value observed in the Sample) is found in the 2.5% "tail" of the distribution; we therefore conclude that PS = K/N is statistically significantly different from PLE. Based upon our definition of reasonableness, it is therefore unreasonable to conclude that PLE = Ps. If it is unreasonable to conclude that PLE = Ps (at α = 5%), then it is unreasonable to consider PLE − PUE to be the 95% confidence interval. As evidenced by this example, Exact confidence intervals and limits can be unreasonable.

Let's now introduce formulas for the "Reasonable" confidence limits being introduced in this article. The limits of a "Reasonable binomial confidence interval" are defined as follows (where the subscript "R" indicates the "Reasonable" method):

PR = the value of P (P > PS) needed to ensure that the MS Excel function Binomdist(K−1, N, P, True) outputs a probability value of α2 precisely (to the desired number of significant digits).
**PLR** = the value of P (P < Ps) needed to ensure that the MS Excel function

Binomdist(K, N, P, True) outputs a probability value of 1– α/2 precisely
(to the desired number of significant digits).

Notice that the first term in the MS Excel functions for "Reasonable" limits is changed by a value of 1 from its corresponding "Exact" function (in the PU definitions, the change is from K to K–1, and in the PL formulas it is from K–1 to K). That was done to ensure that Ps is not statistically significantly different from either PLR or PUR. In effect, those two formulas identify the widest possible confidence interval such that the observed Sample proportion is not statistically significantly different from any point in the interval, most especially the highest and lowest points, namely PUR and PLR (the meaning of "statistically significantly different" is discussed further, later in this article).

If those Reasonable method formulas are applied to the situation we evaluated previously (N = 100, K = 10, Ps = 0.10, and α = 5%), they result in the following Reasonable limits:

PLR = 0.056207020 and PUR = 0.163982255. The probability distribution for a population whose proportion equals that PLR is shown in Figures 4A and 4B. Notice that, in Figure 4A, the probability of occurrence of the observed Sample result or a more extreme result (i.e., K ≥ 10) equals approximately 5.5 %; in Figure 4B, the entire histogram bar representing K=10 is found in the (1 − α/2)% body+lower_tail (i.e., in the lower 97.5% of the distribution); therefore, the observed Sample proportion (Ps = K/N = 0.10) is not statistically significantly different from PLR, and therefore PLR could possibly be Pp.

Similarly, as seen in Table 1, it is reasonable to conclude that PUR could possibly be Pp, since the observed Sample result or a more extreme result (i.e., K ≤ 10) is equal to approximately 4.9% and the probability of K ≤ 9 is precisely 2.5%. Therefore, because both PLR and PUR could possibly be Pp (at α = 5%), it is indeed reasonable to consider PLR–PUR to be the 95% confidence interval.

At first glance, the Wald method has an advantage over Reasonable and Exact ones: Calculation of Wald limits can be performed without trial-and-error and therefore can be automated using simple computer applications such as MS Excel functions. Upon further investigation, we discover that the Beta-distribution formulas that have been used to approximate Exact limits without using trial-and-error can be modified to also approximate Reasonable limits. As shown in Table 2, the following MS Excel formulas approximate the output of the Reasonable binomial formulas shown above, to at least a millionth of a probability unit (= 6 places to the right of the decimal point) (α, N, K, PUR, and PLR are as defined above, and "BetaInv" is an MS Excel function):

\[ \text{PUR (BETA)} = \text{BetaInv} (1 - \alpha/2, K, N - K + 1) \]

\[ \text{PLR (BETA)} = \text{BetaInv} (\alpha/2, K+1, N - K) \]

NOTE: When either K = 0 (Ps = 0.000) or K = N (Ps = 1.000), those two beta-formulas produce error messages or nonsense results (see next paragraph, below).
Let's now turn our attention to the most extreme values of $P_s$: When $P_s$ equals 0.000 precisely, no method can calculate a Lower Confidence Limit ($P_L$), because proportions (i.e., probabilities) lower than zero are undefined; and when $P_s$ equals 1.000 precisely, no method can calculate $P_U$, because proportions above unity are likewise undefined. Not surprisingly, the $P_L$ for $P_s = 1.000$ and the $P_U$ for $P_s = 0.000$ are called "one-sided limits".

Such one-sided limits can be calculated by the Exact method but not by Wald or Reasonable ones. Wald limits for $P_s = 0.000$ or 1.000 cannot be calculated because in either case the binomial standard deviation itself equals 0.000, no matter what the sample size is (recall that $SD_{PS} = \sqrt{P_s(1-P_s)/N}$). Similarly, Reasonable limits cannot be calculated because if $K = 0$ ($P_s = 0.000$), then the $P_{UR}$ binomdist formula value "$K−1$" is meaningless; and if $K = N$ ($P_s = 1.000$), then the $P_{LR}$ binomdist formula always results in a probability of 1.000 and therefore can never equal the sought-after "value of $P (P < P_s)$".

On the other hand, Reasonable and Exact methods (but not the Wald method) share the following advantage: They can calculate two-sided confidence limits for any $P_s$ greater than zero and less than unity, no matter how small or large (e.g., if sample size is a million and $K = 1$, then $P_s = 0.000001 = 1.000E-6$; and $P_{LR} = 0.242E-6$).

A common criterion for evaluation of confidence interval methods is "coverage". That term refers to the percentage of time the interval can be expected to include $P_P$, the "true underlying value". Typically, that percentage is determined experimentally by generating confidence intervals for thousands of random samples drawn from populations of known proportions (i.e., known $P_P$'s), and then determining what percentage of those intervals contain the corresponding $P_P$. As seen in Figure 5, Reasonable confidence limits are completely contained within Exact ones, and therefore Reasonable intervals can be expected to have slightly less "coverage"; experimental results support that conclusion (see Table 3).

The distinctiveness of Reasonable Intervals and their Limits is not apparent when the limits are plotted in the still-common manner introduced in 1934 by Neyman (see also Figure 6). However, alternate plotting methods (e.g., see Figure 7) clearly demonstrate that Reasonable Intervals are narrower than Wald or Exact ones. As a result, only Reasonable Intervals are narrow enough to have Upper and Lower limits that are truly "reasonable". As Neyman insisted: "confidence intervals should be as narrow as possible."  

**RECOMMENDATIONS**

With variables data, a test of significance is mathematically equivalent to use of a confidence interval; i.e., a borderline value being compared to the sample result is considered significantly different if it is outside the sample's confidence interval, but non-significant if inside. Likewise, a borderline binomial proportion is classically viewed as being inside or outside the sample proportion's confidence interval. But such a view is misleading for proportions, since they are based upon counts; because counts can take only discrete or "discontinuous" values, the observed result's probability distribution
histogram bar appears as if it spans the border between significance and non-significance, when the histogram is based upon a borderline value.

For example: In the case of N=100 and observed result $K=10$, we've already discussed (see Figure 3) that probability distribution histograms based upon $P \leq P_L = 0.0490$ result in the histogram bar for $K=10$ being in the upper 2.5% tail of the distribution. Similarly, we've discussed (see Figure 4B) that if the histogram is based upon $P \geq P_{LR} = 0.0562$, then $K=10$ is in the 97.5% body+lower_tail of the distribution. The problem is that histograms based upon $P$ values between $P_L$ and $P_{LR}$ result in the observed-$K$ histogram-bar being neither fully in the 2.5% upper tail nor fully in the 97.5% body+lower tail (see Figures 8A and 8B). In such cases, how do we objectively conclude significance or non-significance? Either conclusion could be viewed as unreasonably subjective and arbitrary.

A solution, introduced in this article, is to always use not one but two confidence intervals: the Exact and the Reasonable.

Before explaining that solution, a brief history lesson is useful (in the following discussion, "$L_\delta$" is the likelihood of obtaining the observed result "$\delta"$, assuming the null hypothesis is true). The concept of "statistical significance" has undergone much change since the precursors to modern tests of significance were developed in the 1800s; in that century, $\delta$ was not considered significant unless it was extremely unlikely. As decades passed, the requirement for significance became less extreme. By the 1930's it could be said that "it is conventional among certain workers to adopt the following rule: If $L_\delta \geq 0.05$, $\delta$ is not significant; if $L_\delta \leq 0.01$, $\delta$ is significant; if $0.05 > L_\delta > 0.01$, our conclusions about $\delta$ are doubtful and we cannot say with much certainty whether the deviation is significant or not until we have additional information" (the original text uses a different symbol than "L" and does not include underlining). In effect, the region between 0.01 and 0.05 was considered a "zone of uncertainty" (a term not in the original text).

On the basis of that history lesson, perhaps the best solution to the problem of borderline proportions is to consider the range of values between $P_L$ and $P_{LR}$, and between $P_{UR}$ and $P_{UE}$, to be "zones of uncertainty" (see Figure 9). Using that approach, and assuming your sample size has the "power" to detect a clinically significant difference, if the Null Hypothesis proportion ($P_{NH}$) being compared to the study result ($P_S$) is outside the study result's Exact confidence interval (i.e., $P_{NH} \leq P_L$ or $P_{UE} \leq P_{NH}$), you can claim statistical significance. On the other hand, if $P_{NH}$ is inside the corresponding Reasonable interval (i.e., $P_{LR} \leq P_{NH} \leq P_{UR}$), you can claim statistical non-significance. However, if $P_{NH}$ is in either of the zones of uncertainty (that is, either $P_L < P_{NH} < P_{LR}$ or $P_{UR} < P_{NH} < P_{UE}$), then your results are statistically inconclusive.

FDA NOTES

A search of the website FDA.gov finds recent submissions, panel opinions, and guidance documents that include a variety of binomial confidence interval methods, the most common being the Exact and the "Score" (the Score formula is an elaborated classic Wald, from the perspective that it uses a binomial standard deviation and not one but a
few copies of a normal distribution Z-table value; Score confidence limits are found midway between Exact and Reasonable ones). In a 2007 "Statistical Guidance" document, the FDA recommends "score confidence intervals, and alternatively, exact (Clopper-Pearson) confidence intervals." 27 Exact vs. Score methods were compared briefly in a 2009 FDA publication: "An advantage with the Score method is that...it can be calculated directly. Score confidence bounds tend to yield narrower confidence intervals than Clopper-Pearson [i.e., Exact] confidence intervals, resulting in a larger lower confidence bound." 28 It is interesting to note that that recommendation was based on ease-of-calculation rather than on theoretical correctness.

On the other hand, in 2003, Medtronic Neurological received approval for its Active Dystonia Therapy ("deep brain stimulation system") by using the classic Wald method almost exclusively: "Exact 95% confidence intervals were used when the # (%) of patients was 0 (0%) because the normal approximation to the binomial does not provide a confidence interval. In every other case, the normal approximation to the binomial was used to calculate confidence intervals" even though in very many of those cases the "N(Ps) > 5" criterion (mentioned above) was not met 29 (underlining was not in the original text).

The new methods proposed in this article have not yet been used in any regulatory submission to the FDA, Health Canada, or a Notified Body, as far as this author knows.

CONCLUSION

The most commonly used binomial confidence intervals can have confidence limits that are statistically significantly different from the random sample proportion from which they were derived. Therefore, such confidence limits are unreasonable to consider as being the proportion of the population from which the random sample was taken. A more defensible choice of confidence intervals is presented in this article, namely a "Reasonable Confidence Interval for a Binomial Proportion", the extreme values of which are "Reasonable Confidence Limits". Although such an interval is slightly narrower than other intervals and thus offers less "coverage", it is more reasonable to use because it is the widest possible range that contains no values that are statistically significantly different from the sample proportion on which basis the interval was calculated. Alternatively, an even more reasonable approach is to use a combination of Exact and Reasonable limits, coupled with the concept of "zones of uncertainty".
REFERENCES


11. FDA Website Document, "Two Group Confidence Interval & Power Calculator" version 1.6 (Issued March 26, 1998).

12. Statgraphics™ Centurion XV, version 15.2.12 (Copyright 1982-2007 by StatPoint, Inc.)


Example of a Probability Distribution Derived from an "Unreasonable" Lower Confidence Limit on an Observed 10 Successes in a Sample Size = N

Observed result = 10 differs significantly from the Lower Confidence Limit, because K ≥ 10 has probability ≤ (0.5)Alpha%.

Figure 1A
Example of a Probability Distribution Derived from an "Unreasonable" Upper Confidence Limit on an Observed 10 Successes in a Sample Size = N

Observed result = 10 differs significantly from the Upper Confidence Limit, because K ≤ 10 has probability ≤ (0.5)Alpha%.

Figure 1B
Figure 2

Probability Distribution for Binomial Proportion $P_{LW} = 0.041201116$

This tail contains 0.834% of the distribution.
Probability Distribution for Binomial Proportion = $P_{LE} = 0.049004689$

![Probability Distribution](image)

Figure 3
Probability Distribution for Binomial Proportion $P_{LR} = 0.056207020$

Figure 4A
Probability Distribution for Binomial Proportion = PLR = 0.056207020

Figure 4B
Table 1  

<table>
<thead>
<tr>
<th></th>
<th>K ≤ 9</th>
<th>K ≤ 10</th>
<th>K ≥ 10</th>
<th>K ≥ 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wald:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PuW = 0.158798884</td>
<td>0.033816</td>
<td>≈ 3.4 %</td>
<td>0.064560</td>
<td>≈ 6.5 %</td>
</tr>
<tr>
<td>PLW = 0.041201116</td>
<td>0.008340</td>
<td>≈ 0.8 %</td>
<td>0.002809</td>
<td>≈ 0.3 %</td>
</tr>
<tr>
<td><strong>Exact:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PuE = 0.176222598</td>
<td>0.011762</td>
<td>≈ 1.2 %</td>
<td>0.025000</td>
<td>≈ 2.5 %</td>
</tr>
<tr>
<td>PLF = 0.049004689</td>
<td>0.025000</td>
<td>≈ 2.5 %</td>
<td>0.009978</td>
<td>≈ 1.0 %</td>
</tr>
<tr>
<td><strong>Reasonable:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRU = 0.163982255</td>
<td>0.025000</td>
<td>= 2.5 %</td>
<td>0.049303</td>
<td>≈ 4.9 %</td>
</tr>
<tr>
<td>PLR = 0.056207020</td>
<td>0.054862</td>
<td>≈ 5.5 %</td>
<td>0.025000</td>
<td>= 2.5 %</td>
</tr>
</tbody>
</table>

Assuming a population has $P_{xx}$ proportion of successes, this table lists the probability for the listed range of $K$ successes, in a sample of size $N = 100$. 

Wald: 

$P_{UW} = 0.158798884 \\ \approx 3.4 \% \quad 0.064560 \quad \approx 6.5 \%$

$P_{LW} = 0.041201116 \\ \approx 0.8 \% \quad 0.002809 \quad \approx 0.3 \%$

Exact: 

$P_{UE} = 0.176222598 \\ \approx 1.2 \% \quad 0.025000 \quad \approx 2.5 \%$

$P_{LE} = 0.049004689 \\ \approx 2.5 \% \quad 0.009978 \quad \approx 1.0 \%$

Reasonable: 

$P_{UR} = 0.163982255 \\ \approx 2.5 \% \quad 0.049303 \quad \approx 4.9 \%$

$P_{LR} = 0.056207020 \\ \approx 5.5 \% \quad 0.025000 \quad \approx 2.5 \%$
<table>
<thead>
<tr>
<th>K</th>
<th>K / N</th>
<th>Identified using...</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>Binomial distribution trial+error: 0.002431337</td>
<td>0.002431333</td>
<td>0.036216693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.002431333</td>
<td>0.036216736</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.10</td>
<td>Binomial distribution trial+error: 0.056207020</td>
<td>0.056206942</td>
<td>0.163982255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.056206942</td>
<td>0.163982391</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>Binomial distribution trial+error: 0.177394438</td>
<td>0.177394390</td>
<td>0.335735489</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.177394390</td>
<td>0.335735321</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>Binomial distribution trial+error: 0.408036329</td>
<td>0.408036232</td>
<td>0.591963671</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.408036232</td>
<td>0.591963768</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.75</td>
<td>Binomial distribution trial+error: 0.664264511</td>
<td>0.664264679</td>
<td>0.822605562</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.664264679</td>
<td>0.822605610</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.90</td>
<td>Binomial distribution trial+error: 0.836017745</td>
<td>0.836017609</td>
<td>0.943792980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.836017609</td>
<td>0.943793058</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>0.99</td>
<td>Binomial distribution trial+error: 0.963783307</td>
<td>0.963783264</td>
<td>0.997568663</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta distribution formula: 0.963783264</td>
<td>0.997568667</td>
<td></td>
</tr>
</tbody>
</table>
Wald, Exact, and Reasonable 95% Confidence Limits
When N = 100 and Sample Proportion (P_S) is 0.10

Figure 5
Table 3

<table>
<thead>
<tr>
<th>Pp</th>
<th>Wald</th>
<th>Exact</th>
<th>Reasonable</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>93.4%</td>
<td>95.8%</td>
<td>90.8%</td>
</tr>
<tr>
<td>0.30</td>
<td>95.0%</td>
<td>96.4%</td>
<td>93.6%</td>
</tr>
<tr>
<td>0.50</td>
<td>93.9%</td>
<td>96.2%</td>
<td>93.9%</td>
</tr>
</tbody>
</table>
Figure 6
Lengths of Confidence Intervals ( $P_U - P_L$ )
N = 100, Confidence = 95%

Figure 7
Probability Distribution for "Borderline"
Proportion = 0.0493 (vs. PLE = 0.0490)
The bar representing precisely $K = 10$ successes has
an individual probability of 1.6%; thus almost all of its
area ($15/16$ of it) is in a tail of size $2.6\% - 0.1\% = 2.5\%$.

Figure 8A
Probability Distribution for "Borderline"
Proportion = 0.0560 (vs. $PLR = 0.0562$)
The bar representing precisely $K = 10$ successes has an individual probability of 2.9%; thus part of its area (1/29 of it) is in a tail of size $2.4\% + 0.1\% = 2.5\%$.

Figure 8B
Exact and Reasonable 95% Confidence Limits
When $N = 100$ and Sample Proportion ($P_s$) is 0.10

$P_{LE}$ to $P_{LR}$ is the Lower Zone of Uncertainty.

Below $P_{LE}$ and above $P_{UE}$ are areas of Significance.
Between $P_{LR}$ and $P_{UR}$ is the area of Non-significance.

$P_{UR}$ to $P_{UE}$ is the Upper Zone of Uncertainty.

Figure 9