

Contents

1	Formal definitions	2
1.1	Definition of errors and interval lengths	2
2	Additional interval definitions	3
2.1	General notations	3
2.2	Normal and Student approximate confidence intervals	5
2.2.1	Pan and kurto-match intervals	5
2.2.2	Wald's and Wilson's intervals	5
2.2.3	Agresti-Coull interval	6
2.2.4	Recentered Wald intervals	6
2.2.5	Borkowf SAIFS interval	6
2.2.6	Continuity corrected intervals	6
2.3	Intervals based on normal approximation after transformation	9
2.4	Bayesian intervals	11
2.5	Bootstrap intervals	12
2.6	Exact binomial confidence intervals	14
2.7	Skewness corrected normal approximation intervals	18
2.8	Generalized linear model intervals	19
3	Wald's validity condition	20
3.1	Informal description of the method	20
3.2	Formal description of the method	20
3.3	Results	21
3.3.1	Local average errors control	21
3.3.2	Conditional errors control	23
4	Figures of the nine main confidence intervals	24
5	Figures of additional intervals	31
6	Comparison of strictly conservative exact intervals	37
7	References	42

1 Formal definitions

1.1 Definition of errors and interval lengths

We will first define one-sided conditional errors as actual probabilities that the confidence interval is completely below (or above) the true proportion p when the sample size and p are constant from one experiment to another. Second, we will define one-sided local average errors as actual errors that the confidence interval is completely below (or above) the true proportion P when the sample size is constant from one experiment to another but the true proportion P is a random variable fluctuating around an expected true proportion p_0 . Third, we will define one-sided random sample errors as actual errors that the confidence is strictly below (or above) the true proportion p when the true proportion p is constant from one experiment to another but the sample size N is a random variable fluctuating around an average sample size n_0 . Similarly, we define three different expected interval half-widths (conditional, local average and random sample).

Let x be the realization of a variable $X_{n,p} \sim \mathcal{B}(n; p)$ where n represents the sample size, x the observed number of successes and $\hat{p} = x/n$ the observed proportion of successes. Let

$$IC_{1-\alpha}(x, n) = [L_{1-\alpha}(x, n); U_{1-\alpha}(x, n)] \quad (1)$$

denote the confidence interval at the nominal error α for x successes in n trials. For constant n and p , we denote

$$\alpha'_L(p, n, \alpha) = Pr(L_{1-\alpha}(X_{n,p}, n) > p | n, p) \quad (2)$$

the actual error that the interval is completely above p . Similarly, we denote

$$\alpha'_U(p, n, \alpha) = Pr(U_{1-\alpha}(X_{n,p}, n) < p | n, p) \quad (3)$$

the actual error that the interval is completely below p . We denote

$$\alpha'(p, n, \alpha) = \alpha'_U(p, n, \alpha) + \alpha'_L(p, n, \alpha) \quad (4)$$

the actual error that the interval does not contain p . These three errors (α'_L , α'_U and α') are called the *conditional errors*.

Let us define the *conditional expected lower half-width*

$$w'_L(p, n, \alpha) = E \left[\frac{X_{n,p}}{n} - L_{1-\alpha}(X_{n,p}, n) | n, p \right] \quad (5)$$

equal to the expected distance between the point estimate and the confidence interval lower bound. Similarly, we define the *conditional expected upper half-width*

$$w'_U(p, n, \alpha) = E \left[U_{1-\alpha}(X_{n,p}, n) - \frac{X_{n,p}}{n} | n, p \right] \quad (6)$$

Let us note P a random proportion whose logit is normally distributed with a $\log(OR_S)$ standard deviation. Let $p_0 = E[P]$ be the expected proportion. We define the *right local average error*

$$\alpha''_u(p_0, n, \alpha) = E[\alpha'_u(P, n, \alpha) | n] \quad (7)$$

the *left local average error*

$$\alpha''_l(p_0, n, \alpha) = E[\alpha'_l(P, n, \alpha) | n] \quad (8)$$

and the *two-sided local average error*

$$\alpha''(p_0, n, \alpha) = \alpha_L''(p_0, n, \alpha) + \alpha_U''(p_0, n, \alpha) \quad (9)$$

It is the probability that a confidence interval around \hat{p} does not contain p in a two steps experiment. In the first step, a p proportion is realized from the P random variable. In the second step, a \hat{p} proportion of successes is realized in a random sample of size n with an actual proportion of successes p . The sample size n is held constant in all experiments. Similarly, we define $w_L''(p, n, \alpha)$ and $w_U''(p, n, \alpha)$ the *local average half-widths*. The constant OR_S will be set at 1.20, except in sensitivity analyses ($OR_S = 1.10$ and $OR_S = 1.05$).

We define a random N variable having a discrete distribution. This distribution is defined from a latent log-normal variable rounded to the nearest integer. Let us denote $n_0 = E[N]$ the expectancy of the same size N and SR_S the geometric standard deviation of the latent log-normal variable. The geometric standard deviation SR_S will be set at 1.20. We define the *random sample right average error* as

$$\alpha_U'''(p, n_0, \alpha) = E[\alpha_U'(p, N, \alpha)|p] \quad (10)$$

the actual error that the confidence interval is completely below p , and the *random sample left average error* as

$$\alpha_L'''(p, n_0, \alpha) = E[\alpha_L'(p, N, \alpha)|p]. \quad (11)$$

Let the *relative conditional expected left half-width* $v_L'(p, n, \alpha)$ be the ratio between the expected lower half-width of an interval estimator and the expected lower half-width of the Clopper-Pearson mid-P confidence interval for the same set of parameters p , n and α . Similarly, we define v_U' , v_L'' , v_U'' taking the Clopper-Pearson mid-P reference.

The nominal α error will be set at 0.05. The conditional errors and conditional expected half-widths will be computed from the exact binomial distribution $\mathcal{B}(n, p)$. The local average errors and local average half-widths will be approximated by numerical integration from 512 values the conditional errors or conditional expected half-widths. These 512 values will be uniformly spaced on the logistic scale.

2 Additional interval definitions

Many intervals described below have a slight modification when the proportion of successes is close to zero or one. Any confidence limit below zero is set to zero. Any confidence limit above one is set to one. The lower confidence limit for zero success is set to zero. The upper confidence limit for n successes is set to one. Before publication, the statistician would probably fix intervals that do not contain the point estimate or that exceed zero or one.

The original idea of each interval is briefly described and then formulas are given in tables.

2.1 General notations

These notations may be used anywhere in this appendix.

We denote x the number of successes and n the sample size, $\hat{p} = \frac{x}{n}$ the observed proportion and \hat{P} the proportion estimator for a binomial experiment $\mathcal{B}(n, p)$ and α the nominal two-sided error rate of the confidence interval.

Let us denote $BPF(x; n, p)$ the probability mass function of the binomial distribution, $BCDF(k; n, p)$ the cumulative distribution function and $BiCDF(q; n, p)$ the inverse cumulative distribution function (or quantiles function) of the binomial distribution.

$$BPF(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad (12)$$

$$BCDF(k; n, p) = \sum_{x=0}^k BPF(x; n, p) \quad (13)$$

$$BiCDF(q; n, p) = \min \{k | BCDF(k; n, p) \geq q\} \quad (14)$$

Let us denote $\beta PF(x; \alpha, \beta)$ the probability density function, $\beta CDF(x; \alpha, \beta)$ the cumulative density function and $\beta iCDF(q; \alpha, \beta)$ the inverse cumulative density function of the beta distribution with α and β shape parameters.

$$\beta PF(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta) x^{\alpha-1} (1-x)^{\beta-1}}{\Gamma(\alpha) \Gamma(\beta)} \quad (15)$$

$$\beta CDF(x; \alpha, \beta) = \int_0^x \beta PF(t; \alpha, \beta) dt \quad (16)$$

$$\beta iCDF(q; \alpha, \beta) = \beta CDF^{-1}(q; \alpha, \beta) \quad (17)$$

where the gamma function is defined as

$$\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx \quad (18)$$

Let us denote $\chi_{q,df}^2$ the q quantile of the χ^2 distribution with df degrees of freedom.

Let us denote $t_{q,df}$ the q quantile of the Student t distribution with df degrees of freedom.

Let us denote $NPDF(x)$ the probability density function of the $\mathcal{N}(0,1)$ normal distribution:

$$NPDF(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (19)$$

Let us denote $NCDF(x)$ the cumulative distribution function of the $\mathcal{N}(0,1)$ normal distribution:

$$NCDF(x) = \int_{-\infty}^x NPDF(t) dt \quad (20)$$

Let us denote $NiCDF(q)$ or z_q the quantile function of the $\mathcal{N}(0,1)$ normal distribution:

$$NiCDF(x) = z_x = NCDF^{-1}(x) \quad (21)$$

and κ the quantile $1 - \alpha/2$ of the normal distribution

$$\kappa = z_{1-\alpha/2} \quad (22)$$

2.2 Normal and Student approximate confidence intervals

Normal and student approximate confidence intervals are defined in Table 1. Auxiliary functions are defined in Table 2.

2.2.1 Pan and kurto-match intervals

The Student t with $n - 1$ degrees of freedom is based on the pivot statistic $T = \frac{\hat{M}}{\sqrt{\widehat{VAR}}}$ where \hat{M} is the mean estimator and \widehat{VAR} is the variance estimator. The Student distribution assumes normality of \hat{M} and independence between the \hat{M} and \widehat{VAR} estimators. Pan¹ tried to compute the degrees of freedom of the Student distribution that best fits the following pivot statistic distribution

$$T = \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1 - \hat{P})}{n}}} \quad (23)$$

Pan made the assumption that \hat{P} and $V(\hat{P}, n) = \frac{\hat{P}(1 - \hat{P})}{n}$ are approximately independent (page 145). That assumption is always wrong, especially when p is small, in which case $\hat{P} \approx \frac{1}{n} V(\hat{P}, n)$. Degrees of freedom computed with Pan's formula leads to a Student distribution with a kurtosis that don't match the binomial distribution kurtosis. The skewness doesn't match either, as it's always zero for all Student distributions. For instance, the excess kurtosis of a binomial distribution $\mathcal{B}(100, 0.08)$ computed from the exact probability mass function is 3.08, which matches a t distribution with 5.95 degrees of freedom. For the same binomial distribution, the Pan approximate t distribution has an excess kurtosis equal to 0.348 with 21.24 degrees of freedom. That's why a kurto-match t interval has been inferred from the binomial probability mass function for $\mathcal{B}(n, \hat{p})$, excluding occurrences of zero successes or zero failures. When $\min(x, n - x) \leq 3$, the excess kurtosis of the binomial distribution is negative and no Student distribution can match it. Therefore, when $\min(x, n - x) \leq 3$, the Clopper-Pearson interval has been used.

Pan proposes to apply his procedure to the Agresti-Coull interval, also known as the adjusted Wald's interval. The add4 Pan interval is obtained in the same way as the Pan interval but, $\frac{\kappa^2}{2}$ successes and $\frac{\kappa^2}{2}$ failures are added before computation where κ is the $1 - \alpha$ quantile of the $\mathcal{N}(0, 1)$ distribution, equal to 1,96 for $\alpha = 0,05$. The kurtosis of the kurto-match add4 t interval is based on the exact computation of the studentized pivot variable obtained by adding $\frac{\kappa^2}{2}$ successes and failures.

All those Student approximations are dubious: The pivot statistic is undefined for zero successes and zero failures and matching the kurtosis while ignoring the skewness may not get the best Student approximation.

2.2.2 Wald's and Wilson's intervals

Wald's and Wilson's intervals rely on normality of the \hat{P} estimator but Wald's interval approximates variance of $\mathcal{B}(n, p)$ to variance of $\mathcal{B}(n, \hat{p})$ while Wilson's interval does not. Indeed, Wilson's interval can be built by observed to theoretical χ^2 test inversion. The second degree equation can be solved so that Wilson's formula is closed form. Unlike Wald's interval, Wilson's interval is not symmetrical around the estimated proportion \hat{p} . Wilson's interval is also known as the score interval.

Wilson's interval has been modified by Brown ² in 2001. When the number of successes or failures is too small, the binomial distribution is approximated to a Poisson distribution. It substitutes Garwood's interval ³ for extreme values.

2.2.3 Agresti-Coull interval

Agresti-Coull interval also known as the adjusted Wald interval ^{4,5} is an approximation of Wilson's interval by a simpler formula. Agresti-Coull interval has the same center as Wilson's interval but is always longer ² making it more conservative. It may be calculated by Wald's formula after adding $\frac{\kappa^2}{2}$ successes and $\frac{\kappa^2}{2}$ failures to the sample where κ is the $1 - \alpha$ quantile of the $\mathcal{N}(0,1)$ distribution. For a 95% confidence interval, $\kappa^2 = 3,84$ may be rounded to four. Agresti-Coull formula can be obtained by adding two successes and two failures before using Wald's formula. The formula shown in Table 1 keeps the full precision $\kappa^2 = 3,84$.

2.2.4 Recentered Wald intervals

Brown ⁶ noticed that Wald's interval bias is mainly due to the fact that it has the "wrong" center. He proposed to recenter Wald's interval to Wilson's interval center without changing Wald's interval length.

2.2.5 Borkowf SAIFS interval

Borkowf ⁷ described his method in 2006. It always contains Wald's interval and so, is strictly more conservative. Indeed, the lower boundary of Borkowf's interval is obtained by adding one imaginary failure before using Wald's formula. Consequently, the lower boundary is always lower than Wald's lower boundary. The upper boundary is obtained by adding one imaginary success before using Wald's formula. The upper boundary is always above Wald's upper boundary.

2.2.6 Continuity corrected intervals

For a binomial variable $X \sim \mathcal{B}(n, p)$, the cumulative distribution function $XBCDF(x) = Pr(X \leq x)$ may be approximated by a normal cumulative distribution function $XNCDF(x) = Pr(Y \leq x)$ where $Y \sim \mathcal{N}(np, np(1 - p))$ is a normal variable having the same expectancy and variance as X . Continuity corrected intervals are based on the approximation of $XBCDF(x)$ to $XCCDF(x) = Pr\left(Y \leq x + \frac{1}{2}\right)$. The $XCCDF$ is a better approximation of $XBCDF$ than $XNCDF$ but is a worse approximation of $XMCDP(x) = Pr(X < x) + \frac{1}{2}Pr(X = x)$. Basically, continuity corrected interval are more conservative in the same way that the Clopper-Pearson confidence interval is more conservative than the Clopper-Pearson mid-P confidence interval.

Table 1 : Definition of the lower bounds of approximate normal and Student confidence intervals, the upper bounds being defined by equivariance $U_{1-\alpha}(x, n) = 1 - L_{1-\alpha}(n - x, n)$.

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	
^{2,4,8-10} Wald	(22)	$\max\left(0, \hat{p} - \kappa \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$	(24)
¹¹ Student t with $n - 1$ df		$\max\left(0, \hat{p} - t_{1-\frac{\alpha}{2}, n-1} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\right)$	(25)
^{8,12} Wald with cc	(22)	$\max\left(0, \hat{p} - \kappa \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} - \frac{1}{2n}\right)$	(26)
⁶ Recentered Wald	(22)(38)	$\min\left(\hat{p}, \max\left(0, \tilde{p} - \kappa \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right)\right)$	(27)
¹⁰ Recentered Wald with cc	(22) (38)	$\min\left(\hat{p}, \max\left(0, \tilde{p} - \kappa \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} - \frac{1}{2n}\right)\right)$	(28)
^{1,10} Pan 2002 Wald t interval	(39)	$\begin{cases} \max\left(0, \hat{p} - t_{1-\alpha/2, dfp} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) & \text{if } x > 0 \\ \text{Undefined if } x = 0 \end{cases}$	(29)
Kurto-match t	(41)	$\begin{cases} \hat{p} - t_{1-\alpha/2, dfg(0)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \\ \text{Undefined if } x \leq 3 \end{cases}$	(30)
⁴ Agresti-Coull	(22) (38)	$\max\left(0, \tilde{p} - \kappa \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + \kappa^2}}\right)$	(31)
¹ Pan 2002 add4 t interval	(22) (44)	$\max\left(0, \tilde{p} - t_{1-\frac{\alpha}{2}, df\tilde{p}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + \kappa^2}}\right)$	(32)
Add4 kurto- match t	(22) (41)	$\begin{cases} \tilde{p} - t_{1-\alpha/2, dfg(\frac{\kappa^2}{2})} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + \kappa^2 - 1}} \\ \text{Undefined if } x \leq 5 \end{cases}$	(33)
¹³ Wilson	(22)	$\frac{x + \frac{\kappa^2}{2} - \kappa \sqrt{\frac{x(n-x)}{n} + \frac{\kappa^2}{4}}}{n + \kappa^2}$	(34)
² Modified Wilson	(22)	$\begin{cases} \frac{1}{2n} \chi_{\alpha, 2x}^2 & \text{if } 1 \leq x \leq x^* \\ \frac{x + \frac{\kappa^2}{2} - \kappa \sqrt{\frac{x(n-x)}{n} + \frac{\kappa^2}{4}}}{n + \kappa^2} & \text{otherwise} \end{cases}$	(35)
⁸ Wilson with cc	(22)	$\begin{cases} \frac{1}{2(n + \kappa^2)} \left(2x + \kappa^2 - 1 - \kappa \sqrt{\kappa^2 - 2 - \frac{1}{n} + \frac{4x}{n}(n-x+1)}\right) & \text{if } x > 0 \\ \text{if } x = 0 \end{cases}$	(36)
⁷ Borkowf SAIFS	(22)	$\max\left(0, p' - \kappa \sqrt{\frac{p'(1-p')}{n}}\right) \text{ where } p' = \frac{x}{n+1}$	(37)

Table 2: auxiliary functions used in Table 1.

Function description	Depends on equation	Function definition	
Wilson's interval center	(22)	$\tilde{p} = \frac{x + \frac{1}{2}\kappa^2}{n + \kappa^2}$	(38)
Pan 2002 ¹ degrees of freedom with exponents of \tilde{p} fixed by Pires 2008 ¹⁰	(40)	$dfp = \frac{2 \left(\frac{\tilde{p}(1-\tilde{p})}{n} \right)^2}{\Omega(\tilde{p}, n)}$	(39)
Pan 2002 ¹ Ω function		$\begin{aligned} \Omega(\tilde{p}, n) &= \frac{\tilde{p} - \tilde{p}^2}{n^3} - 2 \frac{\tilde{p} + (2n-3)\tilde{p}^2 - 2(n-1)\tilde{p}^3}{n^4} \\ &+ \frac{\tilde{p} + (6n-7)\tilde{p}^2 + 4(n-1)(n-3)\tilde{p}^3 - 2(n-1)(2n-3)\tilde{p}^4}{n^5} \end{aligned}$	(40)
Degree of freedom of a Student t distribution whose kurtosis matches the studentized binomial distribution with c successes and c failures added	(42)	$dfg(c) = 4 \frac{6}{kur(x, n, c) - 3}$	(41)
Kurtosis of the studentized binomial distribution $\mathcal{B}\left(n, \frac{x}{n}\right)$ with c successes and c failures added	(13)(43)	$kur(x, n, c) = \frac{\sum_{i=1}^{n-1} BPF(i; n, x)(y_i - \bar{y})^4}{\left(\sum_{i=1}^{n-1} BPF(i; n, x)(y_i - \bar{y})^2\right)^2}$	(42)
Studentized proportion for i success in a binomial distribution $\mathcal{B}\left(n, \frac{x}{n}\right)$ with c successes and c failures added		$y_i = \frac{i - x}{\sqrt{\frac{(i+c)(n-i+c)}{n+2c}}}$	(43)
Like dfp but \tilde{p} substitutes \hat{p} and $\tilde{n} = n + \kappa^2$ substitutes n .	(22)(40)	$df\tilde{p} = \frac{2 \left(\frac{\tilde{p}(1-\tilde{p})}{n + \kappa^2} \right)^2}{\Omega(\tilde{p}, n + \kappa^2)}$	(44)
Poisson approximation threshold of Wilson's interval modified by Brown ²		$x^* = \begin{cases} 2 & \text{when } n \leq 50 \\ 3 & \text{when } n > 50 \end{cases}$	(45)

2.3 Intervals based on normal approximation after transformation

Table 3 shows intervals based on normal approximations after transformations stabilizing variance or improving the normal approximation.

The logit transformation and its reciprocal logitinv are defined below

$$\text{logit}(\hat{p}) = \log\left(\frac{\hat{p}}{1 - \hat{p}}\right) \quad (46)$$

$$\text{logitinv}(t) = \frac{\exp(t)}{1 + \exp(t)} \quad (47)$$

The logistic Anscombe transformation la and its reciprocal lainv are defined below

$$\text{la}(\hat{p}, n) = \log\left(\frac{n\hat{p} + \frac{1}{2}}{n(1 - \hat{p}) + \frac{1}{2}}\right) \quad (48)$$

$$\text{lainv}(t, n) = \frac{\left(n + \frac{1}{2}\right) \exp(t) - \frac{1}{2}}{n(1 + \exp(t))} \quad (49)$$

The Arc-Sine variance stabilizing transformation, and a slight modification, adding a half success and half failure have been described by Bartlett ¹⁴. Improved Arc-Sine transformations that better stabilize variance have been described by Anscombe in 1948 ¹⁵, Freeman and Tukey in 1950 ¹⁶ and have been reviewed by Yu in 2009 ¹⁷. The modified Wald logit interval is equivalent to Wald's interval of a logistic regression except when $\min(x, n - x) = 0$ where the Clopper-Pearson interval has been substituted. Rubin logit+0.5 interval is equivalent to adding a half success and a half failure before computing the Wald logit interval.

Table 3 : Definition of the lower bounds of confidence intervals based on normal approximation after variance-stabilizing transformation, the upper bounds being defined by equivariance $U_{1-\alpha}(x, n) = 1 - L_{1-\alpha}(n - x, n)$.

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	
¹⁴ Arc-Sine	(22)	$\begin{cases} \sin^2 \left(\operatorname{asin} \left(\sqrt{\frac{x}{n}} \right) - \frac{\kappa}{2\sqrt{n}} \right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(50)
^{14,15} Bartlett Arc-sine	(22)	$\sin^2 \left(\max \left(0, \operatorname{asin} \left(\sqrt{\frac{x + \frac{1}{2}}{n + 1}} \right) - \frac{\kappa}{2\sqrt{n + \frac{1}{2}}} \right) \right)$	(51)
^{18,19} Anscombe Arc-Sine	(22)	$\sin^2 \left(\max \left(0, \operatorname{asin} \left(\sqrt{\frac{x + \frac{3}{8}}{n + \frac{3}{4}}} \right) - \frac{\kappa}{2\sqrt{n + \frac{1}{2}}} \right) \right)$	(52)
¹⁰ Anscombe Arc-Sine with cc	(22)	$\sin^2 \left(\operatorname{asin} \left(\sqrt{\max \left(0, \frac{x - \frac{1}{8}}{n + \frac{3}{4}} \right)} \right) - \frac{\kappa}{2\sqrt{n + \frac{1}{2}}} \right)$	(53)
¹⁶ Freeman-Tukey Arc-Sine	(22)	$\begin{cases} \sin^2 \left(\frac{1}{2} \left(\operatorname{asin} \left(\sqrt{\frac{x}{n + 1}} \right) + \operatorname{asin} \left(\sqrt{\frac{x + 1}{n + 1}} \right) - \frac{\kappa}{\sqrt{n + \frac{1}{2}}} \right) \right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(54)
² Modified Wald logit	(22)(51)	$\begin{cases} \operatorname{logitinv} \left(\log \left(\frac{x}{n - x} \right) - \kappa \sqrt{\frac{n}{x(n - x)}} \right) & \text{if } 0 < x < n \\ \sqrt{n\alpha/2} & \text{if } x = n \\ 0 & \text{if } x = 0 \end{cases}$	(55)
^{2,18} Anscombe logit	(22)(52)(53)	$\max \left(0, \operatorname{lainv} \left(\operatorname{la} \left(\frac{x}{n}, n \right) - \kappa \sqrt{\frac{(n + 1)(n + 2)}{n(x + 1)(n - x + 1)}}, n \right) \right)$	(56)
^{2,20} Rubin logit+0.5	(22)(51)(52)	$\begin{cases} \operatorname{logitinv} \left(\operatorname{la} \left(\frac{x}{n}, n \right) - \kappa \sqrt{\frac{n + 1}{\left(x + \frac{1}{2}\right) \left(n - x + \frac{1}{2}\right)}} \right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(57)

2.4 Bayesian intervals

These Bayesian confidence intervals are slightly modified Bayesian credible intervals and are analyzed as frequentist confidence intervals. We denote

$$\beta QJ(q) = \beta iCDF\left(q; x + \frac{1}{2}, n - x + \frac{1}{2}\right) \quad (58)$$

the q quantile of the *posterior* distribution for a beta Jeffreys *prior*. Table 4 shows Bayesian confidence intervals based on these definitions.

For zero successes or zero failures, Bayesian credible intervals would not contain the observed proportion of successes: therefore, the boundary is set to 0 or 1 in order to get good frequentist properties. Brown made an *ad hoc* modification of the equal-tailed Jeffreys interval in order to remove low coverage spikes. Brown modification is based on the Clopper-Pearson confidence interval for zero successes or zero failures. The shortest Bayesian confidence interval with Jeffreys prior is based on the same beta distribution as the equal-tailed Jeffreys interval but the confidence interval used different quantiles of the beta distribution in order to reduce the interval length as much as possible while keeping the two-sided Bayesian credibility level.

Table 4: Definition of the lower bounds of Bayesian confidence intervals, the upper bounds being defined by equivariance $U_{1-\alpha}(x, n) = 1 - L_{1-\alpha}(n - x, n)$.

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	
² Equal-tailed Jeffreys	(17)	$\begin{cases} \beta iCDF\left(\frac{\alpha}{2}; x + \frac{1}{2}, n - x + \frac{1}{2}\right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(59)
² Modified equal-tailed Jeffreys	(17)	$\begin{cases} \beta iCDF(\alpha/2; x + 1/2, n - x + 1/2) & \text{if } 2 \leq x < n \\ \sqrt[n]{\alpha/2} & \text{if } x = n \\ 0 & \text{if } x \leq 1 \end{cases}$	(60)
¹⁰ Equal-tailed, uniform prior	(17)	$\beta iCDF\left(\frac{\alpha}{2}; x + 1, n - x + 1\right)$	(61)
Shortest, Jeffreys prior	(58)	$\begin{cases} \beta QJ(\inf\{\beta QJ(1 - \alpha + r) - \beta QJ(r) r \in [0, \alpha]\}) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(62)

2.5 Bootstrap intervals

Bootstrap confidence intervals shown in Table 5, with auxiliary functions defined in Table 6, are computed from the exact binomial distribution so that, they are equivalent to the non-parametric bootstrap intervals with an infinite number of resampling.

The bootstrap distribution is the binomial distribution $\mathcal{B}\left(n, \frac{x}{n}\right)$, so the percentile and basic bootstrap intervals can be computed from the binomial cumulative distribution function. Studentization is a monotonous transformation for the binomial distribution, making it possible to compute the pivot statistic quantiles from the binomial quantiles function. The studentized proportions are undefined for zero successes or zero failures, making the studentized interval impossible to compute when $\max\left(BPF\left(0; n, \frac{x}{n}\right), BPF\left(n; n, \frac{x}{n}\right)\right) > \frac{\alpha}{2}$. For 95% confidence interval, the condition is $\min(x, n - x) \geq 5$.

The bias corrected and accelerated (BC_a) bootstrap interval as computed from Efron²¹ method, is not equivariant. This is due to the definition of the empirical cumulative distribution function $\hat{G}(s) = Pr_{\hat{\theta}}(\hat{\theta}^* < s)$ shown in equation 3.2 of Efron's article. This $\hat{G}(s)$ function is based on an inequality $\hat{\theta}^* < s$ which is equivalent to $\hat{\theta}^* \leq s$ for continuous bootstrap distributions. The binomial distribution being discrete, this should be changed to $\hat{G}(s) = Pr_{\hat{\theta}}(\hat{\theta}^* < s) + \frac{1}{2}Pr_{\hat{\theta}}(\hat{\theta}^* = s)$. This modification gives the equivariance property $L_{1-\alpha}(x, n) = 1 - U_{1-\alpha}(n - x, n)$.

As for percentile, basic and studentized bootstrap intervals, the BC_a bootstrap interval is computed from the exact binomial distribution, getting an exact $\hat{\theta}^*$ distribution as if an infinite number of bootstrap samples had been drawn.

It should be noted that the bounds of the basic, percentiles, BC_a and modified BC_a intervals, are always integer multiples of $\frac{1}{n}$.

The normal bootstrap interval computed from the exact binomial distribution as if an infinite number of samples had been drawn, is equal to Wald's interval.

In order to check these simplified formulae, calculations of the exact binomial bootstrap intervals were compared to the approximate R 'boot' package computations on a finite number of bootstrap samples. Results were identical on the significant decimal places.

Table 5 : Definition of bootstrap confidence intervals

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	Upper bound $U_{1-\alpha}(x, n)$
²² Percentile bootstrap	(14)	$\frac{BiCDF\left(\frac{\alpha}{2}; n, \frac{x}{n}\right)}{n}$	(63) equivariant
²² Basic bootstrap	(14)	$\max\left(0, \frac{2x - BiCDF\left(1 - \frac{\alpha}{2}; n, \frac{x}{n}\right)}{n}\right)$	(64) equivariant
²² Studentized bootstrap	(14)	$\hat{p} - (l - \hat{p})\sqrt{\frac{\hat{p}(1-\hat{p})}{l(1-l)}}$ with $l = \frac{1}{n}BiCDF(1 - \frac{\alpha}{2})$ <i>Undefined for $\min(x, n-x) \leq 4$ if $\alpha = 0.05$</i>	(65) equivariant
²¹ BC _a bootstrap	(69)	$EF\left(\frac{\alpha}{2}, n, x, 0\right)$ <i>Undefined if $x = 0$ or $x = n$</i>	(66) $EF\left(1 - \frac{\alpha}{2}, n, x, 0\right)$ <i>Undefined if $x = 0$ or $x = n$</i> (67)
²¹ Modified BC _a bootstrap	(69)	$EF\left(\frac{\alpha}{2}, n, x, 1\right)$ <i>Undefined if $x = 0$ or $x = n$</i>	(68) equivariant
Smoothed BC _a bootstrap		Non parametric BC _a bootstrap with random $\mathcal{N}\left(0, \left(\frac{1}{2}\right)^2\right)$ normal noise added to X . The acceleration constant is computed on the distribution without noise. <i>Undefined if $x = 0$ or $x = n$</i>	

Table 6 : Auxiliary functions for Table 5.

Function description	Depends on equation	Function definition
BC _a boundary based on Efron ²¹	(20)(70)	$EF(q, n, x, d) = \frac{1}{n}BiCDF\left(NCDF\left(bias(x, n, d) + \frac{bias(x, n, d) + NiCDF(q)}{1 - acc(x, n) \times (bias(x, n, d) + NiCDF(q))}\right); n, \frac{x}{n}\right)$ (69)
Based on equation 4.1 page 174 Efron ²¹	(12)(13)(21)	$bias(x, n, d) = NiCDF\left(BCDF\left(x - 1; n, \frac{x}{n}\right) + \frac{d}{2}BPF\left(x; n, \frac{x}{n}\right)\right)$ (70)
Based on equation 4.4 page 174 Efron		$acc(x, n) = \frac{n - 2x}{6\sqrt{nx(n-x)}}$ (71)

2.6 Exact binomial confidence intervals

These intervals are based on the exact binomial distribution. Like any deterministic interval, they have coverage oscillations. They are described in Table 7 with auxiliary functions defined in Table 8.

The Sterne and Blaker intervals are based on non-contiguous confidence regions. The intervals are defined as the shortest interval containing the confidence region (convex hull). The Clopper-Pearson interval can be defined as

$$\left\{q \mid \min(BCDF(x; n, q), 1 - BCDF(x - 1; n, q)) \geq \frac{\alpha}{2}\right\} \quad (72)$$

which is equal to $\beta iCDF\left(\frac{\alpha}{2}; x, n - x + 1\right)$. The interval described by Blaker²³ in 2000 had been previously proposed by Cox and Hinkley in 1974²⁴. It is a two-sided unequal tailed interval that is always containing the Clopper-Pearson interval, unlike Sterne's interval that is sometimes longer. Sterne's interval, like Blaker's interval is nested: the 90% confidence interval is always contained in the 95% confidence interval.

The Schilling-Doi²⁵ interval is a two-sided unequal-tailed interval based on Sterne acceptance regions. Acceptance regions can be interpreted as prediction intervals. For a theoretical p proportion, an acceptance region AR is an interval of x values so that $P(X \in AR) \geq 1 - \alpha$ where $X \sim \mathcal{B}(n, p)$. Sterne acceptance regions optimize the following criteria from most important to least important: Minimal extent, highest coverage probability and highest lower bound. Schilling-Doi modified Sterne acceptance regions when an acceptance region bound (lower or upper) is not monotonous along p . Schilling-Doi algorithm memory usage and computing time is exponential with the number of requested decimal places. When analyzing the expected number of successes $\lambda = np$, the precision required on p is proportional to the sample size. This makes Schilling-Doi algorithm unpractical to compute on sample sizes of a few thousands observations. Based on the knowledge that Sterne acceptance regions, analyzed as functions of the theoretical proportion p , may only change (along p) when likelihood curves (one being defined for each x value between 0 and n) cross the confidence level or cross one another, an updated algorithm based on finding those hot points has been written.

The Wang interval is based on sequential squeezing of Clopper-Pearson intervals²⁶ while keeping the confidence interval set strictly conservative. Like Sterne, Schilling-Doi and Blaker, it's a strictly conservative unequal-tailed interval. Wang's algorithm is slow. It moves interval boundaries at small steps proportional to the decimal precision requested. At each step, it computes the actual coverage for n theoretical values while only one confidence interval has changed. The algorithm has been modified to update only coverage values that may change between two steps. Numerical results are identical. Both Wang and Shilling-Doi intervals, for a specified sample size, tend to shorten the arithmetic mean of all confidence interval lengths but not necessarily the geometric mean.

The Blyth-Still-Casella interval has been defined by Blyth and Still with a fixed decimal precision algorithm²⁷. Casella improved the algorithm to get better decimal precision²⁸. Winstein's implementation in the C++ programming language has been used²⁹.

The Zieliński interval³⁰ defined in 2009, is based on changes of the Clopper-Pearson formula, adding a bias in the beta distribution quantiles. It's not strictly conservative as Zieliński acknowledged in 2017³¹. This is due to the dependence of the γ bias to the number of successes x . Zieliński proposed

a randomized unequal-tailed shortest confidence interval in 2017³¹. This randomized interval is not presented here.

The exact likelihood ratio confidence set³² is defined by inversion of an exact likelihood ratio test. The confidence set is not always contiguous, as the test p-value is not bi-monotonous. The confidence interval is the convex hull of the confidence set³³.

The exact score interval defined by Sakakibara or Thulin^{32,33} is equivalent to the alternative exact confidence interval defined by Cai and Krishnamoorthy in equation 2³⁴. Indeed

$$Pr\left(\frac{(X - np)^2}{np(1 - p)} \geq \frac{(k - np)^2}{np(1 - p)} \middle| p\right) = Pr((X - np)^2 \geq (k - np)^2 | p) \quad (73)$$

where $X \sim \mathcal{B}(n, p)$. The confidence set is not always contiguous. The confidence interval is the convex hull of the confidence set. The Cai-Krishnamoorthy combined interval is based on inversion of the more liberal of the Clopper-Pearson or exact-score tests³⁴. The confidence interval is the convex hull of the confidence set. The procedure is not strictly conservative but is mostly conservative.

The Pratt interval was described by Blyth¹² in 1986 as the Paulson-Camp-Pratt approximation (Equation D); this is an approximation of the Clopper-Pearson interval. The mean Pratt interval was proposed by Vollset⁸ in 1993 as an approximation of the Clopper-Pearson mid-P interval. It is based on the average of two Pratt intervals, for x and for $x + 1$. It is not based on the exact binomial distribution but it approximates an interval constructed as such. Minimal adjustments to the Vollset formula have been applied to avoid aberrant behaviors when $x = 0$ or $x = n$. Rather than evaluating the Pratt interval at x and $x + 1$, it is evaluated at x and $\min(x + 1, n)$. Moreover, when $x = 0$, the lower bound is replaced by 0 and conversely, the upper bound is replaced by 1 when $x = n$.

Let us define the *absolute odds ratio* between two values a and b both between 0 and 1 as $\exp(|\text{logit}(a) - \text{logit}(b)|)$. That is, if the odds ratio OR between a and b is greater than 1, the absolute odds ratio is equal to OR , otherwise it is equal to its inverse $1 / OR$. The Clopper-Pearson mid-P approximation by the mean Pratt interval is fairly good since for $n = 32$ and $x = 0, \dots, 32$, the mean absolute odds ratio between the lower bounds of the Clopper-Pearson mid-P and mean Pratt intervals is 1.08. The approximation was poor for $x = 1$ (Absolute odds ratio = 2.41) and middling for $x = 2$ (Odds absolute ratio = 1.25) and for $x = 32$ (Absolute odds ratio = 1.24). Two modifications improve these approximations. The first modification is the use of the logistic mean between Pratt's bounds for x and $\min(x + 1, n)$, that is,

$$\text{logitinv}\left(\frac{\text{logit}(LPratt_{1-\alpha}(x, n)) + \text{logit}(LPratt_{1-\alpha}(\min(n, x + 1), n))}{2}\right) \quad (74)$$

where logitinv is the reciprocal of the logit function. The second modification is the use of the $LPratt_{1-2\alpha}(x, n)$ lower boundary when $x = n$ rather than $LPratt_{1-\alpha}(x, n)$. This *modified mean Pratt interval*, for $n = 32$, had a mean absolute odds ratio of the lower bound equal to 1.027 compared to the Clopper-Pearson mid-P interval with a maximal absolute odds ratio equal to 1.11. This modified mean Pratt interval is not shown in the results, but its behavior would be very similar to that of Clopper-Pearson mid-P.

Problems of the two-sided unequal-tailed intervals have been described by Vos and Hudson in 2008, then Thulin and Zwanzig in 2017^{33,35}.

Table 7 : Definition of the lower bounds of exact binomial confidence intervals, the upper bounds being defined by equivariance $U_{1-\alpha}(x, n) = 1 - L_{1-\alpha}(n - x, n)$.

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	
³⁶ Sterne	(12)	$\inf\{q \text{stpval}(x, n, q) > \alpha\}$	(75)
²³ Blaker	(88)	$\inf\{q \text{bpval}(q, x, n) > \alpha\}$	(76)
^{2,37} Clopper-Pearson	(17)	$\beta iCDF\left(\frac{\alpha}{2}; x, n - x + 1\right)$	(77)
^{38,39} Clopper-Pearson mid-P	(12)(89)	$\inf\{q \text{cpval}(x, n, q) - BPF(x; n, q) > \alpha\}$	(78)
⁸ mean Pratt (Vollset 1993)	(90)	$\begin{cases} \frac{LPratt_{1-\alpha}(x, n) + LPratt_{1-\alpha}(\min(n, x + 1), n)}{2} & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(79)
²⁵ Schilling-Doi		Inversion of Sterne acceptance regions updated to get monotonous acceptance boundaries along the theoretical proportion p	(80)
²⁶ Wang 2014		Squeezing Clopper-Pearson intervals step down from $x = \frac{n}{2}$ to $x = 0$ while keeping the two-sided interval strictly conservative while using equivariance to complete the interval definition for $x = \frac{n}{2} + 1$ to $x = n$.	(81)
²⁷⁻²⁹ Blyth-Still-Casella		Based on a set of shortest-length acceptance regions of X for every theoretical proportion	(82)
³⁰ Zieliński	(91)(92)	$\begin{cases} \beta QCL(\inf\{\beta QCU(1 - \alpha + \gamma) - \beta QCL(\gamma) \gamma \in [0, \alpha]\}) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(83)
³² Exact likelihood ratio	(93)	$\inf\{q \text{lrpval}(x, n, q) > \alpha\}$	(84)
^{32,40} Exact score test	(95)	$\inf\{q \text{spval}(x, n, q) > \alpha\}$	(85)
³⁴ Cai-Krishnamoorthy combined	(89)	$\inf\{q \min(\text{cpval}(x, n, q), \text{spval}(x, n, q)) > \alpha\}$	(86)

^a $\beta iCDF(q; \alpha, \beta)$ is the q quantile of the beta distribution with α and β shape parameters.

Table 8 : Auxiliary functions for Table 7

Function description	Depends on equation	Function definition	
Sterne's P-value	(12)	$\text{stpval}(x, n, p) = \sum_{k \in \{i BPF(i; n, q) \leq BPF(x; n, q)\}} BPF(k; n, p)$	(87)
Blaker's P-value	(13)(14)	$\text{bpval}(q, x, n) = \begin{cases} 1, BPDF(x; n, q) + \\ \min \left(1 - BPDF(BiCDF(1 - BPDF(x; n, q); n, q); n, q) \right) \text{ if } q \geq x/n \\ \text{bpval}(1 - q, n - x, n) \text{ if } q < x/n \end{cases}$	(88)
Clopper-Pearson P-value	(13)	$\text{cpval}(x, n, p) = 2 \times \min \left(\frac{1}{2}, BPDF(x; n, p), 1 - BPDF(x - 1, n, p) \right)$	(89)
Pratt's lower bound		$LPratt_{1-\alpha}(x, n) = 1 - \left(\frac{1 + \left(\frac{n - x + 1}{x} \right)^2}{\times \left(\frac{81x(n - x + 1) - 9n - 8 - 3\kappa \sqrt{9x(n - x + 1)(9n + 5 - \kappa^2) + n + 1}}{81(n - x + 1)^2 - 9(n - x + 1)(2 + \kappa^2) + 1} \right)^3} \right)^{-1}$	(90)
Clopper-Pearson like lower bound	(17)	$\beta QCL(q) = \beta iCDF(q; x, n - x + 1)$	(91)
Clopper-Pearson like upper bound	(17)	$\beta QCU(q) = \beta iCDF(q; x + 1, n - x)$	(92)
Exact likelihood ratio P-value	(12)	$\text{lrpval}(x, n, p) = \sum_{k \in \{i LIKR(i, n, p) \geq LIKR(x, n, p)\}} BPF(k; n, p)$	(93)
Likelihood ratio statistic	(12)	$LIK R(x, n, p) = \frac{BPF\left(x, n, \frac{x}{n}\right)}{BPF(p, n, p)}$	(94)
Exact score P-value	(12)	$\text{spval}(x, n, p) = \sum_{k \in \{i (i - np)^2 \geq (x - np)^2\}} BPF(k; n, p)$	(95)

2.7 Skewness corrected normal approximation intervals

These intervals are based on normal approximations with skewness bias corrections making them equal-tailed two-sided intervals. Table 9 defines these intervals and Table 10 contains auxiliary functions.

Kott-Liu interval ⁴¹ updates Hall's interval variance but doesn't change its center.

Cai's interval is similar to Hall's interval but keeps terms of smaller asymptotic orders.

Table 9 : Definition of the lower bounds of skewness corrected closed-from confidence intervals, the upper bounds being defined by equivariance $U_{1-\alpha}(x, n) = 1 - L_{1-\alpha}(n - x, n)$.

Name	Depends on equation	Lower bound $L_{1-\alpha}(x, n)$	
⁴² Hall 1982	(22)(99)	$\begin{cases} \max\left(0, \hat{p} + \delta - \kappa \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(96)
⁴³ Cai 2005	(22)(100)(101)	$\begin{cases} \max\left(0, \check{p} - \kappa \sqrt{\frac{\hat{p}(1-\hat{p}) + \gamma}{n}}\right) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$	(97)
⁴¹ Kott-Liu	(22)(99)	$\max\left(0, \hat{p} + \delta - \sqrt{\kappa^2 \frac{\hat{p}(1-\hat{p})}{n-1} + \delta^2}\right)$	(98)

Table 10 : Auxiliary functions for Table 7

Function description	Depends on equation	Function definition	
Hall center adjustment	(22)	$\delta = \frac{\left(\frac{\kappa^2}{3} + \frac{1}{6}\right)(1-2\hat{p})}{n}$	(99)
Cai adjusted center	(22)	$\check{p} = \frac{n\hat{p} + \left(\frac{\kappa^2}{3} + \frac{1}{6}\right)}{n + 2\left(\frac{\kappa^2}{3} + \frac{1}{6}\right)}$	(100)
Cai variance adjustment	(22)	$\gamma = \frac{2\kappa^2 + 7 - \hat{p}(1-\hat{p})(26\kappa^2 + 34)}{36n}$	(101)

2.8 Generalized linear model intervals

Logistic, Poisson and log-binomial regressions are generalized linear models (GLM). These may be multivariate, bivariate or univariate. Two estimation methods are widely used. The first is Wald's method, $B \pm z_{1-\alpha/2} \times SE$ where B is the raw model coefficient, such as $\log(\text{odds ratio})$ for a logistic regression and SE is the standard error, got from the variance-covariance model matrix. It's based on normal approximation of the regression coefficient. The second method is inversion of a χ^2 likelihood-ratio test. Likelihood-ratio test inversion may be performed by the profile-likelihood algorithm ⁴⁴. The R and SAS statistical software use the profile-likelihood confidence intervals by default for generalized linear model. The R software has been used to compute confidence intervals of a proportion through intercept-only GLMs. Both Wald's and likelihood-ratio (LR) intervals have been computed. The interval estimators presented in Table 11 have been analyzed. Since the intervals were not calculable for zero successes, the Clopper-Pearson interval was used in this case. When the upper bound of the confidence interval exceeded 100%, for the Poisson regression or the log-binomial regression, it was replaced by 100%. The log-binomial and Poisson intervals are not equivariant.

Table 11 : Definition of generalized linear models confidence interval estimators computed in the R statistical software.

Name	Description	R software command
LR logit GLM	Logistic regression profile-likelihood interval	<code>confint(glm(family=binomial, cbind(x, n-x)~1))</code>
LR Poisson GLM	Poisson regression profile-likelihood interval	<code>confint(glm(family=poisson, x~offset(log(n))))</code>
Wald logit GLM	Logistic regression Wald's interval	<code>confint.default(glm(family=binomial, cbind(x, n-x)~1))</code>
Wald log binom GLM	Log-binomial regression Wald's interval	<code>confint.default(glm(family=binomial(log), cbind(x, n-x)~1))</code>
Wald Poisson GLM	Poisson regression Wald's interval	<code>confint.default(glm(family=poisson, x~offset(log(n))))</code>

The likelihood ratio interval of a logistic regression "LR logit GLM, CP if k = 0" provided identical results to the "modified likelihood ratio" method that had been defined in the main analysis by the following formula:

$$L_{1-\alpha} = \begin{cases} \inf \left\{ q \left| \log \left(\left(\frac{x}{nq} \right)^x \left(\frac{n-x}{n(1-q)} \right)^{n-x} \right) \leq \frac{1}{2} \kappa^2 \right\} & \text{if } x < n \\ \sqrt{n\alpha/2} & \text{if } x = n \text{ (i.e. Clopper - Pearson)} \end{cases} \quad (102)$$

Moreover, the unmodified likelihood ratio method is presented too:

$$L_{1-\alpha} = \inf \left\{ q \left| \log \left(\left(\frac{x}{nq} \right)^x \left(\frac{n-x}{n(1-q)} \right)^{n-x} \right) \leq \frac{1}{2} \kappa^2 \right\} \quad (103)$$

It's defined for all x even for $x = 0$ or $x = n$.

3 Wald's validity condition

New validity conditions of Wald's interval have been sought. When Wald's interval validity threshold is not reached, the statistician is assumed to compute a less approximate interval such as Clopper-Pearson mid-P. Indeed, not reporting the interval at all when the validity threshold is not reached, would overestimate proportions when they are shown (statistical inference bias). Consequently, we need to evaluate the two steps procedure: if the number of successes is above a threshold, Wald's confidence interval is computed; otherwise, Clopper-Pearson mid-P interval (or Clopper-Pearson for conditional errors) is computed.

3.1 Informal description of the method

The validity conditions of the Wald interval estimator will be defined to control that one-sided local average errors, do not exceed 1.50 times the nominal error, i.e. 0.0375 for a nominal one-sided error at 0.025. The conditions will apply to the observed number of successes and failures as validity conditions applying to the actual proportion cannot be used in practice because the actual proportion is unknown. In order to define a validity condition, it is necessary to specify the procedure to be applied when the condition is not verified. Therefore, it will be assumed that the Clopper-Pearson mid-P interval will be used when Wald's interval validity conditions are not satisfied. The evaluation will focus on a hybrid estimator consisting of the use of one or the other of the interval estimators according to whether a threshold is exceeded or not. For a given sample size n , the validity threshold ξ'' will be defined as the smallest threshold of the hybrid procedure which guarantees one-sided local average errors less than 0.0375 for a nominal one-sided error equal to 0.025. An empirical table of validity thresholds ξ'' will be built for various sample sizes from 2^5 to 2^{11} and the limit case where $n \rightarrow +\infty$. Actual errors will be computed for hybrid estimators and a dichotomy algorithm will find the smallest threshold keeping the actual errors below 0.0375.

The table of validity thresholds being impractical to memorize in practice, a formula approximating the values of this table will be sought. The approximate formula will be sought empirically in order to get thresholds as close as possible to the values of the table. If there is a discrepancy between the values, a conservative formula will be preferred, that is, a formula providing slightly higher validity thresholds. The elaboration of this formula will not necessarily be based on a mathematical theory since it is only a matter of getting close to the values of the table. Simplified formulas will be sought in order to make their teaching possible in practice. Although the maximum tolerable error threshold 0.0375 will serve as a reference, other thresholds will also be analyzed: 0.030, 0.035 and 0.05. By expressing these thresholds in the form of equal-tailed two-sided errors they will be equal to 0.060, 0.070 and 0.10. Although the main analysis will focus on local average errors, a secondary analysis will be carried out on the validity thresholds controlling the maximum conditional errors for a Wald/Clopper-Pearson hybrid interval. Hybridization of Wald's interval with the Clopper-Pearson mid-P interval, for one-sided local average errors control, and with the Clopper-Pearson interval, for one-sided conditional errors, has been chosen for the good behavior of these two intervals on the judgment criteria involved in the analysis.

3.2 Formal description of the method

Let $ICW_{1-\alpha}$ be Wald's confidence interval estimator and $ICM_{1-\alpha}$ the Clopper-Pearson mid-P confidence interval estimator. For an integer threshold χ , we define the hybrid confidence interval estimator $ICMW_{1-\alpha}(\chi)$ equal to $ICM_{1-\alpha}$ if $\min(x, n-x) < \chi$ and $ICW_{1-\alpha}$ if $\min(x, n-x) \geq \chi$. For a nominal controlled local average error α''_{max} , we define the validity threshold $\xi''(\alpha''_{max}, n)$ as

the smallest χ such as for all p_0 , the errors $\alpha'_l(n, p_0, \alpha)$ and $\alpha'_u(n, p_0, \alpha)$ of the $ICMW_{1-\alpha}(\chi)$ estimator are smaller or equal to $\alpha''_{max}/2$. For this analysis, OR_S will be set to 1.10, so that the logit of the actual proportion $\text{logit}(P)$ is a normal distribution with a standard deviation $\log(OR_S)$ and the expectation of P is p_0 .

$$\xi''(\alpha''_{max}, n) = \min \left\{ \chi \left| \begin{array}{l} \forall p_0 \in [0; 1] \ Pr(ICMW_{1-\alpha}(\chi) < P) \leq \alpha''_{max}/2 \text{ and } \\ \Pr(ICMW_{1-\alpha}(\chi) > P) \leq \alpha''_{max}/2 \\ \text{where } \text{logit}(P) \sim \mathcal{N}(\mu, (\log(OR_S))^2) \text{ and } E(P) = p_0 \end{array} \right. \right\} \quad (104)$$

Similarly, we define the $ICCW(\chi)$ hybrid estimator between the Clopper-Pearson estimator and Wald's estimator at threshold χ . The conditional validity threshold $\xi'(\alpha'_{max}, n)$ of the $ICCW(\chi)$ hybrid estimator is defined as the smallest χ threshold such that, for all theoretical proportions p , the one-sided conditional errors $\alpha'_l(n, p, \alpha)$ and $\alpha'_u(n, p, \alpha)$ are both less than or equal to $\alpha'_{max}/2$.

$$\xi'(\alpha'_{max}, n) = \min \left\{ \chi \left| \begin{array}{l} \forall p \in [0; 1] \ Pr(ICCW_{1-\alpha}(\chi) > p) \leq \alpha'_{max}/2 \text{ and } \\ \Pr(ICCW_{1-\alpha}(\chi) < p) \leq \alpha'_{max}/2 \end{array} \right. \right\} \quad (105)$$

Firstly, a table of validity ξ' and ξ'' thresholds will be numerically computed according to the sample size n and the maximum α''_{max} and α'_{max} tolerance. These thresholds will be found by a dichotomy algorithm. Secondly, two empirical validity formulas will be determined in order to approximate the ξ' and ξ'' thresholds by $\tilde{\xi}'$ and $\tilde{\xi}''$ thresholds. These empirical formulas will be built to fit the table data as well as possible while keeping them as simple as possible.

3.3 Results

3.3.1 Local average errors control

Table 11 presents the validity thresholds ξ'' and their approximations $\tilde{\xi}''$ obtained by an empirical formula specifically constructed to best fit the ξ'' numerical values observed in the table. For instance, the fourth column of this table ($\alpha''_{max} = 0.075$) corresponds to a one-sided local average errors that do not exceed 1.50 times the nominal error, i.e. 0.0375 for the lower boundary and 0.0375 for the upper boundary. Thus, for a sample of 512 observations $\xi'' = 31$. This means that at least 31 successes and 31 failures are required to apply the Wald interval estimator for a sample size 512 in order to keep the actual one-sided local average errors below 1.50 times the nominal error. When either the number of successes or the number of failures is less than this threshold, the Clopper-Pearson mid-P interval estimator must be used. For the limit case $n \rightarrow +\infty$ corresponding to Poisson distributions, a number of successes or failures greater or equal to 41 is needed to keep Wald's interval actual one-sided local average errors below 1.50 times the nominal one-sided error.

Applying the following validity condition:

$$nx(n-x) \geq \xi''(\alpha''_{max}, \infty) \times (n-2x)^2 \quad (106)$$

we find all the thresholds in parentheses in Table 12. If 1.50 times the nominal error is accepted (fourth column of the table), the threshold $\xi''(\alpha''_{max}, \infty)$ is equal to 41 and the formula becomes:

$$nx(n-x) \geq 41(n-2x)^2 \quad (107)$$

This formula is a threshold of the absolute value of the skewness which is equal to $\left| \frac{1-2p}{\sqrt{np(1-p)}} \right|$ for a binomial proportion, estimated as $\left| \frac{n-2x}{\sqrt{nx(n-x)}} \right|$ on the sample. By squaring this skewness, we can find a

condition of the form $\frac{(n-2x)^2}{nx(n-x)} \leq threshold$; that translates to $nx(n-x) \geq \frac{1}{threshold} \times (n-2x)^2$.

Finally, the threshold is chosen such that ξ'' and its approximation $\bar{\xi}''$ converge when $n \rightarrow +\infty$. This condition, requires 7 mathematical operations (function application, multiplication, addition, division, subtraction), which is also equal to the number of operations required to compute Wald's interval itself $\left[p - 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; p + 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$. This complex condition can be simplified by applying the worst case scenario where $n \rightarrow +\infty$:

$$\min(x, n-x) > 40 \quad (108)$$

In a less formal form, the number of successes and failures must both be strictly greater than 40.

The choice not to exceed 1.50 times the nominal error presented in the fourth column of Table 12 is arbitrary. The other columns of Table 12, based on other tolerance thresholds, serve as a sensitivity analysis. This sensitivity analysis shows a great influence of the choice of α''_{max} on the validity thresholds, especially when $n \rightarrow +\infty$ (last row in the table). Another way of presenting the α''_{max} thresholds is to express them as a ratio between the maximum one-sided local average error tolerated and the nominal one-sided local average error. These ratios are, for columns 2 to 5 of Table 12, equal respectively to 1.20, 1.40, 1.50 and 2.00.

A sensitivity analysis was applied for the selection of the substitution confidence interval estimator when the validity threshold is not reached. Replacing the Clopper-Pearson mid-P interval by Bartlett's Arc-Sine interval did not change any of the ξ'' thresholds presented in Table 12 suggesting that the substitution interval does not matter as long as it has a low bias.

Table 14 summarizes validity conditions guaranteeing the control of one-sided conditional errors (α'_{max}) or one-sided local average errors (α''_{max}) for several error rate bias tolerance limits. In our opinion, the condition $\min(x, n-x) > 40$ is a good compromise between simplicity and quality of control and usefulness, but this choice is subjective.

Table 12 : Validity thresholds ξ'' of Wald's 95% confidence interval expressed as the number of successes and failures required for the application of Wald's estimator, based on the maximum tolerated two-sided equal-tailed local average error α'_{max} specified in column and the sample size n specified in row. The thresholds ξ'' specified in parentheses are approximations of the thresholds ξ'' obtained from the empirical condition $nx(n-x) \geq \xi''(\alpha'_{max}, \infty) \times (n-2x)^2$. Actual errors are calculated for a random proportion P for which the typical odds ratio is $OR_S = 1, 10$. Below these thresholds, the Clopper-Pearson mid-P confidence interval estimator must be used.

n	$\alpha'_{max} = 0.060$ $\xi'' (\xi'')$	$\alpha'_{max} = 0.070$ $\xi'' (\xi'')$	$\alpha'_{max} = 0.075$ $\xi'' (\xi'')$	$\alpha'_{max} = 0.10$ $\xi'' (\xi'')$
32	11 (14)	9 (11)	8 (10)	5 (6)
64	21 (25)	15 (18)	13 (16)	7 (8)
128	37 (43)	24 (27)	19 (22)	8 (9)
256	62 (70)	34 (38)	26 (29)	9 (10)
512	96 (107)	44 (47)	31 (34)	10 (11)
1024	136 (147)	51 (54)	36 (37)	10 (11)
2048	174 (182)	57 (58)	38 (39)	11 (11)
∞	245 (245)	63 (63)	41 (41)	11 (11)

3.3.2 Conditional errors control

Table 13 presents the validity thresholds of Wald's interval to control one-sided conditional errors. The interpretation is the same as in Table 12, but conditional errors rather than local average errors are controlled, and when the threshold is not reached, the substitution interval is the Clopper-Pearson interval (strictly conservative) rather than the Clopper-Pearson mid-P interval. It is assumed that the theoretical proportion and the sample size are constant among experiments. For instance, for a sample size 512, applying Wald's method when the number of successes and the number of failures are both above 70 (fourth column of the table) and the Clopper-Pearson method otherwise, keeps the actual one-sided conditional errors below 1.50 the nominal error rate. The empirical formula provides a slightly different threshold $\xi'' = 72$ in place of $\xi' = 70$ for the same sample size.

The empirical condition ensuring the control of the conditional error is

$$\min(x, n-x) \geq c(\alpha'_{max}) + \frac{n}{2\xi'(\alpha'_{max}, \infty) + n} \quad (109)$$

where

$$c(\alpha'_{max}) = \begin{cases} 6 & \text{si } \alpha'_{max} = 0,060 \\ 1 & \text{si } \alpha'_{max} = 0,070 \\ 0 & \text{si } \alpha'_{max} \geq 0,075 \end{cases} \quad (110)$$

and $\xi'(\alpha'_{max}, \infty)$ is given by the last line of Table 13. The value $c(\alpha'_{max})$ is an empirical adjustment constant dependent on α'_{max} , the maximum error tolerance. This empirical adjustment function $c(\alpha'_{max})$ is only applicable to the nominal error $\alpha = 0.05$. In the case where the α'_{max} tolerance is 0.075, the condition is

$$\min(x, n-x) \geq \frac{n}{198 + n} \quad (111)$$

If an error tolerance α'_{max} slightly exceeding 0.075 is tolerated, this condition may be rounded to

$$\min(x, n-x) \geq \frac{n}{200 + n} \quad (112)$$

The following simplified condition is always sufficient and is easier to teach

$$\min(x, n - x) \geq 100 \quad (113)$$

The other maximum conditional tolerated errors α'_{max} presented in columns 2, 3 and 5 of Table 13 give different results. The maximum conditional error that one is willing to tolerate has a strong influence on the validity threshold of Wald's interval.

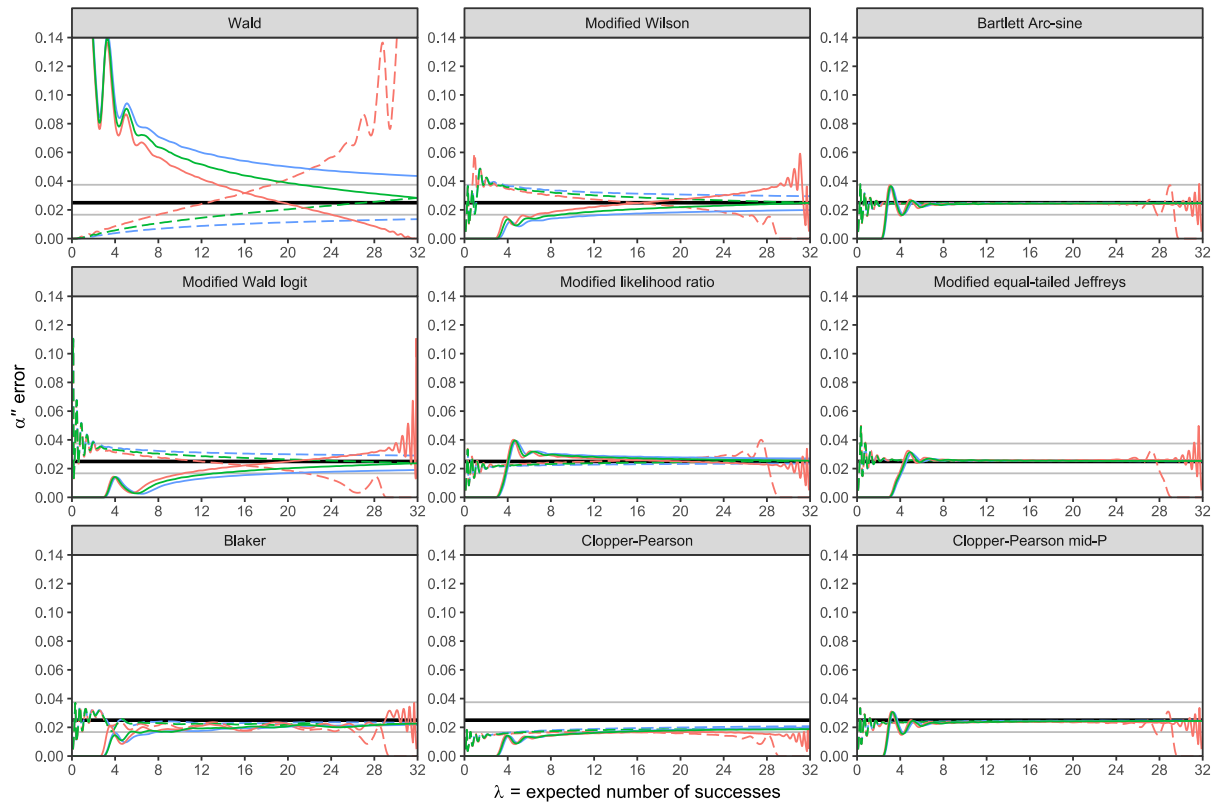
Table 13 : Validity thresholds ξ' of Wald's 95% confidence interval expressed as the number of successes and failures required for the application of Wald's estimator, based on the maximum tolerated two-sided equal-tailed conditional error α'_{max} specified in column and the sample size n specified in row. The thresholds ξ' specified in parentheses are approximations of the thresholds ξ' obtained from the empirical condition $\min(x, n - x) > c + \frac{n}{2\xi'(\alpha'_{max}, \infty) + n}$ where $c = 6$ for $\alpha'_{max} = 0.060$, $c = 1$ for $\alpha'_{max} = 0.070$ and $c = 0$ for $\alpha'_{max} \geq 0.075$. Actual errors are calculated for a constant proportion p . Below these thresholds, the Clopper-Pearson confidence interval estimator must be used.

n	$\alpha'_{max} = 0.060$ $\xi' (\xi')$	$\alpha'_{max} = 0.070$ $\xi' (\xi')$	$\alpha'_{max} = 0.075$ $\xi' (\xi')$	$\alpha'_{max} = 0.10$ $\xi' (\xi')$
32	Never (Never)	15 (16)	14 (14)	10 (11)
64	Never (Never)	27 (28)	25 (25)	14 (15)
128	62 (64)	46 (46)	39 (39)	19 (20)
256	110 (111)	69 (70)	54 (56)	22 (23)
512	181 (182)	93 (96)	70 (72)	25 (26)
1024	271 (274)	115 (117)	82 (83)	26 (27)
2048	362 (369)	130 (132)	89 (91)	27 (28)
∞	560 (566)	149 (150)	99 (99)	28 (28)

Table 14: Simplified validity conditions of Wald's 95% confidence interval according to the maximum tolerated two-sided equal-tailed conditional error α'_{max} for a constant theoretical proportion, or the maximum tolerated two-sided equal-tailed local average error α''_{max} for a random theoretical proportion P for which the typical odds ratio is $OR_S = 1.10$.

Control	Validity for $n \rightarrow \infty$	Validity for any n
$\alpha'_{max} \leq 0.060$	$\min(x, n - x) \geq 560$	$\min(x, n - x) \geq 6 + \frac{n}{1120 + n}$
$\alpha'_{max} \leq 0.070$	$\min(x, n - x) \geq 149$	$\min(x, n - x) \geq 1 + \frac{n}{298 + n}$
$\alpha'_{max} \leq 0.075$	$\min(x, n - x) \geq 99$	$\min(x, n - x) \geq \frac{n}{198 + n}$
$\alpha'_{max} \leq 0.100$	$\min(x, n - x) \geq 28$	$\min(x, n - x) \geq \frac{n}{56 + n}$
$\alpha''_{max} \leq 0.060$	$\min(x, n - x) \geq 245$	$nx(n - x) \geq 245(n - 2x)^2$
$\alpha''_{max} \leq 0.070$	$\min(x, n - x) \geq 63$	$nx(n - x) \geq 63(n - 2x)^2$
$\alpha''_{max} \leq 0.075$	$\min(x, n - x) \geq 41$	$nx(n - x) \geq 41(n - 2x)^2$
$\alpha''_{max} \leq 0.100$	$\min(x, n - x) \geq 11$	$nx(n - x) \geq 11(n - 2x)^2$

4 Figures of the nine main confidence intervals



Sample size
 — $n = 32$ — $n = 64$ — $n = 2048$

Actual local average error
 - - α''_L (lower bound) — α''_U (upper bound)

$OR_S = 1.10$

Figure 1 : One-sided local average errors of nine 95% confidence interval estimators according to different sample sizes (red for $n = 32$, green for $n = 64$ and blue for $n = 2048$), with a random true P proportion following a logit-normal distribution with a typical odds ratio of the true proportion between two experiments equal to $OR_S = 1.10$. The abscissa is the expected number of successes np_0 and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (left local average error: dashed lines) or the probability that the upper bound of the confidence interval is lower than the true proportion p (right local average error: solid line). This figure is the same as Figure I in the main article except that $OR_S = 1.10$ rather than 1.20.

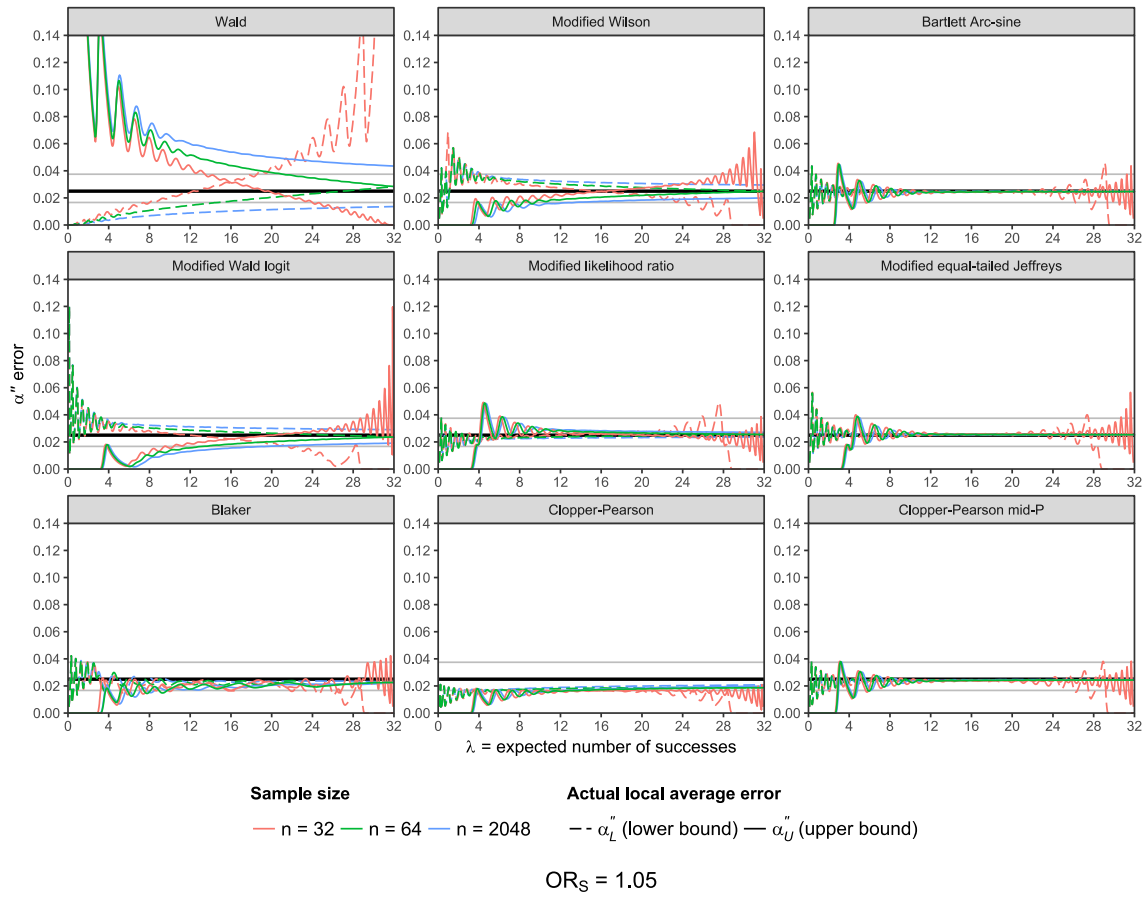
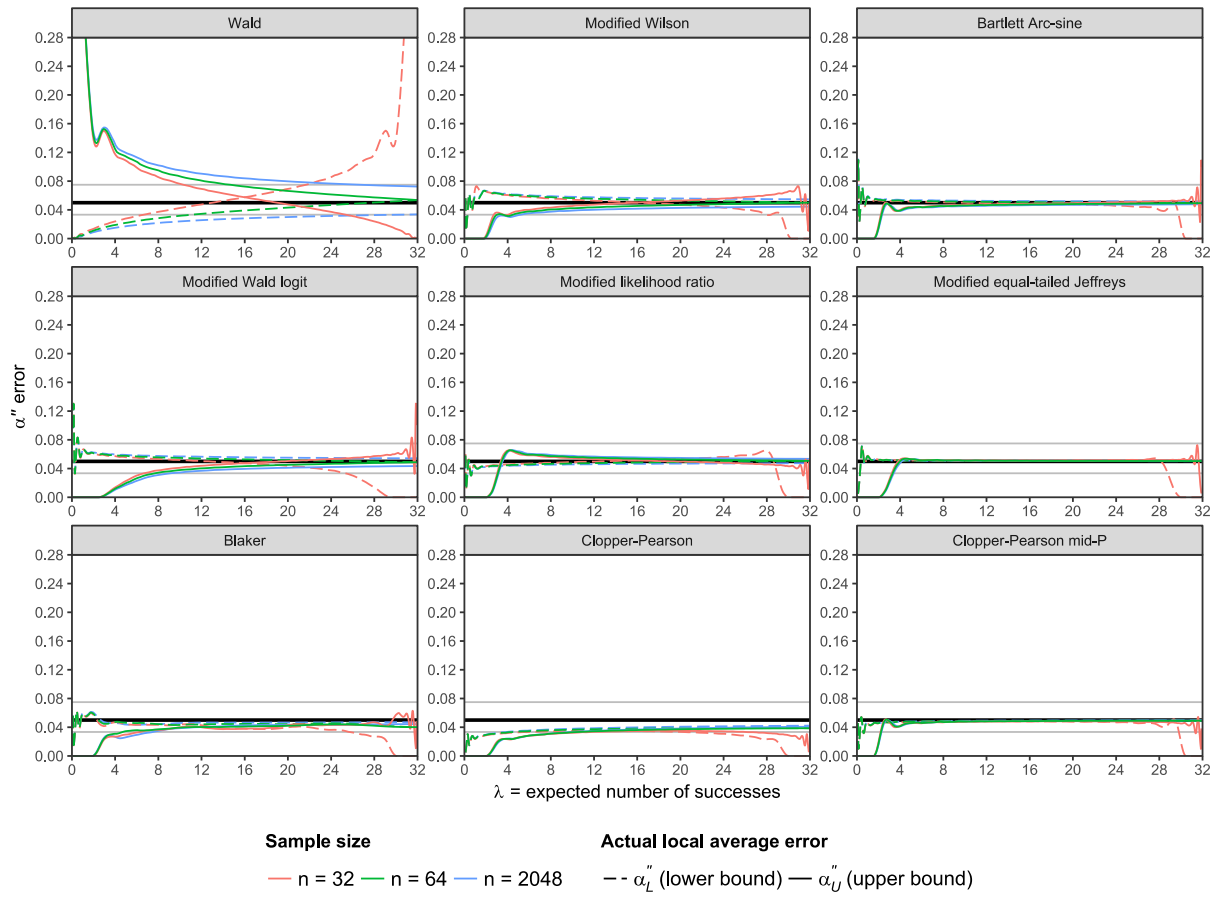
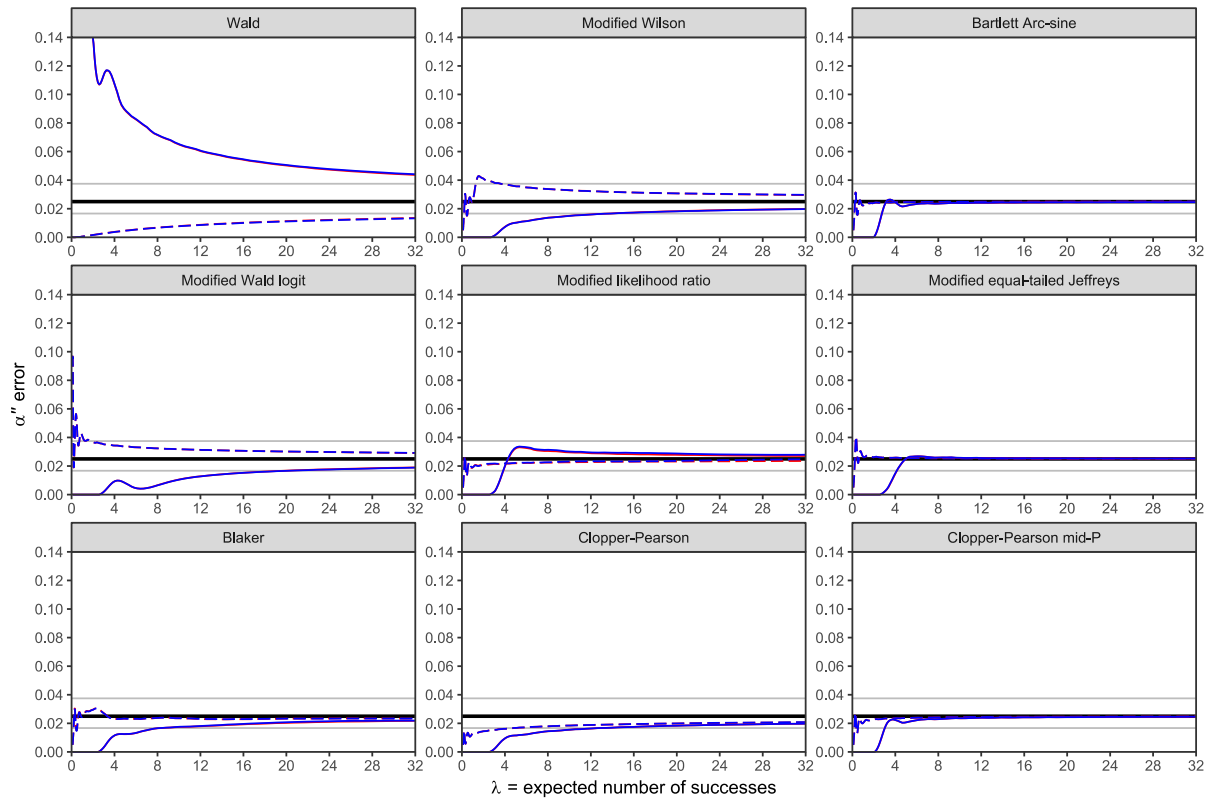


Figure 2 : One-sided local average errors of nine 95% confidence interval estimators according to different sample sizes (red for $n = 32$, green for $n = 64$ and blue for $n = 2048$), with a random true P proportion following a logit-normal distribution with a typical odds ratio of the true proportion between two experiments equal to $OR_S = 1.05$. The abscissa is the expected number of successes np_0 and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (left local average error: dashed lines) or the probability that the upper bound of the confidence interval is lower than the true proportion p (right local average error: solid line). This figure is the same as Figure I in the main article except that $OR_S = 1.05$ rather than 1.20.



$OR_S = 1.20$

Figure 3 : One-sided local average errors of nine 90% confidence interval estimators according to different sample sizes (red for $n = 32$, green for $n = 64$ and blue for $n = 2048$), with a random true P proportion following a logit-normal distribution with a typical odds ratio of the true proportion between two experiments equal to $OR_S = 1.20$. The abscissa is the expected number of successes np_0 and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (left local average error: dashed lines) or the probability that the upper bound of the confidence interval is lower than the true proportion p (right local average error: solid line). This figure is the same as Figure I in the main article except that the confidence level is 0.90 rather than 0.95 and the ordinate scale is doubled making relative biases easy to compare between figures.



Sample size

— $n = 2048$ — $n \rightarrow +\infty$

Actual local average error

--- α_L'' (lower bound) — α_U'' (upper bound)

$OR_S = 1.20$

Figure 4 : One-sided local average errors of nine 95% confidence interval estimators according for a finite sample size $n = 2048$ (red) and an infinite sample size $n \rightarrow +\infty$ (blue) for which the binomial distribution becomes a Poisson distribution. The true P proportion follows a logit-normal distribution with a typical odds ratio of the true proportion between two experiments equal to $OR_S = 1.20$. The abscissa is the expected number of successes np_0 and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (left local average error: dashed lines) or the probability that the upper bound of the confidence interval is lower than the true proportion p (right local average error: solid line).

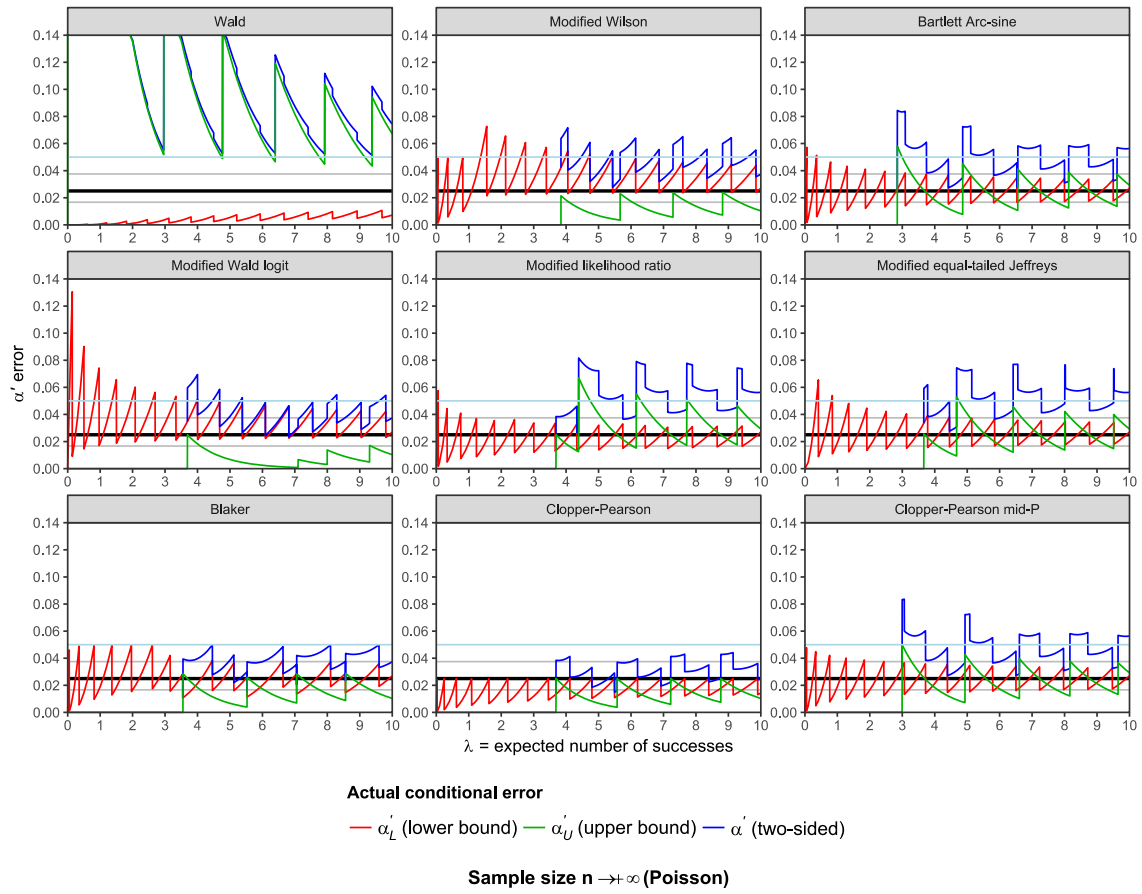


Figure 5 : One-sided and two-sided conditional errors of nine 95% confidence interval estimators for an infinite sample size (Poisson distribution) and a constant theoretical p proportion. The abscissa is the expected number of successes np and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (left conditional error: red), the probability that the upper bound of the confidence interval is lower than the true proportion p (right conditional error: green) or the probability that the confidence interval does not contain the true proportion p (two-sided error: blue). This figure is the same as Figure II of the main article except that $n \rightarrow +\infty$ rather than $n = 2048$.

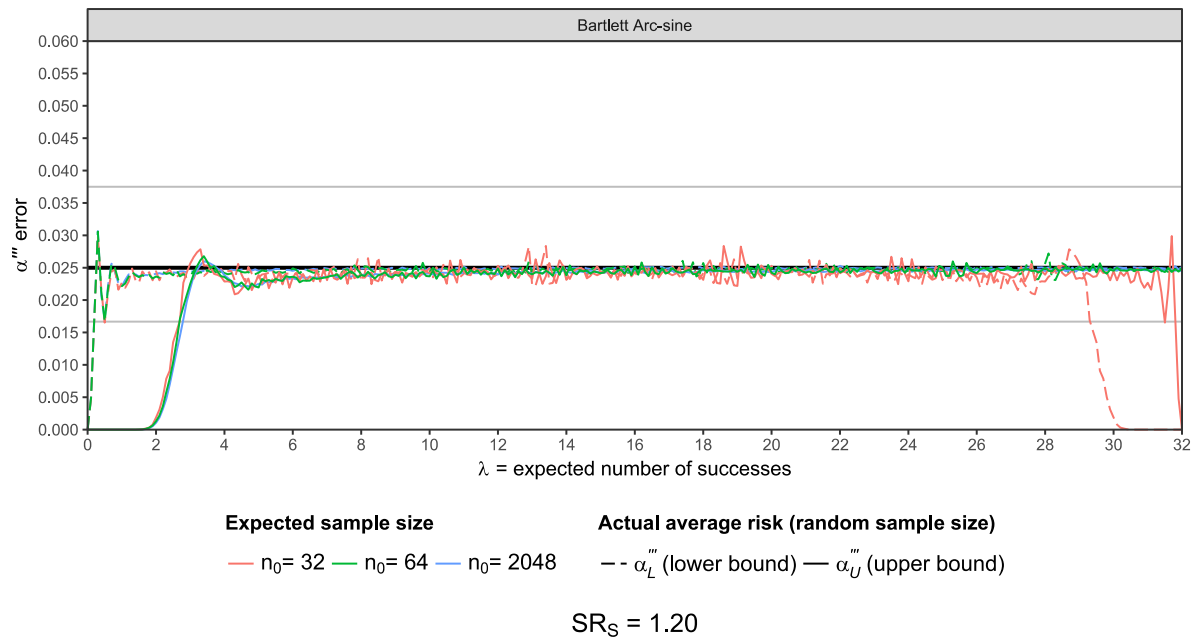


Figure 6 : One-sided local average errors of the 95% Bartlett Arc-Sine interval estimator assuming a constant theoretical proportion p and a random sample size N following a log-normal distribution with expectancy n_0 and a geometric standard deviation $SR_S = 1.20$, rounded to the nearest whole number. The abscissa is the expected number of successes $n_0 p$ and the ordinate is the probability that the lower bound of the confidence interval is greater than the true proportion p (random sample left one-sided error: dashed lines) or the probability that the upper bound of the confidence interval is lower than the true proportion p (random sample right one-sided error: solid line).

5 Figures of additional intervals

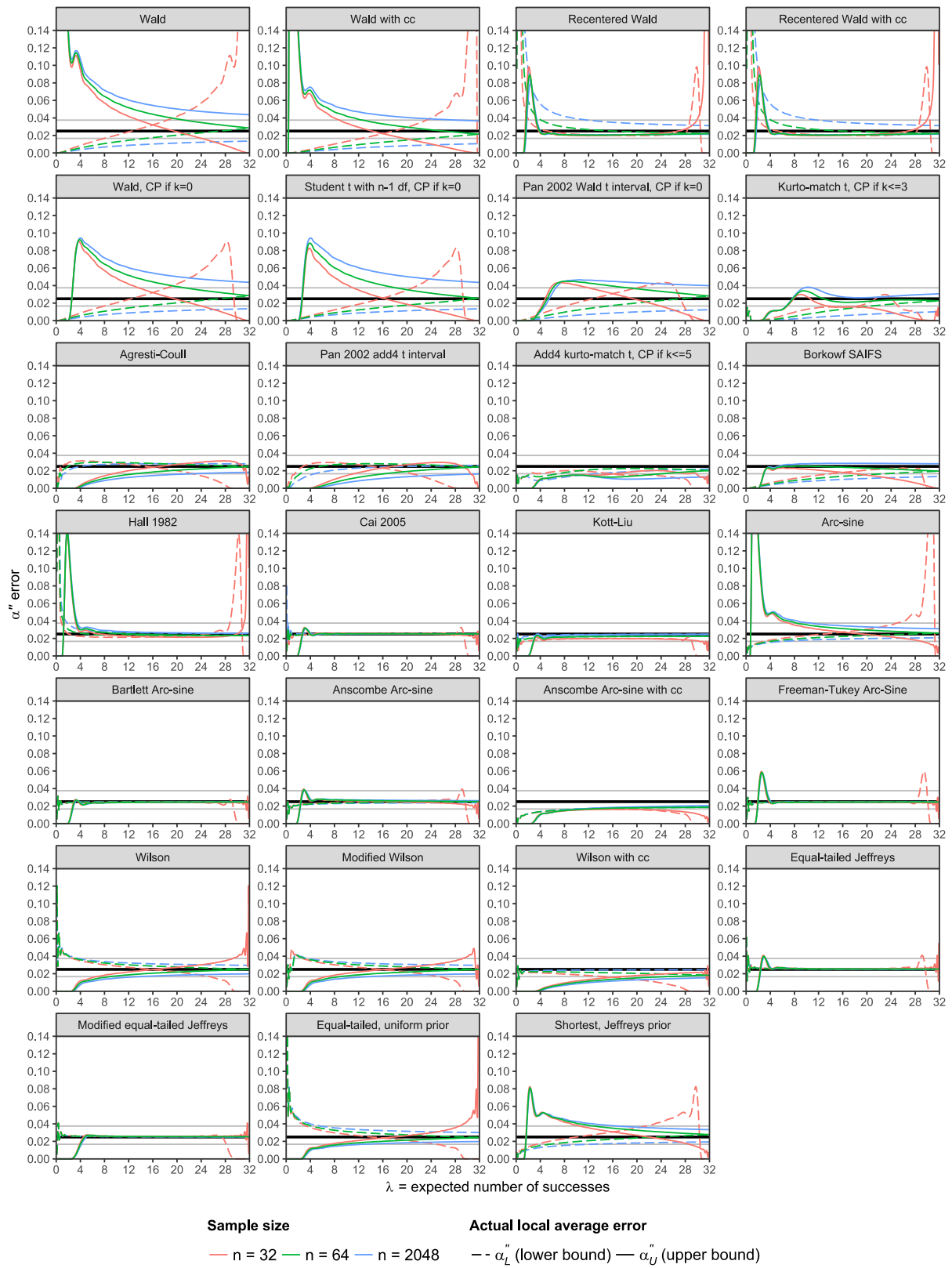
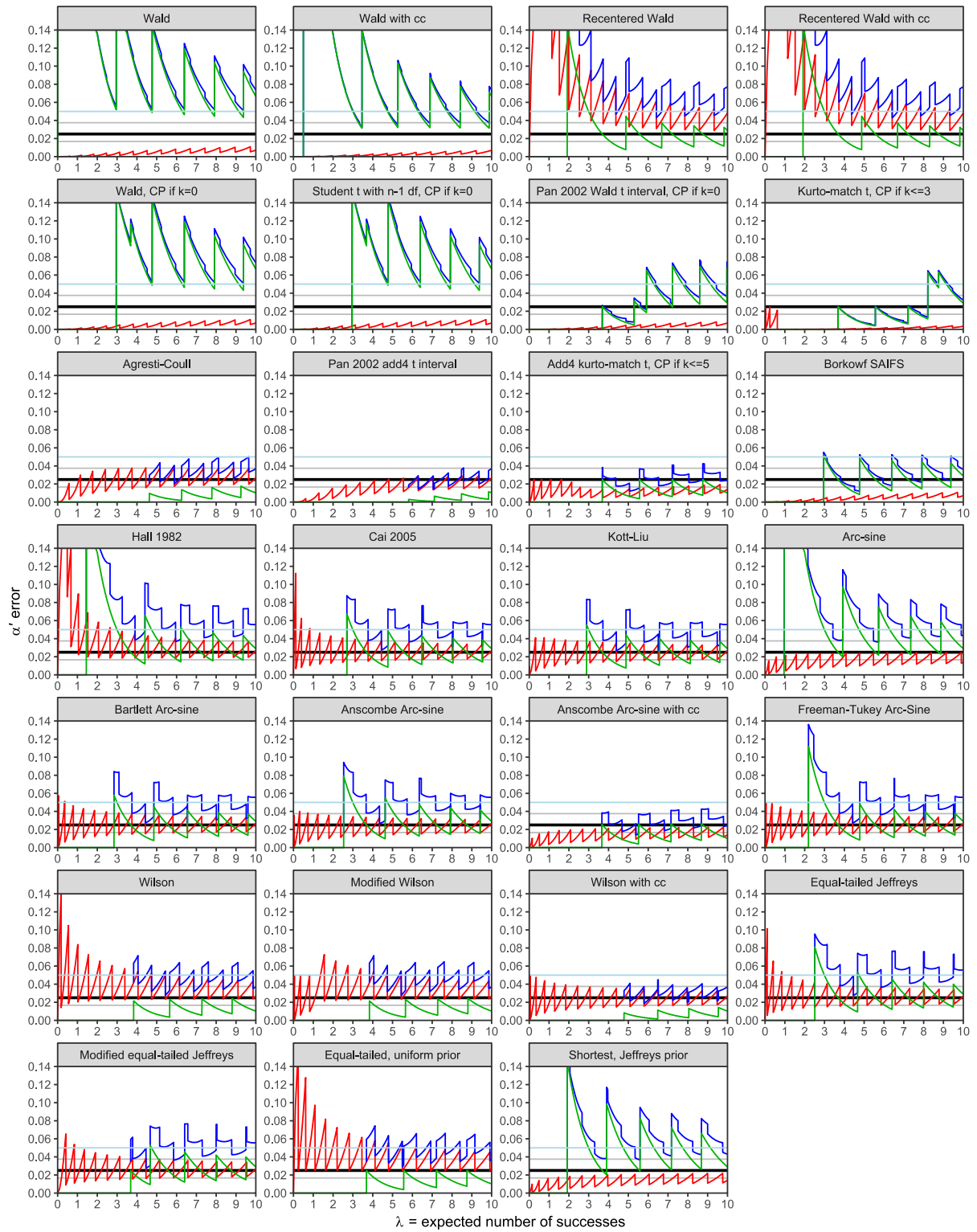


Figure 7 : Local-average α''_L and α''_U errors for 27 two-sided 95% confidence interval estimators, according to the expected number of successes and sample size. The value k equals $\min(x, n - x)$. CP means “Clopper-Pearson”.



Figure 8 : Local-average α_L'' and α_U'' errors for 28 two-sided 95% confidence interval estimators, according to the expected number of successes and sample size. The value k equals $\min(x, n - x)$. CP means “Clopper-Pearson”.



Actual conditional error

— α'_L (lower bound) — α'_U (upper bound) — α' (two-sided)

Sample size $n = 2048$

Figure 9 : Conditional α'_L and α'_U errors for 27 two-sided 95% confidence interval estimators, according to the expected number of successes for a sample size $n=2048$. The value k equals $\min(x, n - x)$. CP means “Clopper-Pearson”.

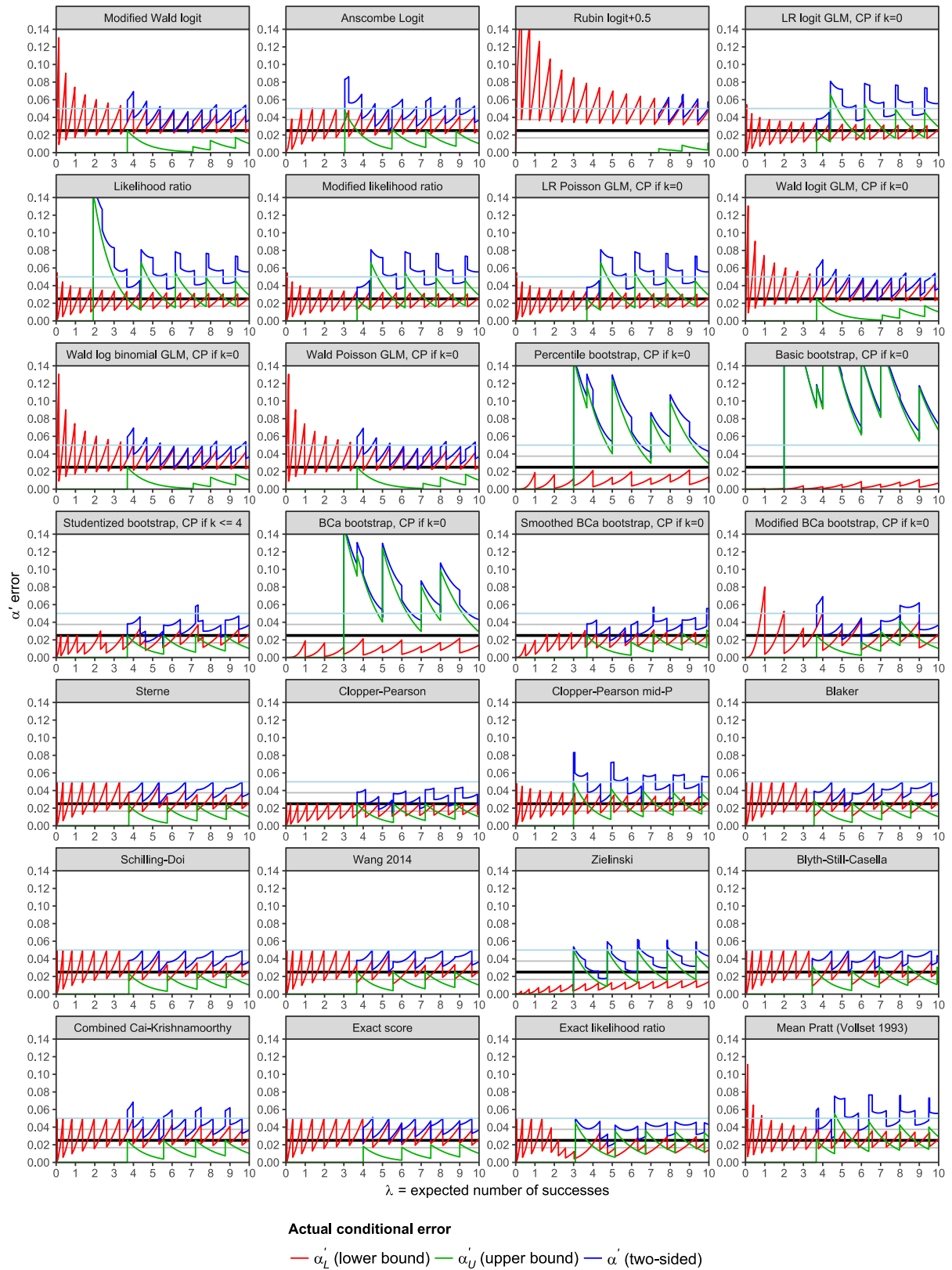


Figure 10 : Conditional α'_l and α'_u errors for 28 two-sided 95% confidence interval estimators, according to the expected number of successes for a sample size n=2048. The value k equals $\min(x, n - x)$. CP means “Clopper-Pearson”.

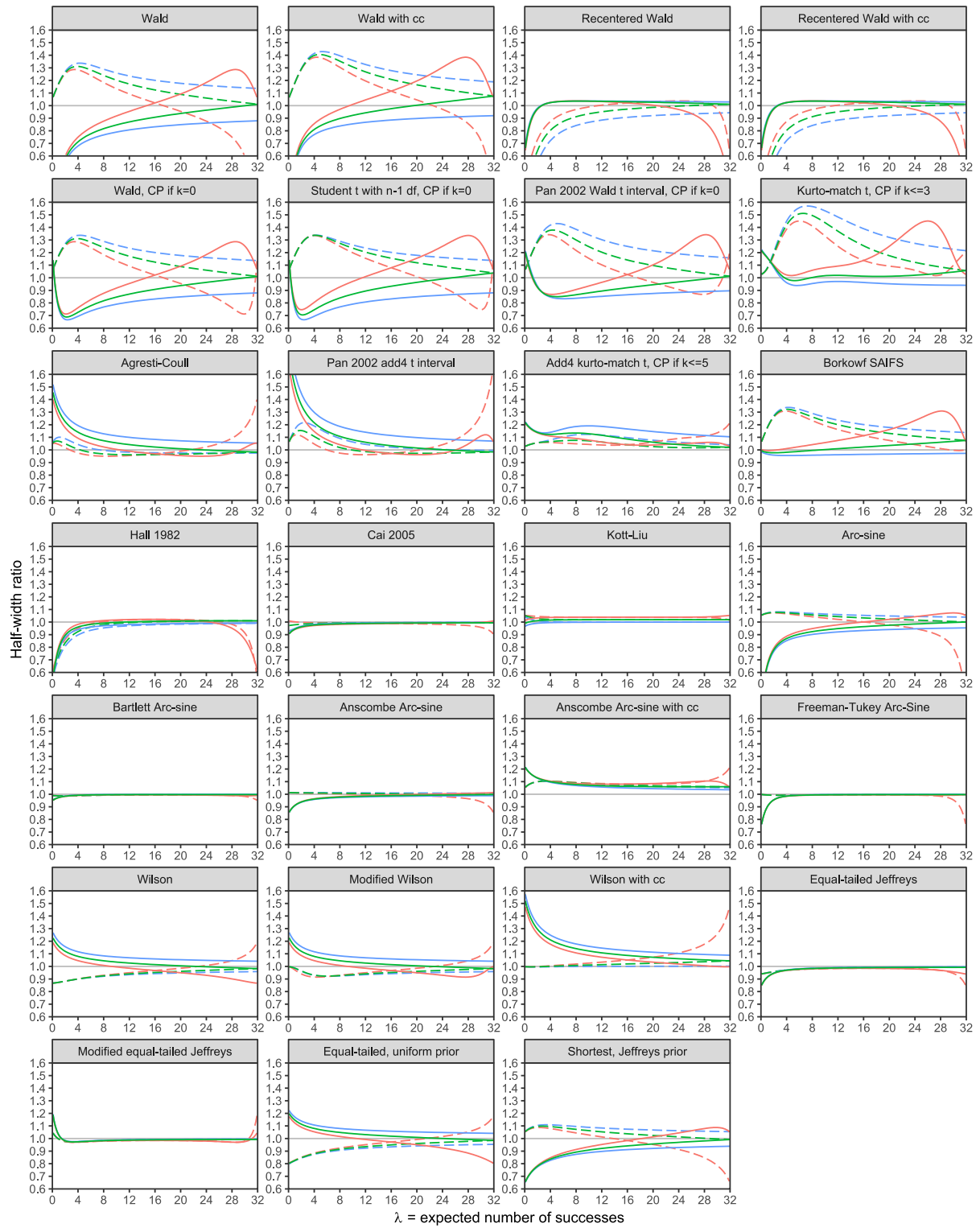


Figure 11 : Local average half-lengths ratio v_L'' and v_U'' of twenty-seven 95% confidence intervals relative to the Clopper-Pearson mid-P interval, according to the expected number of successes $\lambda = np$ and the sample size n .

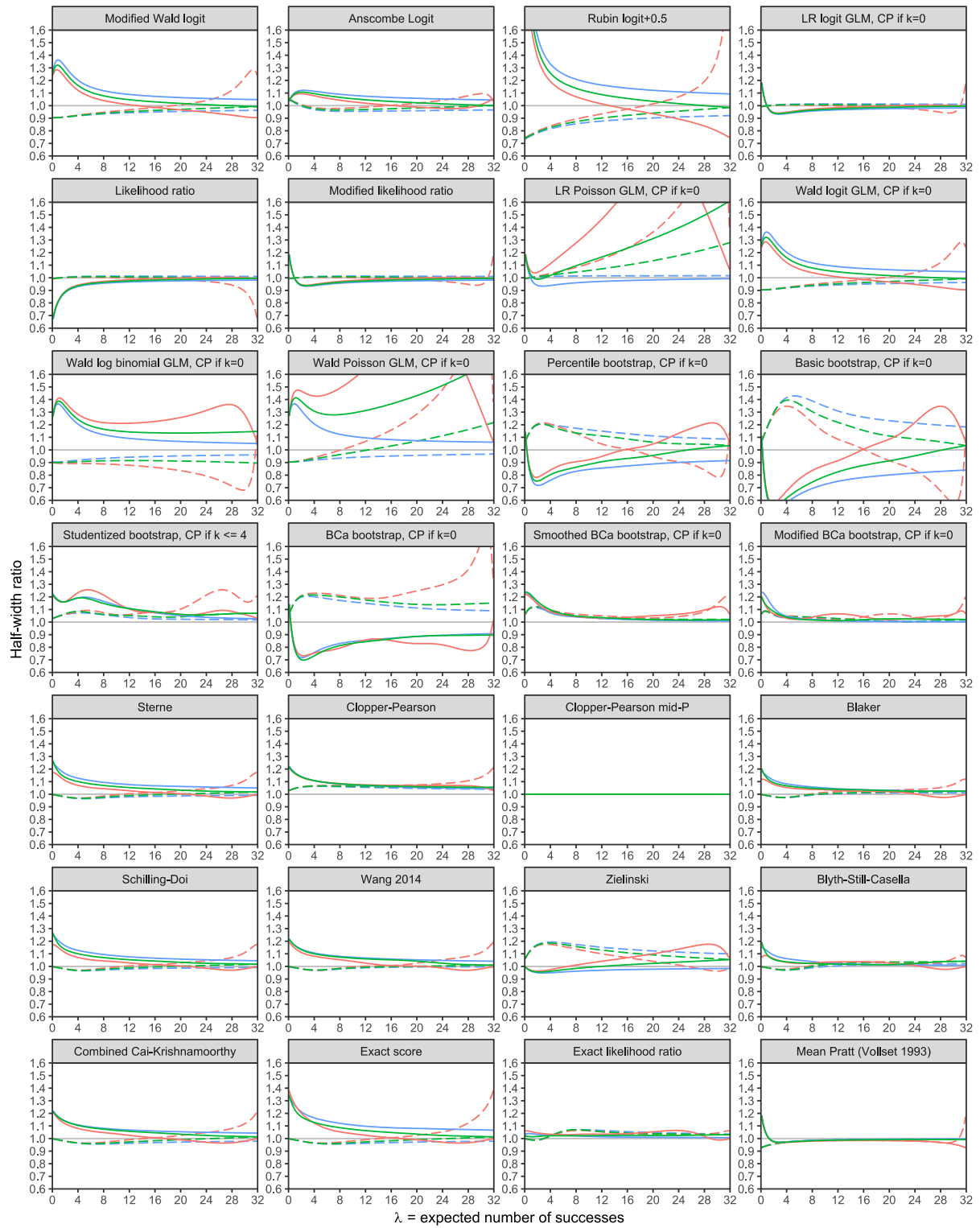


Figure 12 : Local average half-lengths ratio v_L'' and v_U'' of twenty-eight 95% confidence intervals relative to the Clopper-Pearson mid-P interval, according to the expected number of successes $\lambda = np$ and the sample size n .

6 Comparison of strictly conservative exact intervals

Unequal tailed strictly conservative exact intervals tend to be as short as possible. The most recent, Wang ²⁶ and Schilling-Doi ²⁵ requires heavy computation but are designed to be shorter on average. They have been compared to each other on samples of 174 observations. The sample size 174 has been selected because it's large enough to show a difference between Schilling-Doi and Sterne intervals, but otherwise, is not special. Wang and Schilling-Doi intervals optimize the interval length arithmetic mean, but the geometric mean may be more relevant. Squeezing $[0 ; 0.036]$ to $[0 ; 0.034]$ (5.6% relative length reduction) is more interesting than squeezing $[0.398 ; 0.602]$ to $[0.398 ; 0.600]$ (0.98% relative length reduction) while the absolute reduction is equal. The ratio of geometric means is the geometric mean of ratios.

The arithmetic and geometric mean lengths of intervals are shown in Table 15. The Clopper-Pearson interval length geometric mean is 2.6% higher than that of the shortest interval (Blaker). All unequal tailed interval lengths are very close to each other. The Schilling-Doi interval is almost equal to Sterne's interval for all values (Figure 13-A). There are a few spikes where they differ, otherwise, the intervals are equal. The difference is negligible for practical use. Blaker's interval is slightly longer than Schilling-Doi in arithmetic mean and shorter in geometric mean (Table 15), thanks to shorter lengths for proportions close to 0 and 1 (Figure 13-B). Wang's interval has arithmetic and geometric mean lengths almost equal to those of the Schilling-Doi interval (Table 15), but it's shorter for proportions close to 0.50, and longer for most proportions between 0.20 and 0.40 (Figure 13-C). This can be explained by Wang's algorithm, sequentially squeezing intervals from $x = n/2$ down to $x = 0$. Intervals that are squeezed first have more space to be.

The analysis of half lengths show that for small proportions, the Schilling-Doi interval and Wang's interval have upper bounds close to Clopper-Pearson's interval (Figure 14, panels A and B). These intervals have greater lower bounds for small proportions. For instance, the Clopper-Pearson 95% confidence interval for 6 successes out of 174 draws is $[0.0128 ; 0.0735]$ while the Schilling-Doi interval is $[0.0151 ; 0.0738]$. Both boundaries are higher in the Schilling-Doi interval, increasing α'_l and decreasing α'_u . Blaker's intervals are always contained in Clopper-Pearson intervals ²³. Consequently, Blaker's intervals half lengths are smaller or equal for all proportions. For 6 successes out of 174 draws, Blaker's interval is $[0.0151 ; 0.0725]$.

Table 15 : Mean interval lengths of strictly conservative exact 95% confidence intervals for sample size $n = 174$.

Interval name	Arithmetic mean of length	Geometric mean of length
Clopper-Pearson	0.12137	0.11470
Blaker	0.11855	0.11182
Wang	0.11848	0.11196
Sterne	0.11852	0.11216
Schilling-Doi	0.11846	0.11199
Exact score	0.11804	0.11192
Exact likelihood ratio (LR)	0.11882	0.11185

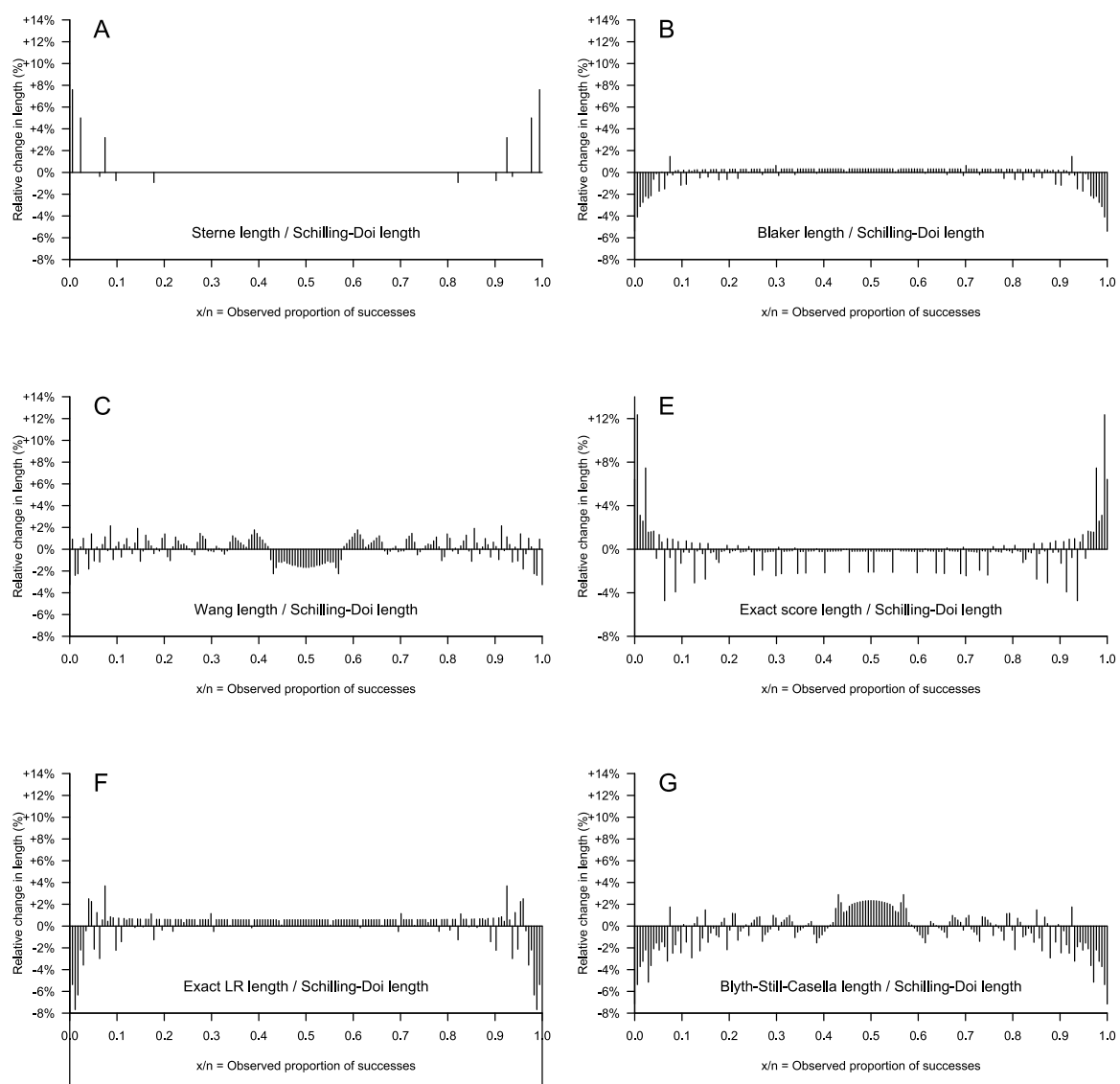


Figure 13 : Relative lengths of 95% unequal-tailed confidence intervals for sample size $n = 174$, for any integer number of successes from 0 to 174. It is not a random experiment; it is a representation of the 175 confidence intervals for $x = 0, \dots, 174$ and $n = 174$.

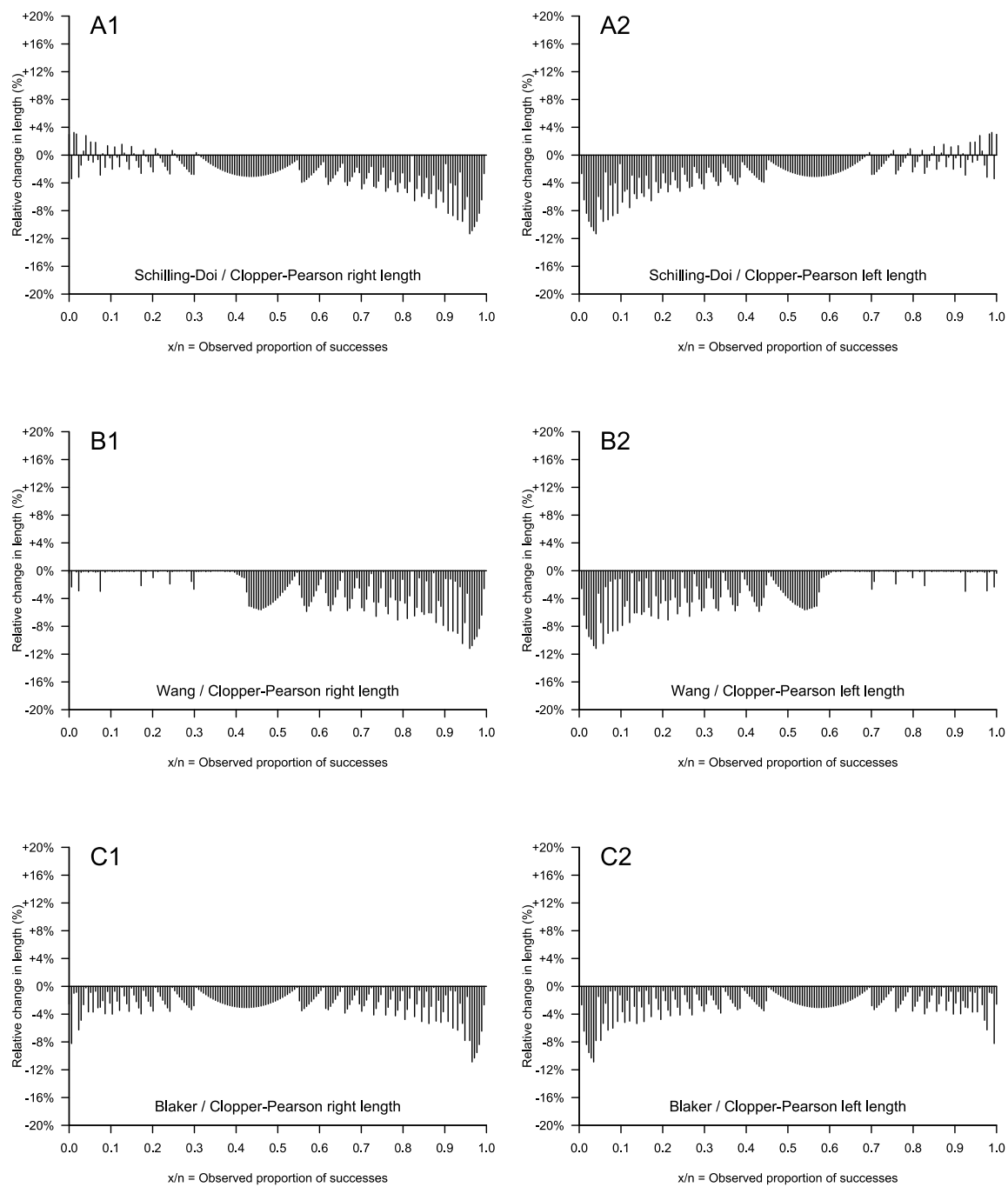


Figure 14 : Relative half-lengths of 95% unequal-tailed confidence intervals (Schilling-Doi Wang, Blaker) compared to the Clopper-Pearson interval, for sample size $n = 174$, for any integer number of successes from 0 to 174. It is not a random experiment; it is a representation of the 175 confidence intervals for $x = 0, \dots, 174$ and $n = 174$.

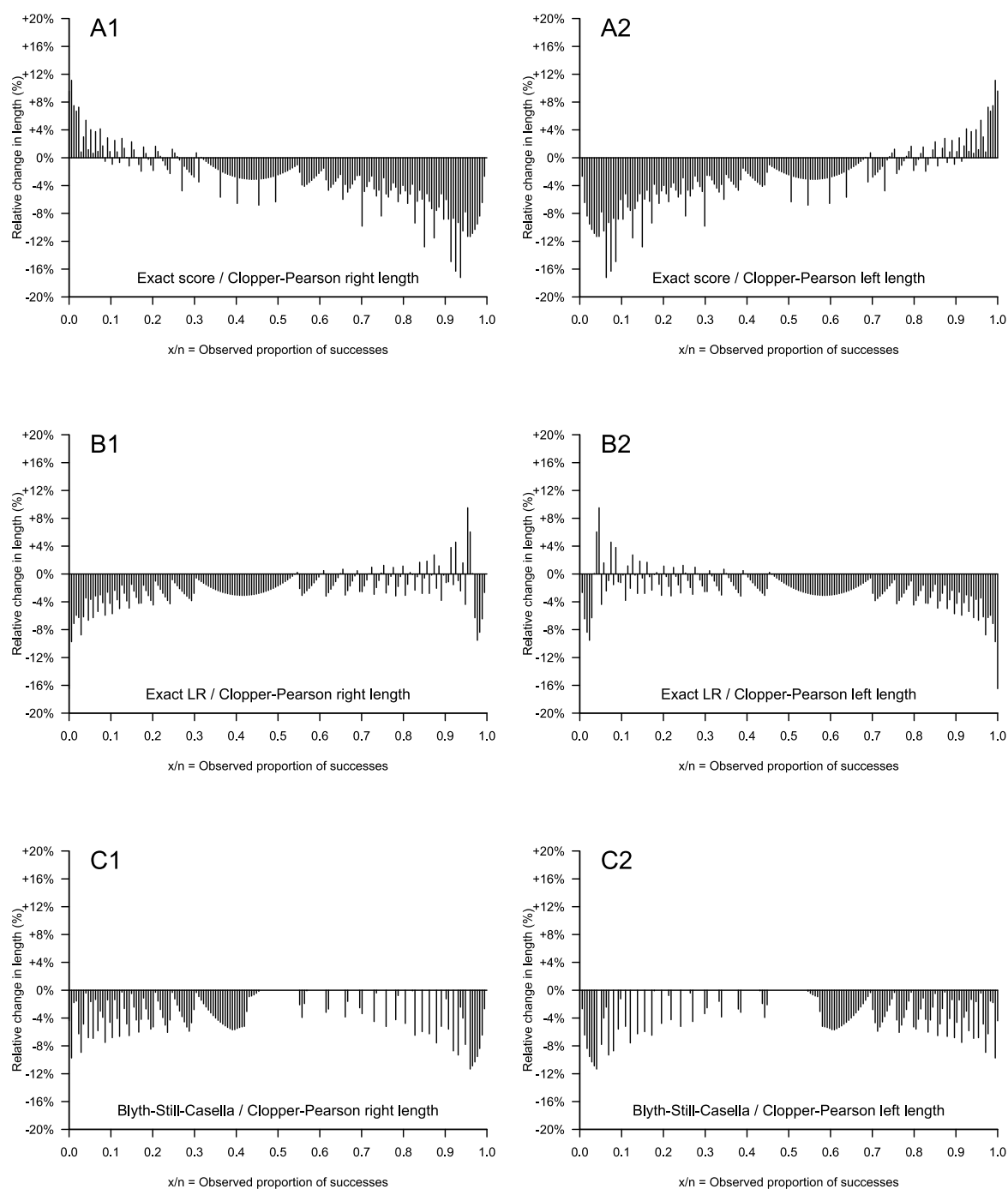


Figure 15 : Relative half-lengths of 95% unequal-tailed confidence intervals (Exact score, Exact LR and Blyth-Still-Casella) compared to the Clopper-Pearson interval, for sample size $n = 174$, for any integer number of successes from 0 to 174. It is not a random experiment; it is a representation of the 175 confidence intervals for $x = 0, \dots, 174$ and $n = 174$.

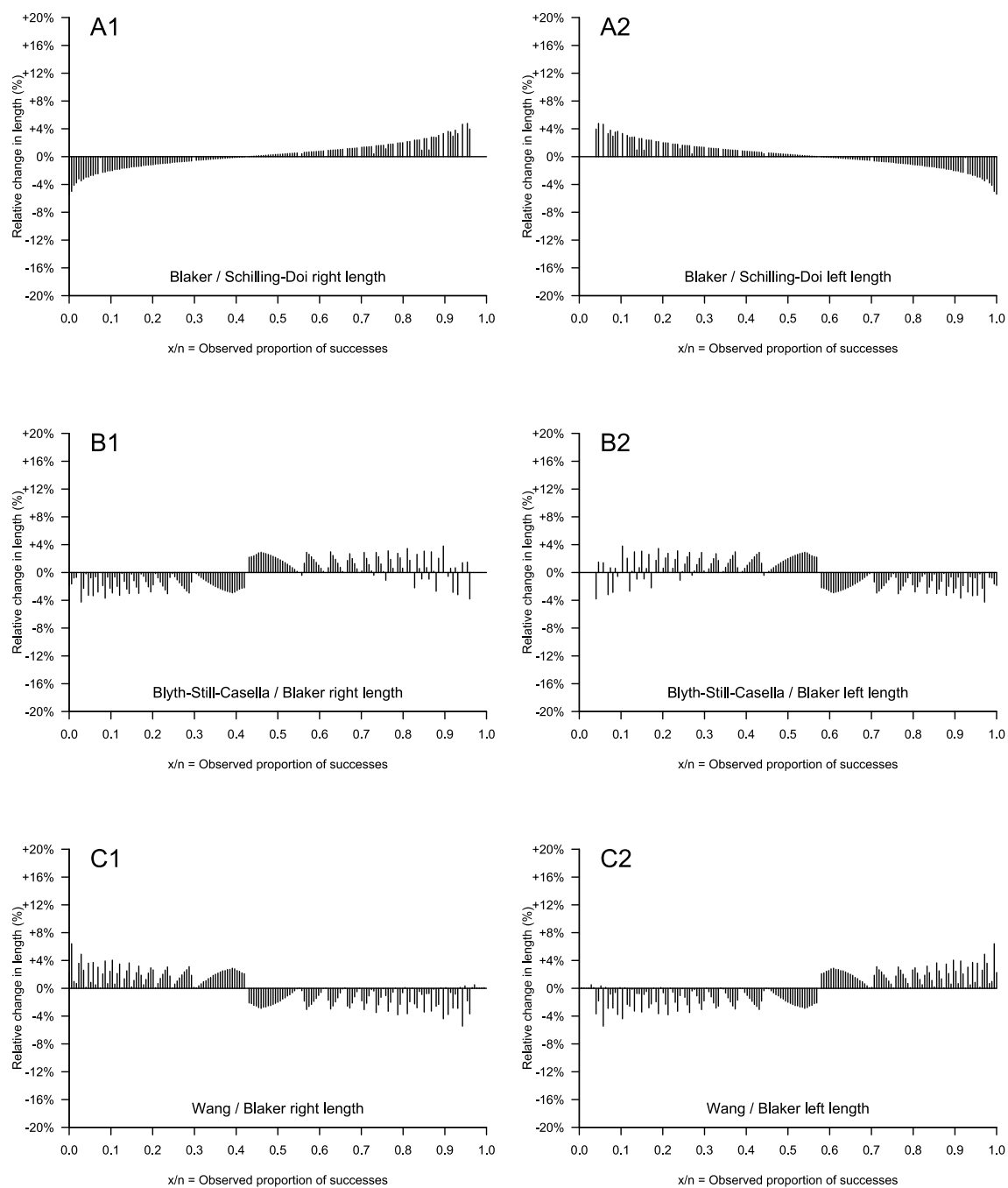


Figure 16 : Relative half-lengths of 95% unequal-tailed confidence intervals (Blaker, Schilling-Doi, Blyth-Still-Casella, Wang) compared to each other, for sample size $n = 174$, for any integer number of successes from 0 to 174. It is not a random experiment; it is a representation of the 175 confidence intervals for $x = 0, \dots, 174$ and $n = 174$.

7 References

1. Pan W. Approximate confidence intervals for one proportion and difference of two proportions. *Computational statistics & data analysis* 2002; **40**:143–157.
2. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statistical science* 2001; **16**:101–117.
3. Garwood F. Fiducial limits for the Poisson distribution. *Biometrika* 1936; **28**:437-442.
4. Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician* 1998; **52**:119–126.
5. Agresti A, Caffo B. Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. *The American Statistician* 2000; **54**:280-288.
6. Brown LD, Cai TT, DasGupta A. Confidence Intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics* 2002; **30**:160-201.
7. Borkowf CB. Constructing binomial confidence intervals with near nominal coverage by adding a single imaginary failure or success. *Statistics in Medicine* 2006; **25**:3679-3695.
8. Vollset SE. Confidence intervals for a binomial proportion. *Statistics in Medicine* 1993; **12**:809-824.
9. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine* 1998; **17**:857–872.
10. Pires AM, Amado C. Interval estimators for a binomial proportion: Comparison of twenty methods. *REVSTAT–Statistical Journal* 2008; **6**:165–197.
11. Student. The probable error of a mean. *Biometrika* 1908; **6**:1-25.
12. Blyth CR. Approximate Binomial Confidence Limits. *Journal of the American Statistical Association* 1986; **81**:843-855.
13. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* 1927; **22**:209-212.
14. Bartlett MS. The Square Root Transformation in Analysis of Variance. *Supplement to the Journal of the Royal Statistical Society* 1936; **3**:68-78.
15. Anscombe FJ. The transformation of poisson, binomial and negative-binomial data. *Biometrika* 1948; **35**:246-254.
16. Freeman MF, Tukey JW. Transformations Related to the Angular and the Square Root. *The Annals of Mathematical Statistics* 1950; **21**:607-611.
17. Yu G. Variance stabilizing transformations of Poisson, binomial and negative binomial distributions. *Statistics & Probability Letters* 2009; **79**:1621-1629.
18. Anscombe FJ. On Estimating Binomial Response Relations. *Biometrika* 1956; **43**:461-464.

19. Matuszewski A, Sotres D. A basic statistical problem: Confidence interval for the Bernoulli parameter. *Computational Statistics & Data Analysis* 1985; **3**:103-114.
20. Rubin DB, Schenker N. Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological methodology* 1987; **17**:131–144.
21. Efron B. Better bootstrap confidence intervals. *Journal of the American statistical Association* 1987; **82**:171–185.
22. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 2000; **19**:1141-1164.
23. Blaker H. Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics* 2000; **28**:783–798.
24. Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman & Hall.; 1974.
25. Schilling MF, Doi JA. A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure. *The American Statistician* 2014; **68**:133-145.
26. Wang W. AN ITERATIVE CONSTRUCTION OF CONFIDENCE INTERVALS FOR A PROPORTION. *Statistica Sinica* 2014; **24**:1389-1410.
27. Blyth CR, Still HA. Binomial Confidence Intervals. *Journal of the American Statistical Association* 1983; **78**:108-116.
28. Casella G. Refining binomial confidence intervals. *Canadian Journal of Statistics* 1986; **14**:113–129.
29. Winstein K. Efficient routines for biostatistics. URL <https://github.com/keithw/biostat> Published September 30, 2016 (accessed July 26, 2017).
30. Zieliński W. The Shortest Clopper–Pearson Confidence Interval for Binomial Probability. *Communications in Statistics - Simulation and Computation* 2009; **39**:188-193.
31. Zieliński W. The Shortest Clopper-Pearson Randomized Confidence Interval. *REVSTAT–Statistical Journal* 2017; **15**:141–153.
32. Sakakibara I, Haramo E, Muto A, Miyajima I, Kawasaki Y. Comparison of five exact confidence intervals for the binomial proportion. *American Journal of Biostatistics* 2014; **4**:11.
33. Thulin M, Zwanzig S. Exact confidence intervals and hypothesis tests for parameters of discrete distributions. *Bernoulli* 2017; **23**:479-502.
34. Cai Y, Krishnamoorthy K. A simple improved inferential method for some discrete distributions. *Computational statistics & data analysis* 2005; **48**:605–621.
35. Vos PW, Hudson S. PROBLEMS WITH BINOMIAL TWO-SIDED TESTS AND THE ASSOCIATED CONFIDENCE INTERVALS. *Australian & New Zealand Journal of Statistics* 2008; **50**:81–89.
36. Sterne TE. Some remarks on confidence or fiducial limits. *Biometrika* 1954; **41**:275-278.
37. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; **26**:404-413.

38. Lancaster HO. The combination of probabilities arising from data in discrete distributions. *Biometrika* 1949; **36**:370-382.
39. Berry G, Armitage P. Mid-P confidence intervals: a brief review. *The Statistician* 1995; **44**:417–423.
40. Krishnamoorthy K, Thomson J, Cai Y. An exact method of testing equality of several binomial proportions to a specified standard. *Computational statistics & data analysis* 2004; **45**:697–707.
41. Liu YK, Kott PS. Evaluating alternative one-sided coverage intervals for a proportion. *Journal of Official Statistics* 2009; **25**:569.
42. Hall P. Improving the Normal Approximation when Constructing One-Sided Confidence Intervals for Binomial or Poisson Parameters. *Biometrika* 1982; **69**:647-652.
43. Cai TT. One-sided confidence intervals in discrete distributions. *Journal of Statistical planning and inference* 2005; **131**:63–88.
44. Venzon DJ, Moolgavkar SH. A method for computing profile-likelihood-based confidence intervals. *Applied statistics* 1988:87–94.