Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

1310

Wilson confidence intervals for the two-sample log-odds-ratio in stratified 2 x 2 contingency tables

B.M. Brown, Thomas Suesse and Von Bing Yap

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Wilson confidence intervals for the two-sample log-odds-ratio in stratified 2×2 contingency tables

Short title: Wilson confidence intervals for 2×2 contingency tables

B. M. Brown[1], Thomas Suesse[2] and Von Bing Yap[3]

[1] *Department of Statistics, University of New South Wales,*

Sydney, NSW 2052, Australia. email: bruce.brown@unsw.edu.au

[2] *School of Mathematics and Applied Statistics, University of Wollongong, NSW 2500,*

Australia. email: tsuesse@uow.edu.au

[3] *Department of Statistics and Applied Probability, National University of Singapore,*

6 Science Drive 2, Singapore 117546. email: stayvb@nus.edu.sg

Correspondence to

Thomas Suesse

School of Mathematics and Applied Statistics,

University of Wollongong, NSW 2500, Australia

email: tsuesse@uow.edu.au

tel: (61) 2 4221 4173; fax: (61) 2 4221 4845

Abstract

Large-sample Wilson-type confidence intervals (**CIs**) are derived for a parameter of interest in many clinical trials situations: the log-odds-ratio, in a two sample experiment comparing binomial success proportions, say between cases and controls. The methods cover several scenarios: (i) results embedded in a single 2×2 contingency table, (ii) a series of K 2×2 tables with common parameter, or (iii) K tables, where the parameter may change across tables under the influence of a covariate. The calculations of the Wilson CI require only simple numerical assistance, and for example are easily carried out using Excel. The main competitor, the exact CI, has two disadvantages: It requires burdensome search algorithms for the multi-table case and results in strong over-coverage associated with long confidence intervals. All the application cases are illustrated through a well-known example. A simulation study then investigates how the Wilson CI performs among several competing methods. The Wilson interval is shortest, except for very large odds ratios, while maintaining coverage similar to Wald-type intervals. An alternative to the Wald CI is the Agresti-Coull CI, calculated from Wilson and Wald **CIs**, which has same length as **the** Wald CI but improved coverage.

Some key words: binomial proportions, case-control studies, exponential families, hyper-geometric distributions, sufficient statistics.

1 Introduction

The elementary problem of comparing two binomial success probabilities θ_1 and θ_2 is an old one in Statistics, with many applications in medical and related fields. It can be parametrized in several ways, for example through the difference $\theta_1 - \theta_2$, the ratio θ_1/θ_2, the complementary ratio $(1-\theta_1)/(1-\theta_2)$, or the log-odds-ratio,

$$\alpha = \log\left\{\frac{\theta_1(1-\theta_2)}{(1-\theta_1)\theta_2}\right\}.$$

Each of the above parametrizations has a corresponding practical setting for which it is appropriate. A good discussion is in (Agresti 2002, p. 43 ff). Among these different methods, the use of log-odds-ratio has some further advantages in terms of interpretations within prospective studies; see Agresti (2002, p. 45 ff). The present paper considers the problem of developing large sample confidence intervals (CIs) for α, for several different experimental settings, in a way that parallels the Wilson CIs for a single binomial probability; see Wilson (1927) or Brown et al. (2001).

Wilson-type intervals use an approximately normal pivotal statistic, standardized by the correct standard deviation expression rather than a standard error estimate, and as such, are generally superior to Wald intervals. Despite the elementary nature of the problem under consideration, Wilson CIs appear not to have been derived previously, current popular intervals **being** the Wald CI and the exact CI. The former is based on a asymptotic result, as **is** the Wilson CI, and is usually provided as standard output by statistical packages, e.g. the routine `clogit` of the `R` (R-Development-Core-Team 2006) package

`survival` (Therneau and Lumley 2010) provides the Wald CI. The latter, the exact CI, strictly maintains the nominal confidence level, but at two **costs, namely** strong over-coverage **plus a** huge computational burden in order to compute coefficients necessary to determine the conditional distribution for the multi-table case (Mehta et al. 1985). This explains **why the exact CI is** not widely implemented in statistical packages, for example it is not implemented in `R`. The phrase "exact" refers to the exact (and not approximate) underlying discrete distribution, but the discreteness makes it impossible to attain **exact nominal confidence level, this explaining over-coverage**, see Agresti (2002, p.18-19) for a discussion for the binomial case. The **Agresti-Coull** CI, the symmetric interval around the mid point of the Wilson CI with length equal to Wald CI, is computational attractive, since it only requires the Wald and the Wilson **CIs, both requiring relatively simple computations** for the scenarios considered here.

The necessary theory for the Wilson CI is discussed in Section 2. Conditioning is used to remove a nuisance parameter $\beta = \log\{\theta_1\theta_2(1-\theta_1)^{-1}(1-\theta_2)^{-1}\}$. This conditioning creates a 2×2 table with fixed marginals, and hence one degree of freedom. The single random observable has a large sample approximate distribution, assuming all marginal totals are large, of known form. The result is re-stated as Theorem 1 in a form applicable to the CI problems considered here, with a simple proof. From it, a CI method **is formulated, in Section 3**, for a single 2×2 table, with very regular numerical and computational properties. The calculations are easy to implement, for example, with Excel.

The method is also applied, in Section 4, to find a CI for the assumed common value of α across a series of K 2×2 stratified tables. The numerical calculation properties are not as strong as for the single table case, but are reliable nevertheless.

A more complex situation still is the case of K stratified tables, where the parameter α changes across tables, through the influence of a covariate u. This creates a regression-style situation where a 'slope' parameter δ, measuring the influence of u on α, is of interest. Once again the proposed procedures can be adapted to this case, in Section 5, with reliable numerical convergence.

A well-known example is used in Section 6 to illustrate all cases of the methods. Section 7 investigates the performance of the Wilson CI, the Wald CI, the Agresti-Coull CI and the exact CI by conducting a **simulation study and discusses results along with future research directions.**

2 Theory for a Single Table

For $i = 1, 2$, let θ_i be the ith sample success probability, with sample size n_i and x_i observed successes, and let $\phi_i = \log\{\theta_i/(1-\theta_i)\}$. Then the likelihood is

$$L = \prod_{i=1}^{2} \binom{n_i}{x_i} \theta_i^{x_i} (1-\theta_i)^{n_i - x_i}, \quad \propto \quad \prod_{i=1}^{2} \frac{e^{\phi_i x_i}}{(1+e^{\phi_i})^{n_i}}.$$

Write $\alpha = \phi_1 - \phi_2$ and $\beta = \phi_1 + \phi_2$. Then

$$L \quad \propto \quad \frac{\exp\{x_1(\frac{\alpha+\beta}{2}) + x_2(\frac{\beta-\alpha}{2})\}}{\{1+e^{(\alpha+\beta)/2}\}^{n_1}\{1+e^{(\beta-\alpha)/2}\}^{n_2}},$$

implying that $x_1 + x_2 = s$, the total number of successes, is sufficient for the nuisance

parameter β. Conditioning on s shows that $x_1 - x_2$, or equivalently just $x_1 = x$, is sufficient for α, and that this conditional likelihood, given s, as a function of x is $\propto \exp(\alpha x)$.

Therefore the conditional probability for x, given s, must have the exponential form

$$p_x = P(X_1 = x | X_1 + X_2 = s) = \frac{a_x e^{\alpha x}}{m(\alpha)},$$

where $\{a_x\}$ is the corresponding probability distribution for the null case $\alpha = 0$, which is the hyper-geometric distribution $Hg(n; s, n_1)$, with $N = n_1 + n_2$. Also, m is the moment generating function of this hyper-geometric distribution. With this exponential family form, x is conditionally sufficient for α, and optimal large-sample inference for α will be based upon the corresponding large-sample distribution of x.

The form of this distribution is non-central hyper-geometric, differing from the usual Hg distribution through the factor $e^{\alpha x}$. Under conventional large-sample assumptions n_1, n_2 and s are all assumed to be large, and the usual, well-known normal approximation to the hyper-geometric corresponds to $\alpha = 0$. This result also covers the cases $\alpha = O(N^{-1/2})$, but the present case where $\alpha = O(1)$ is different. The required result is of known form; see Breslow (1981), and **earlier papers by Stevens (1951) and Hannan and Harkness (1963)**. It is re-stated now in a form suitable for extension to confidence interval derivation in the more complex cases of later sections.

Theorem 1. As $N \to \infty$, suppose that $s/N \to p$ and $n_1/N \to q$. Let $y = x/N$, and $z = \sqrt{N}(y - \mu)/\sigma$, where μ is given implicitly for fixed α by

$$\alpha = g(\mu) = \log\left\{\frac{\mu(1 - p - q + \mu)}{(p - \mu)(q - \mu)}\right\}, \tag{1}$$

and σ^2 by

$$\sigma^2\Big\{\frac{1}{\mu}+\frac{1}{p-\mu}+\frac{1}{q-\mu}+\frac{1}{1-p-q+\mu}\Big\}=\sigma^2 g'(\mu)=1. \tag{2}$$

Then $z \to^{\mathcal{D}} N(0,1)$ as $N \to \infty$.

A standard proof for this know result can be found in the **Appendix**.

The approximate mean μ **and variance** σ^2/N **of** y **are determined** implicitly by (1) and (2). Details of iterative confidence interval calculations for a single 2×2 table are outlined in the next Section.

3 Confidence Interval Calculations for a Single Table

The equations (1) and (2), which determine the approximate mean and variance μ and σ^2/N, have the form $\alpha = g(\mu)$ and $\sigma^{-2} = g'(\mu)$, where g' is clearly a convex function, being the sum of four convex functions of μ; see (2). An approximate $100(1-\varepsilon)\%$ CI for μ is found by solving

$$\frac{N(y-\mu)^2}{\sigma^2} = z^2_{1-\varepsilon/2}, \tag{3}$$

where z_ε is the level-ε quantile of $N(0,1)$. The two solutions μ_1, μ_2 are the required CI end-points. A corresponding CI for α comes from (1), ie with end-points $\alpha_i = g(\mu_i)$ for $i = 1, 2$.

Computing the CI for μ through solving (3) is numerically easy, because $\sigma^{-2}(y-\mu)^2 = g'(\mu)(y-\mu)^2$ is also the sum of four convex functions of μ, and hence is convex. To see

this, a typical term is

$$\frac{(y-\mu)^2}{(p-\mu)^2} = p - \mu + 2(y-p) + \frac{(y-p)^2}{p-\mu},$$

convex in μ.

In addition to being convex, the function $g'(\mu)(y-\mu)^2$ asymptotes up to $+\infty$ at the two end points of the permissible range of μ-values, $\max(0, p+q-1) \le \mu \le \min(p,q)$. Thus (3) has only two solutions μ_1, μ_2 for μ, and the CI (μ_1, μ_2) converts easily to the CI $(g(\mu_1), g(\mu_2))$ for α.

To set up a convergent iterative scheme for calculating the roots μ_1, μ_2 of (3), observe that this equation can be re-written as

$$\begin{aligned} 1 + \frac{z^2_{1-\varepsilon/2}}{N} &= 1 + g'(\mu)(y-\mu)^2 \\ &= \frac{y^2}{\mu} + \frac{(y-p)^2}{p-\mu} + \frac{(y-q)^2}{q-\mu} + \frac{(y+1-p-q)^2}{1-p-q+\mu} \\ &= f(\mu) \end{aligned}$$

say, so that

$$f'(\mu) = -\frac{y^2}{\mu^2} + \frac{(y-p)^2}{(p-\mu)^2} + \frac{(y-q)^2}{(q-\mu)^2} - \frac{(y+1-p-q)^2}{(1-p-q+\mu)^2}. \tag{4}$$

We know that $u < \mu_1 < y < \mu_2 < v$, where $u = \max(0, p+q-1)$ and $v = \min(p,q)$. Begin the iterations with a trial value $\mu_{(0)}$ with $u < \mu_{(0)} < \mu_1$ for the lower root μ_1, or $\mu_2 < \mu_{(0)} < v$ for the upper root μ_2. Then use the Newton-Raphson algorithm:

$$\mu_{(j)} = \mu_{(j-1)} - \frac{f(\mu_{(j-1)}) - 1 - z^2/n}{f'(\mu_{(j-1)})}.$$

The convexity of f guarantees that $\{\mu_{(j)}\}$ converges rapidly and monotonically to μ_1 or μ_2, as j increases.

The routine is easy to implement, for example in Excel, and is illustrated in Section 6. It is based upon calculating a CI on the μ scale, where there are strong convexity properties to aid the calculation, then converting to the α scale through use of $\alpha = g(\mu)$. Writing the inverse function as $\mu = h(\alpha)$, the equivalent formulation on the α scale satisfies $N\{y - h(\alpha)\}^2/h'(\alpha) = z^2_{1-\varepsilon/2}$, instead of (3). The function h does not necessarily have the same strong convexity properties as g, but is monotone on either side of its single zero, so CI endpoints calculated on the α scale are monotone in z, and unique. While use of the μ scale is numerically easier, when applications to multiple 2×2 contingency tables are considered in the Sections to follow, it is necessary to revert to calculations on the α scale.

4 Stratified Tables with Common Log-Odds-Ratio

Consider a series of K stratified 2×2 tables, all with a common value of log-odds-difference α. The ith table has nuisance parameter β_i, removed by conditioning on the marginal total of successes, s_i. Thus each table has fixed marginals and a single informative observable. The result of Theorem 1 applies to each table, but to bring all such results under one umbrella, a modified set-up is needed. Using subscripts i to refer to the ith table, $X = \sum_1^K x_i$ is sufficient for the common α value of interest. Letting $N = \sum_1^K N_i$, with $\lambda_i = N_i/N$, it is easy to establish the following result, using the method of proof in Theorem 1. In working through the proof, it is useful to note that for all i we have $0 < p_i < \lambda_i$ and $0 < q_i < \lambda_i$, so that $\max(0, p_i + q_i - \lambda_i) < \mu_i < \min(p_i, q_i)$.

Theorem 2. For each $i = 1, \ldots, K$, we have

$$\alpha = \log\left\{\frac{\mu_i(\lambda_i - p_i - q_i + \mu_i)}{(p_i - \mu_i)(q_i - \mu_i)}\right\} = g_i(\mu_i), \tag{5}$$

with the inverse relation written as $\mu_i = h_i(\alpha)$. Also $\sigma_i^2 = \{g_i'(\mu_i)\}^{-1} = h_i'(\alpha)$ and $y_i = x_i/N$. Then the approximate distribution of each $z_i = \sqrt{N}(y_i - \mu_i)/\sigma_i$ is independent standard normal.

Corollary. The approximate distribution of the sufficient statistic X is normal, with mean $N\sum_1^K \mu_i = N\sum_1^K h_i(\alpha)$, and variance $N\sum_1^K \sigma_i^2 = N\sum_1^K h_i'(\alpha)$.

Note that it is necessary to work on the α-scale rather than μ_i scales as in the case of a single table. The end-points of a $100(1-\varepsilon)\%$ CI for α are the solutions of

$$\frac{\left\{X - N\sum_1^K h_i(\alpha)\right\}^2}{N\sum_1^K h_i'(\alpha)} = z_{1-\varepsilon/2}^2. \tag{6}$$

In the case of a single table $K = 1$, the criterion on the left-hand side is zero at the estimated $\hat{\alpha}$, and monotone, moving away from $\hat{\alpha}$ on either side. The same properties appear to hold for multiple tables, $K \geq 2$, yielding reliable computation of CI endpoints, but a formal proof has not yet been found.

In calculating the left-hand side of (6) for trial values of α, the derivatives $h_i'(\alpha)$ are awkward, and it is easier to revert back to the μ_i values. The expression for each μ_i is given by inverting (5), requiring careful choice of a quadratic equation root. The general result is $A = e^\alpha - 1, \quad B = -\lambda - (p+q)A, \quad C = pq(A+1)$, and

$$\mu = h(\alpha) = \frac{-B - \sqrt{B^2 - 4AC}}{2A},$$

with subscripts i inserted for $\mu_i, h_i(\alpha), p_i, q_i$ and λ_i. Because g, h are inverse functions of each other, we have $\sum_i h_i'(\alpha) = \sum_i \{g_i'(\mu_i)\}^{-1}$, so that (6) becomes

$$\frac{(X - N\sum_{i=1}^K \mu_i)^2}{N\sum_{i=1}^K \{1/g_i'(\mu_i)\}}. \tag{7}$$

The values of each g_i' are easy to calculate from the expression in (2), which is just $\sigma^2 g'(\mu) = 1$.

An example of the calculation of a CI for the common α value in a series of 2×2 tables in given in Section 6.

5 Stratified Tables with Varying Log-Odds-Ratios

Now consider a series of K stratified 2×2 tables as in Section 4, but with the ith table value α_i of the log-odds-ratio-difference given by $\alpha_i = \omega + \delta u_i$, depending on a known covariate u_i. The coefficient δ measures the influence of the covariates $\{u_i\}$ upon $\{\alpha_i\}$, and is of interest. The parameter ω, measuring a common level of $\{\alpha_i\}$ values, is another nuisance parameter, additional to the parameters $\{\beta_i\}$ already removed by conditioning on marginal success totals $\{s_i\}$.

Likelihood calculations as in Section 2 show that $X = \sum_1^K x_i$ is sufficient for ω and $Y = \sum_1^K u_i x_i$ is sufficient for δ. Conditioning upon X will remove the nuisance parameter ω. Then, following carefully the same method of proof as for Theorem 1 gives this result.

Theorem 3. Under the same conditions as Theorem 2, with $y_i = x_i/N$ and $\lambda_i = N_i/N$, the approximate joint distribution of $\{y_i\}$ as $N \to \infty$, conditional on $X = \sum_1^K x_i = Nt$,

is given by $y_i = \mu_i + N^{-1/2}\sigma_i z_i$; the means $\{\mu_i\}$ are given by

$$g_i(\mu_i) = \delta u_i - c, \qquad \text{or} \qquad \mu_i = h_i(\delta u_i - c),$$

where the constant c is chosen so that $\sum_1^K \mu_i = t$ is satisfied; the variances σ_i^2 are given by

$$\sigma_i^2 = \{g_i'(\mu_i)\}^{-1} = h_i'(\delta u_i - c); \tag{8}$$

and the $\{z_i\}$ are zero-mean normal random variables, with idempotent covariance matrix $\mathrm{cov}(z) = I - vv^T$, where $v_i = \sigma_i/\sqrt{\sum_j \sigma_j^2}$.

Corollary. The approximate conditional distribution of the sufficient statistic $Y = \sum_1^K u_i x_i$ is normal, with mean $N\sum_1^K u_i\mu_i$, and variance

$$N\Big\{\sum_1^K \sigma_i^2 u_i^2 - \frac{[\sum_1^K \sigma_i^2 u_i]^2}{\sum_1^K \sigma_i^2}\Big\}. \tag{9}$$

Remarks. (i) Equation (8) is useful in evaluating the variance in (9). The easiest path is to use the explicit expression for h_i in Section 4 to evaluate $\mu_i = h_i(\delta u_i - c)$, and then the expression for g_i' implicit in (2).

(ii) The value of δ is unchanged if each u_i is replaced by $u_i - b$, for some constant b, and the approximate distribution of Y, as described in Theorem 3, is also unchanged, because the constant c therein is adjusted by any choice of b.

The end-points of a $100(1-\varepsilon)\%$ CI for δ are the solutions of

$$z_{1-\varepsilon/2}^2 = \frac{\{Y - N\sum_1^K u_i\mu_i\}^2}{\mathrm{var}(Y)}, \tag{10}$$

where $\mathrm{var}(Y)$ is given by (9). Section 6 contains an example of the calculation of such a CI, using the results in Theorem 3 and its Corollary. Although there is no formal proof, the

example suggests that the right-hand side of (10) has very regular monotonicity properties, making calculation of the CI end-point solutions an easy numerical task.

6 Example

The following example **shown in Table 1** comes from Kraus et al. (1989), and is reproduced in Greenland (1989). The data is from a 1960 case-control study relating the occurrence of sudden infant death syndrome (SIDS) to marital status. Here, the three marital status categories married, single, and separated/divorced are combined into two, partnered or alone. There are $K = 4$ tables, stratified according to four income levels.

To begin, use the first 2×2 table, the lowest income level, to illustrate the single-table confidence interval method for α, the difference in log-odds-ratio for occurrence of SIDS, between 'partnered' and 'alone' states.

We have $y = 19/281 = 0.06762$, $p = 36/281 = 0.12811$, $q = 132/281 = 0.46975$, and for a 95% CI, $1 + z^2/n = 1.01367$. Thus $u = 0$ and $v = 0.12811$. For the lower root μ_1, beginning with $\mu_{(0)} = 0.01$ gives the iterates presented in Table 2. For the upper root μ_2, beginning with $\mu_{(0)} = 0.1$ gives the output shown in Table 3.

Applying the function $\alpha = g(\mu)$ in (1) to $(\mu_1, \mu_2) = (0.049, 0.086)$ gives the 95% CI for α, the log-odds-ratio for SIDS between the two partnered states, as $(-0.425, +0.958)$.

The corresponding Wald, Agresti-Coull and exact CI's are $(-0.433, +0.965)$, $(-0.433, +0.966)$ and $(-0.495, +1.034)$ and the corresponding lengths are 1.383 (Wilson), 1.400 (Wald and Agresti) and 1.529 (exact). We also use the `R` (R-Development-Core-Team

2006) package `survival` (Therneau and Lumley 2010) with conditional ML fitter `clogit` to double check our own program. It offers the option "exact" or "approximate" conditional ML. The difference between the two options is large, it gives the Wald CI $(-0.351, +0.621)$ with length 0.972. Looking at the other CIs **we prefer to rely** on the exact conditional ML option. In fact for this example `clogit` does not provide an "exact" solution (non-convergence), but only an approximate solution, because it seems the search for the number of permutations is cumbersome.

Now consider calculation of a 95% CI for a common value of α for the first two tables, corresponding to the two lowest income groups. The CI criterion (6) is monotone in α on either side of the zero-value, so is easy to solve numerically, in Excel for example. The resulting 95% CI is $(-0.217, +0.776)$ with length 0.993.

When the same method is applied to find a CI for the assumed common α across all four tables, the result is $(-0.348, +0.424)$ with length 0.772. For multiple tables we have not implemented the exact method, because of the complexity involved for computing the resulting coefficients of the underlying distribution. For the first two tables, the alternative CIs are: (i) Wald $(-0.434, +0.965)$ (length 1.40), (ii) Agresti-Coull $(-0.421, 0.978)$ (length 1.40) and across all four tables, the Wald CI is undefined, because conditional ML estimation fails to provide estimates. Using the option "approximate" of `survival`, the Wald CI is $(-0.176, 0.194)$, which again seems far too small, compared to the Wilson CI $(-0.217, +0.776)$. Unconditional ML estimation gives a Wald CI of $(-0.341, 0.428)$ with length 0.770 which is close to the Wilson CI in this instance.

The CI for four tables is narrower than for the one found from the first two tables only, as is expected from the larger total sample size, but centred more around zero. This reflects the fact that the empirical odds-ratios from the first two tables 1.31 **and** 1.34 are different to those for the last two tables, 0.60 **and** 0.82, suggesting the possibility that the effect of being partnered upon prevention of SIDS may be beneficial in the higher income groups, but the opposite in the lower income groups. Thus the assumption of a common odds-ratio across all four tables is called into question, and the method of Section 5, assessing the effect of an income covariate, becomes relevant.

Therefore, as in Section 5 we propose the model $\alpha_i = \omega + \delta u_i$, for $i = 1, \ldots, 4$, where to distinguish between low and high income levels, the covariate values $-1, -1, +1, +1$ are chosen for $\{u_i\}$. The regression parameter δ measures the influence of the income covariates.

To evaluate a 95% CI for δ by solving (10), an extra layer of computation is needed, because each trial value of δ requires calculation of a tuning constant c to enforce the conditioning constraint $\sum \mu_i = t = 124/1915 = 0.06475196$ which removed the nuisance parameter ω. This can be dealt with in Excel by making simple modifications to the columns used for the common α case of Section 4, and then using Excel's Solver. The resulting values of z^2 on the right-hand side of (10) are monotone in δ on either side of the zero-value, or estimate $\hat{\delta}$. Computation is slower than for the methods of earlier Sections, but reliable nevertheless. The results of a numerical search are as **shown in Table 4**.

A 95% CI for δ is $(-0.71, +0.05)$, just including the null point $\delta = 0$, corresponding to

a non-significant P-value 0.092 for $z^2 = 2.85$. The estimate of δ is $\hat{\delta} = -0.329$. Although the result does not conclusively establish a non-zero value of δ, the analysis does open the possibility that any beneficial effect of being partnered upon the prevention of SIDS among upper income levels is reversed among lower income levels. The function `clogit` does only allow conditioning on one variable, neither have we implemented any of the other methods due to the complexity involved. Unconditional ML estimation gives an estimate of -0.44879 and a Wald CI of $(-0.946, 0.048)$.

7 Simulation Study and Discussion

In order to evaluate the performance of 95% Wilson CIs, we conducted a small-scale simulation study focusing on the single table scenario considered in Section 3. We compared the Wilson CI with (i) the exact CI, which is based on inverting an exact test, see for example Mehta et al. (1985), (ii) the standard Wald-type CI based on conditional maximum likelihood (CML), and (iii) the Agresti-Coull CI (Brown et al. 2001), which has the same form as the Wald-type CI, except that its midpoint - the CML estimate - is replaced by the midpoint of the Wilson CI. Brown et al. (2001) considered CIs for binomial proportions and compared the performance of many CIs, among them the Wilson CI, the Agresti-Coull CI and the Wald-type CI.

We consider a small sample situation, single tables with $n_1 = n_2 = 5, 10, 15, 30$. The Wilson CI is based on a large-sample approximation, hence we would not expect it to perform well for small $n_1 = n_2$. **Figures 1 and 2 show** the exact coverage probabilities

for the log odds ratio $\alpha = \log\{\theta_1(1-\theta_2)/(\theta_2(1-\theta_1))\}$ for all four CIs **and** $n_1 = n_2 = 5, 30$. over a dense grid of parameters θ_1 and θ_2 with $\theta_1, \theta_2 \in \{0.001, 0.002, \ldots, 0.999\}$. Due to the shading involved plotting the coverages along a line gives a better picture. Figures 3 and 4 show the coverages along the line $\theta_1 = 1 - \theta_2$ for $n_1 = n_2 = 10, 30$.

Figures 5 and 6 show the differences in expected length between the exact and the Wald-type CI, and between the Wald-type CI and the Wilson CI **for** $n_1 = n_2 = 5, 15$**. The same is illustrated in Figures 7-8 in form of boxplots.** The exact CI is roughly 30% longer than the Wald CI, the Wilson CI is shortest roughly being 5% shorter than the Wald CI, both statements in average and apply approximately for all four scenarios $n_1 = n_2 = 5, 10, 15, 30$.

The coverage pattern of all CIs is very similar to those reported by Brown et al. (2001), an alternating coverage probability fluctuating around 95%, see Figures 1-8. In particular, the downward spikes for the Wald CI near the boundary of the parameter space is problematic, but less severe than anticipated. The infimum coverage of the Wald CI here is larger ($\approx 89\%$) than that the infimum coverage ($< 70\%$) for the problem of estimating the probability of a binomal proportion, see plot Agresti (2002, p.19), and the infimum coverage probability of the other competing methods, the Agresti-Coull CI ($\approx 93\%$) and the Wilson CI ($\approx$ 89%), are also relatively high. Interestingly for $n_1 = n_2 \geq 10$ the infimum coverage probability for the Wilson CI is larger than that of the Wald CI.

The exact CI is the only method considered here that strongly maintains the required confidence level of 95%. In fact the minimum coverage recorded is 97.5% owing **to** the

"exact" discrete distribution making the exact CI strongly conservative. Attaining infimum coverage probability seems to be the best criteria in evaluating CIs. The exact CI does not attain the specified infimum nominal level of 95%, but that of 97.5%. Therefore we do not consider the exact CI as the optimal CI, but are rather interested in a CI with a coverage around the nominal 95% level. This view is shared for example by Agresti (2002, p.19), who does not consider the Wald CI as a good option due to under-coverage, but also more importantly does not consider the exact CI as good due to strong over-coverage. In fact he considers another CI method as "a good method" with fluctuating coverage around 95%. Following the same argumentation as Agresti, we conclude that the approximate methods considered here are favoured over the exact CI.

The expected length of the Wilson CI is shortest and that of the exact CI longest. Only for very large α, the Wald-type CI has shorter expected length, which can be seen from Figures 9-12. The results also show that the Agresti-Coull interval is better than the Wald-type CI, having the same expected length, but higher coverage probabilities.

Figures 9-12 shows that the length of the Wilson CI is in roughly 91% ($n_1 = n_2 = 5$) - 98% ($n_1 = n_2 = 30$) of the cases shorter than the Wald CI.

Even though our proposed Wilson CI is based on a large sample approximation, the method performs well. Conditionally on some fixed parameter combination of θ_1 and θ_2 it might have lower coverage than 95%, but unconditionally, by averaging over a (relatively small) neighbourhood of the true parameter or alternatively over the whole parameter space, the coverage level is maintained; see Figures 1-8,13-16.

If accuracy, ie the expected length of the CI, is more important than conditional coverage probability, then we suggest the Agresti-Coull CI (favouring coverage) for smaller n_1 and n_2, for larger n_1 and n_2 we recommend the Wilson CI; otherwise if maintaining infimum coverage is most important then there is no alternative to the exact CI. This paper showed that the Wilson CI should also be included in standard statistical packages that give CIs for such single and multiple case scenarios to provide the practitioner with another "shorter" alternative.

The results indicate that the Wilson CI is not only useful for the stratified situations considered in this paper, but may also be more generally for conditional ML and logistic regression. However this needs further investigation before such a generalisation can be made. Furthermore, Brown et al. (2001) derived boundary modifications for the Wilson CI to eliminate undesirable downwards coverage spikes near the boundary of the parameter space. Their method reduced the downward bias dramatically. It remains to be seen whether such a modification can also be developed for the two-sample log odds ratio due to the higher complexity, compared to binomial proportions.

References

Agresti, A.: 2002, *Categorical Data Analysis*, Wiley Series in Probability and Statistics, 2nd edition edn, Wiley.

Breslow, N.: 1981, Odds ratio estimators when the data are sparse, *Biometrika* **68**(1), 73–84.

Brown, L. D., Cai, T. T. and DasGupta, A.: 2001, Interval estimation for a binomial proportion - comment - rejoinder, *Statistical Science* **16**(2), 101–133.

Greenland, S.: 1989, Generalized mantel-haenszel estimators for k 2xj tables, *Biometrics* **45**(1), 183–191.

Hannan, J. and Harkness, W.: 1963, Normal approximation to the distribution of two independent binomials, conditional on fixed sum, *Annals of Mathematical Statistiscs* **34**, 1593–1595.

Kraus, J. F., Greenland, S. and Bulterys, M.: 1989, Risk-factors for sudden infant death syndrome in the united-states collaborative perinatal project, *International Journal of Epidemiology* **18**(1), 113–120.

Mehta, C. R., Patel, N. R. and Gray, R.: 1985, Computing an exact confidence-interval for the common odds ratio in several 2x2 contingency-tables, *Journal of the American Statistical Association* **80**(392), 969–973.

R-Development-Core-Team: 2006, R: A language and environment for statistical computing, `http://www.R-project.org`.

Stevens, W.: 1951, Mean and variance of an entry in a contingency table, *Biometrika* **38**, 468–470.

Therneau, T. and Lumley, T.: 2010, R package survival: Survival analysis, including penalised likelihood, `http://CRAN.R-project.org/package=survival`.

Wilson, E.: 1927, Probable inference, the law of succession, and statistical inference, *Journal of the American Statistical Association* **22**, 209–212.

A Proof of Theorem 1:

Proof. Use Stirling's approximation to the factorial function, $k! \sim k^{k+1/2}e^{-k}\sqrt{2\pi}$ as $k \to \infty$. The hypergeometric probability a_x is given by

$$a_x = \frac{s!\,(N-s)!\,n_1!\,n_2!}{x!\,(s-x)!\,(n_1-x)!\,(n_2-s+x)!\,N!},$$

for integer x with $0 \le x \le \min(n_1, s)$, so that, as a function of x,

$$p_x \quad \propto \quad \frac{e^{\alpha x}}{x!(s-x)!(n_1-x)!(n_2-s+x)!}.$$

Use of Stirling's formula then gives

$$\log p_x \propto \alpha x - (x+\frac{1}{2})\log x - (s-x+\frac{1}{2})\log(s-x) - (n_1-x+\frac{1}{2})\log(n_1-x) - (n_2-s+x+\frac{1}{2})\log(n_2-s+x).$$

Substitute $x = n\mu + \sigma z n^{1/2}$, and re-express this formula for $\log p_x$ in terms of z. This is an elementary although lengthy task. Gathering up the terms in decreasing order of magnitude gives:

(i) the coefficients of $N\log(N), \sqrt{N}\log(N), N$ and $\log(N)$ do not depend on z, so are part of a normalizing constant;

(ii) the coefficient of $N^{1/2}$ is

$$\sigma z\Big[\log\Big\{\frac{(p-\mu)(q-\mu)}{\mu(1-p-q+\mu)}\Big\} + \alpha + o(1)\Big], \quad \to 0, \quad \text{from (1)};$$

and

(iii) the coefficient of the constant term is

$$-\frac{\sigma^2 z^2}{2}\left\{\frac{1}{\mu}+\frac{1}{p-\mu}+\frac{1}{q-\mu}+\frac{1}{1-p-q+\mu}\right\}+o(1), \quad \to -z^2/2, \quad \text{from (2)}.$$

Thus, as a function of z, $\log p_x$ is $constant - z^2/2 + o(1)$, implying a $N(0,1)$ limit distribution for z.

Tables

Table 1. Data from the 1960 SIDS study (Kraus et al. 1989)(Kraus, Greenland and Bulterys, 1989)

income(1960 $)	disease status	partnered	alone
$< 1,500$	case	19	17
	control	113	132
$1,501 - 2,500$	case	40	12
	control	283	114
$2,501 - 3,500$	case	27	10
	control	308	69
$> 3,500$	case	38	5
	control	657	71

Table 2. iteration for the lower root μ_1

j	$\mu_{(j)}$
0	0.01
1	0.0179
2	0.0288
3	0.0399
4	0.0466
5	0.0485
≥ 6	0.0486

Table 3. iteration for the upper root μ_2

j	$\mu_{(j)}$
0	0.1
1	0.09129
2	0.08695
3	0.08611
≥ 4	0.08609

Table 4. Searching for end-points of a 95% CI for δ.

δ	c	z^2
-0.7	0.1096	3.61
-0.6	0.088	1.91
-0.5	0.077	0.76
-0.4	0.061	0.13
-0.3	0.042	0.02
-0.2	0.019	0.43
-0.1	-0.007	1.37
0.0	-0.038	2.85
0.1	-0.073	4.89
-0.712	0.097	3.85
-0.329	0.048	0.00
0.0523	-0.056	3.84

Figures

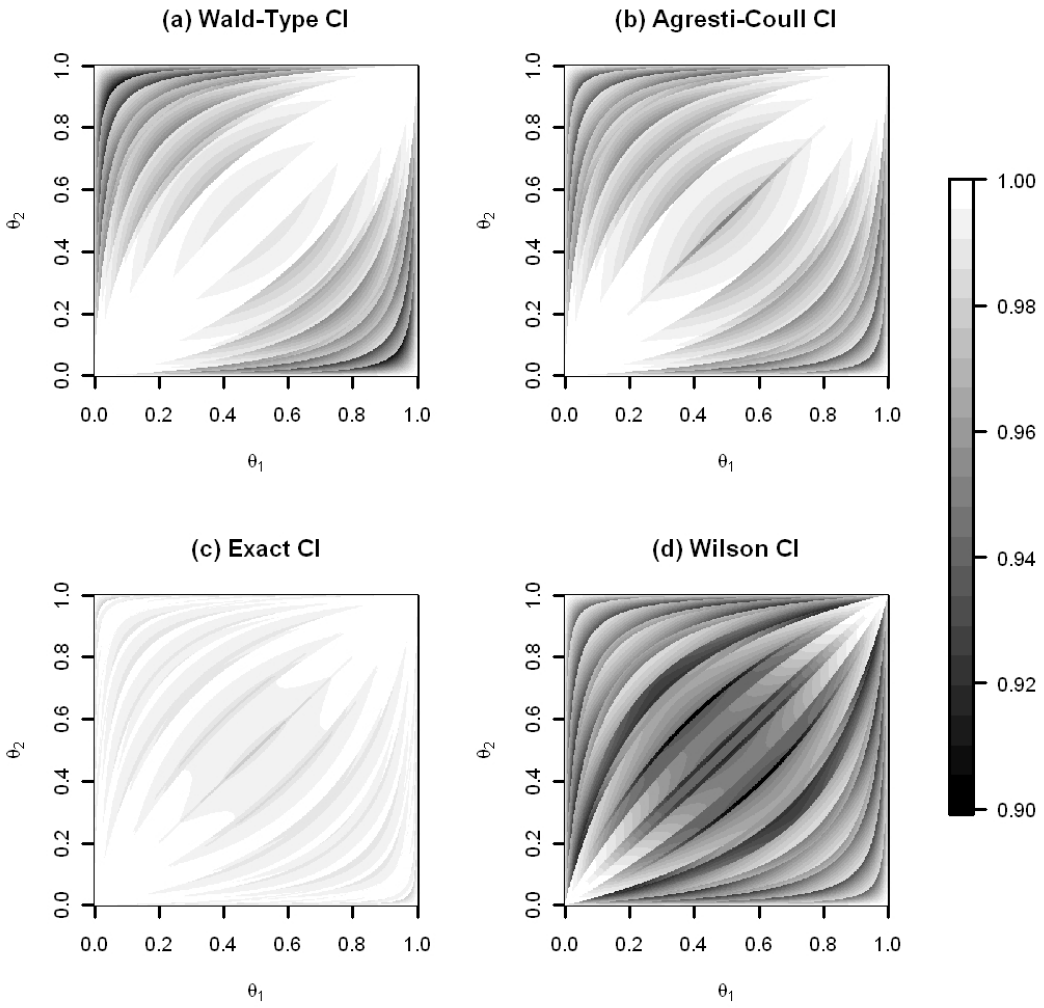

Figure 1: Coverage of confidence intervals for the log odds ratio α for $n_1 = n_2 = 5$: (a) Wald-type, (b) Agresti-Coull, (c) Exact and (d) Wilson.

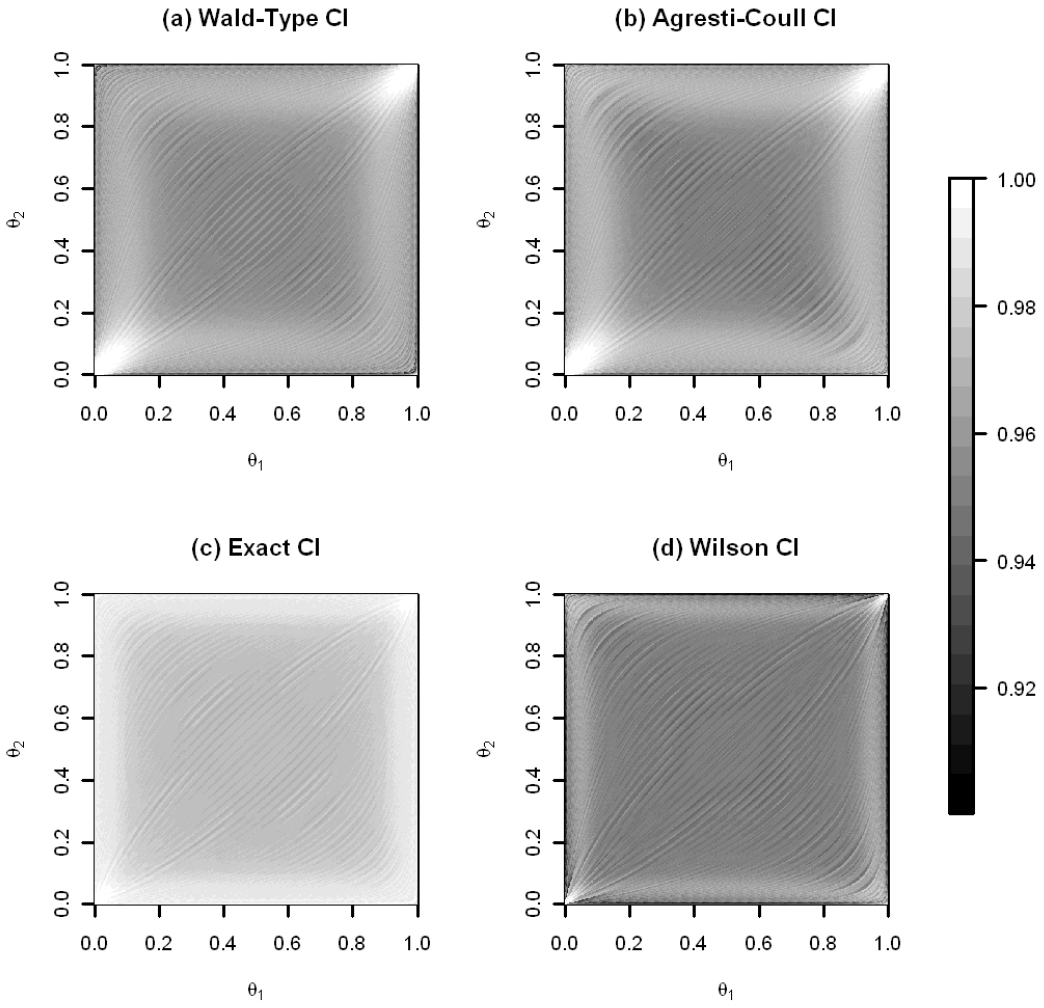

Figure 2: Figure 4: Coverage of confidence intervals for the log odds ratio α for $n_1 = n_2 = 30$: (a) Wald-type, (b) Agresti-Coull, (c) Exact and (d) Wilson.

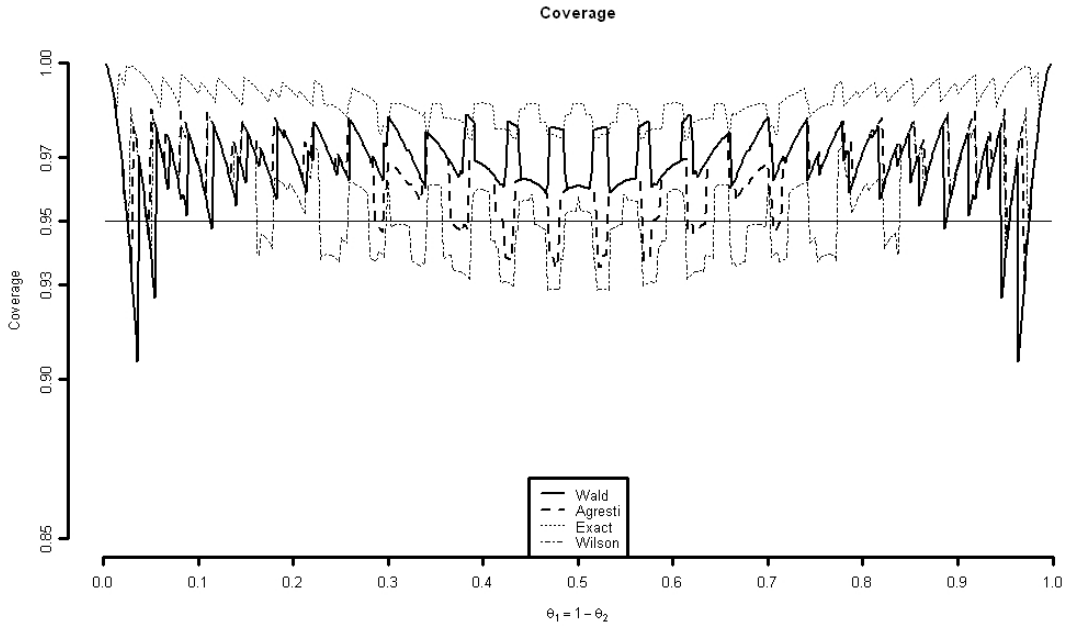

Figure 3: Coverage of confidence intervals for the log odds ratio α for $n_1 = n_2 = 10$ along the line $\theta_1 = 1 - \theta_2$.

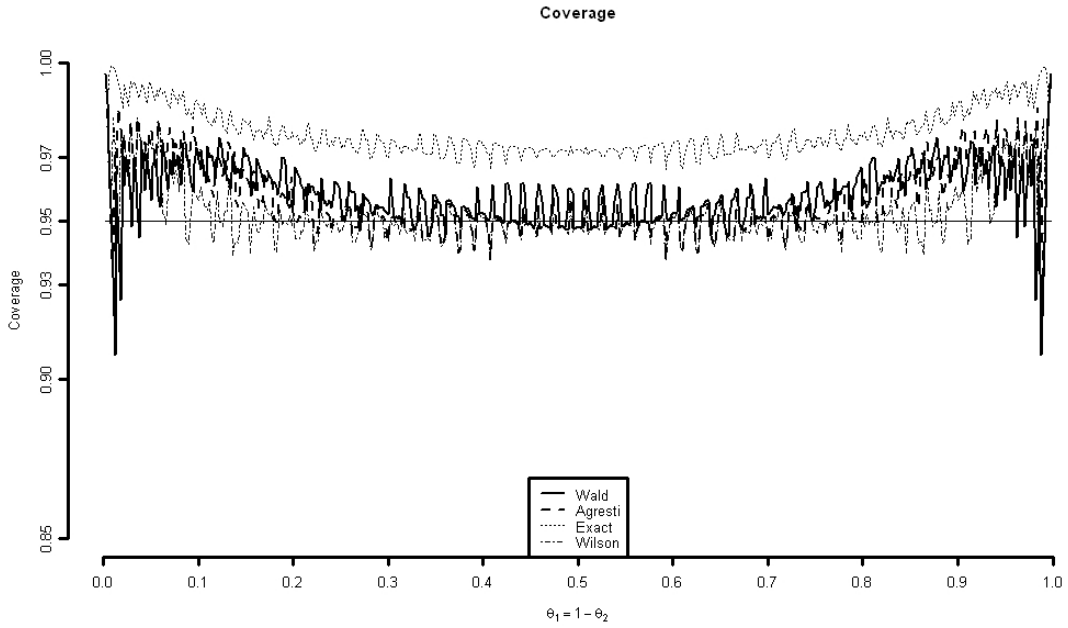

Figure 4: Coverage of confidence intervals for the log odds ratio α for $n_1 = n_2 = 10$ along the line $\theta_1 = 1 - \theta_2$.

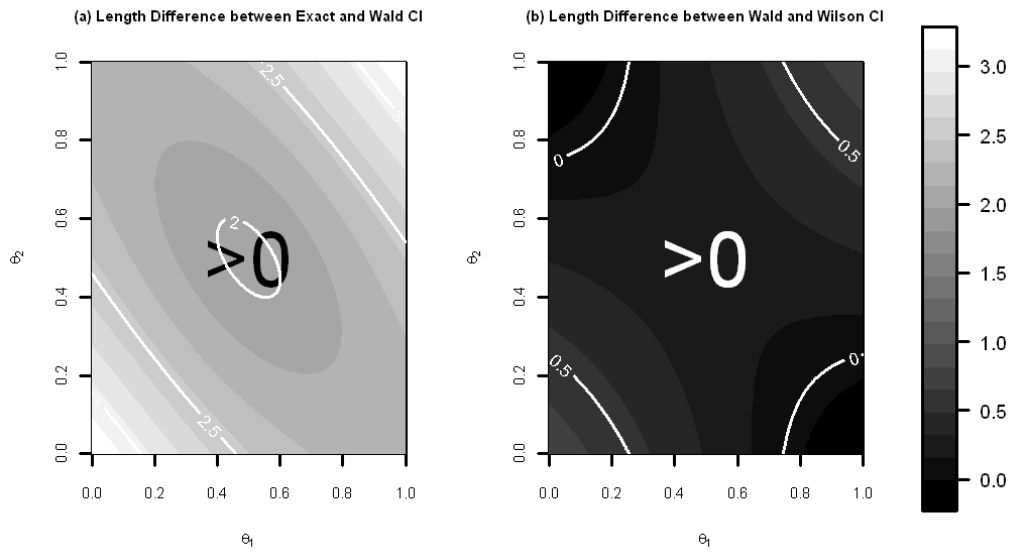

Figure 5: Differences between expected lengths of confidence intervals for the log odds ratio α for $n_1 = n_2 = 5$: (a) Difference between exact and Wald-types, and (b) Difference between the Wald-type and Wilson.

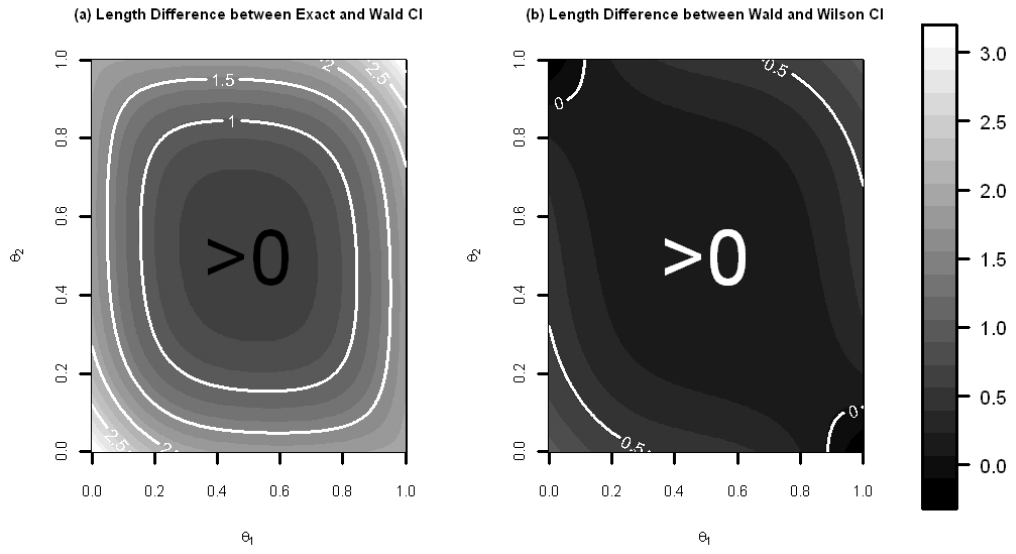

Figure 6: Differences between expected lengths of confidence intervals for the log odds ratio α for $n_1 = n_2 = 15$: (a) Difference between exact and Wald-types, and (b) Difference between the Wald-type and Wilson.

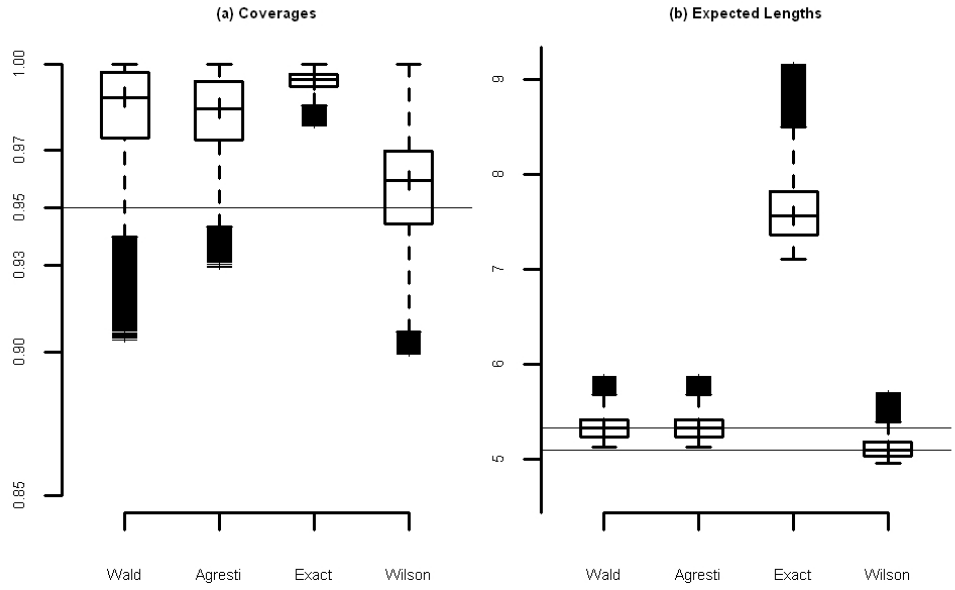

Figure 7: (a) Boxplots for coverage probabilities and (b) Boxplots for the expected lengths for the Wald-type CI, the Agresti-Coull CI, the Exact CI and the Wilson CI for $n_1 = n_2 = 5$.

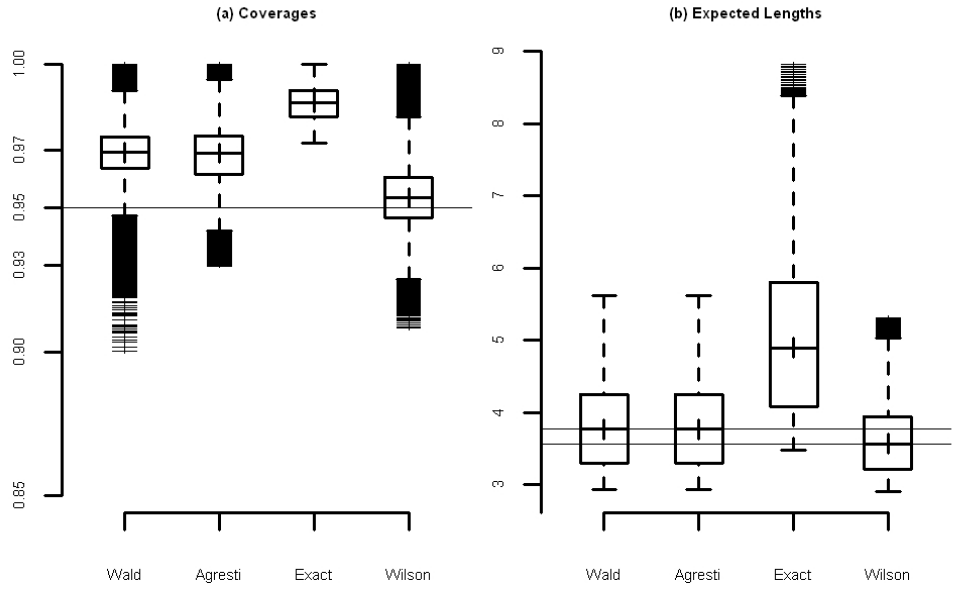

Figure 8: (a) Boxplots for coverage probabilities and (b) Boxplots for the expected lengths for the Wald-type CI, the Agresti-Coull CI, the Exact CI and the Wilson CI for $n_1 = n_2 = 15$.