# ASSESSMENT OF DIFFERENT METHODS OF DETERMINING CONFIDENCE INTERVALS FOR THE DIFFERENCE OF BINOMIAL PROPORTIONS

by

**MARIETTE NEL**

Submitted in fulfillment of the requirements of the degree of

**MASTER IN MEDICAL SCIENCE**

in the

Department of Biostatistics, Faculty of Health Sciences,

University of the Orange Free State, Bloemfontein

Promotor: Ms. G. Joubert

Co-Promotors: Dr. R. Schall

Prof. D.G. Nel

May 1998

**ACKNOWLEDGEMENTS**

I thank

Dr. Robert Schall for his guidance, patience and criticism of my research and thesis.

Prof. Daan Nel for his guidance, patience and criticism of my research and thesis.

Ms. Gina Joubert for her comments during the write-up of this work and willingness to act as promotor at a late stage of the research.

Mrs. Ina Bester for her patience and help with FORTRAN.

Family and friends for their interest and encouragement.

My Heavenly Father for giving me the ability to undertake this study.

## TABLE OF CONTENTS

CHAPTER 3

# LIST OF FIGURES

# LIST OF TABLES

# GLOSSARY

| | |
|---|---|
| $p_i$ | observed population proportion for population $i$ |
| $\hat{p}_i$ | an observed proportion from a sample from population $i$ |
| $X_i$ | number of successes in a Binomial$(n_i, p_i)$, $i = 1,2$ experiment, a random variable |
| $x_i$ | observed number of successes from a Binomial$(n_i, p_i)$ population $i = 1,2$, not a random variable |
| $n_i$ | number of repetitions from population $i = 1,2$ also known as sample size |
| $\rho$ | relative risk |
| $b$ | difference, $p_1 - p_2$ |
| $\theta$ | general notation for a parameter |
| $\psi$ | odds ratio |
| $\beta$ | the probability of failing to find the specified difference to be statistically significant |
| $L$ | lower confidence limit |
| $U$ | upper confidence limit |
| $\Phi$ | the cumulative distribution function of the standard normal distribution |
| $e^Y$ | the exponential function, denoting the inverse procedure to taking logarithms |
| $\Sigma$ | the Greek capital letter sigma, denoting 'sum of' |
| $\infty$ | infinity, the value larger than any imaginable number. Likewise $-\infty$ is the value less than any imaginable negative number. |
| $\alpha$ | significance level of a hypothesis test |

| | |
|---|---|
| $\gamma = 1 - \alpha$ | general notation for the confidence coefficient in a confidence interval |
| $\chi^2$ | a value from the Chi squared distribution, the sampling distribution for test statistics derived from tables of frequencies. |
| $\chi^2_{\nu,\alpha}$ | the $100(1-\alpha)\%$ percentile of the chi-square distribution with $\nu$ degrees of freedom. |
| $z_\alpha$ | the $100(1-\alpha)\%$ percentile of the standard normal distribution |

# CHAPTER 1

# CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

## 1.1 Introduction

The purpose of the statistical analysis of medical research data is to make observations on a sample of subjects and then draw inferences, in some instances about the population of all such subjects from which the sample is drawn. Even a well-designed study can give only an idea of the answer sought because of random variation in the sample. Results from a single sample are thus subject to statistical uncertainty, and this uncertainty is related, among others, to the size of the sample.

An example of the statistical analysis of data would be calculating an estimate of a parameter and a confidence interval for the parameter. The confidence interval indicates the precision of the estimate. For example, if the parameter of interest is a difference between the proportions of patients improving on two different treatments, and if the independent sample proportions of patients improving in two groups of 60 patients each receiving the treatments are respectively 75% and 55%, then the difference of 20% can be reported with a 95% confidence interval of [ 3.3% ; 36.7% ]. This interval was calculated using the Asymptotic normal without continuity correction method. The estimate of 20% is imprecise, but the imprecision is incorporated into the presentation of findings through reporting a confidence interval.

Confidence intervals reflect the uncertainty associated with the findings of a study directly on the scale of the original measurement. This has advantages over the practice of only reporting P-values, which are usually referred to as 'significant' or 'non-significant'. A confidence interval gives a range of values that contains the true population value of the parameter in question with the chosen confidence. The confidence level usually used is 95%.

As pointed out above, a single study gives an imprecise estimate of the true population value in which we are interested. The imprecision is indicated by the width of the confidence interval, the wider the confidence interval the less the precision. The width of the confidence interval depends on three factors. Firstly, larger sample sizes will give more precise results than smaller samples and thus narrower confidence intervals. Thus wide confidence intervals may emphasise the unreliability of conclusions based on small samples. During the planning stage of a study it is possible to estimate the sample size that should be used by stating the width of the confidence interval required at the end of the study and carrying out the appropriate calculation for the sample size. Altman(1992) noted that the determination of an appropriate sample size is a common task in the planning of clinical trials. Secondly, the smaller the variability of the data(between subjects, within subjects, from measurement error and from other sources) the more precise the sample estimate and the narrower the confidence interval will be. Thirdly, the higher the confidence level required, the wider the confidence interval will be.

According to Gardner and Altman(1989) the British Medical Journal(BMJ) expects researchers submitting scientific papers to use confidence intervals when appropriate. The BMJ also wants a reduced emphasis on the presentation of P-values. The Lancet, the Medical Journal of Australia, the American Journal of Public Health, and the British Heart Journal, among others, have adopted this policy, and it has been endorsed by the International Committee of Medical Journal Editors. It is therefore important that, with different analyses of medical data, the user should be able to calculate the relevant confidence intervals accurately. In the present thesis confidence intervals of the difference of binomial proportions based on two independent samples are studied.

In this chapter a general introduction is given. In chapter 2 confidence intervals are discussed in general and then the different methods for calculating confidence intervals for the difference of binomial proportions found in literature are discussed. In chapter 3 the different confidence interval methods discussed in chapter 2 are compared through simulation with respect to length and coverage probability of the respective intervals. In the Appendix the FORTRAN computer program to calculate all the mentioned confidence intervals is given.

## 1.2 Comparison of proportions in medical research

### 1.2.1 Areas of application

A comparison of proportions is performed in various areas of medical research.

*Epidemiology*: In epidemiology the risk of disease in different groups of subjects is often compared. For example, the risk of developing lung cancer (proportion of subjects who develop lung cancer) in a group of smokers is compared to the risk of developing lung cancer in a non-smoker group.
*Clinical trials*: The estimation of response or cure rates is an important aspect in the analysis of many clinical trials; for example, the response rate (proportion of patients that are cured) to two different drug treatments may be compared.

### 1.2.2 Statistical measures for the comparison of proportions

#### 1.2.2.1 Proportions

A proportion is calculated as : $\hat{p}_i = \dfrac{x_i}{n_i}$,

where $x_i$ = Number of successes from population $i = 1,2$

and $n_i$ = Number of repetitions from population $i = 1,2$

Note that $\hat{p}_i$ is an estimate for the population parameter $p_i$ which is the probability for a success in population $i = 1,2$. We assume the number of successes is Binomially distributed $X_i \sim \mathrm{Bin}(n_i, p_i)$. A proportion can lie in the interval [ 0 ; 1 ], so the difference between two proportions can lie only in the interval [ -1 ; 1 ], which we will call the definition interval.

#### 1.2.2.2 Difference, odds ratio and relative risk

Three statistical measures often used to compare two proportions $p_1$ and $p_2$ are:

(i) the difference $b = p_1 - p_2$, also called the risk difference,

(ii) the odds ratio $\psi = p_1(1-p_2)/(1-p_1)p_2$ and

(iii) the relative risk $\rho = p_1 / p_2$.

Santner and Yamagami(1993) noted that the odds ratio is the most difficult of the three to interpret although it is the easiest for which to calculate confidence intervals(at least in the sense that an exact confidence interval can be calculated in some situations). Interpretation of the relative risk and of the difference is relatively easy but confidence intervals are more difficult to construct. Anbar(1983) noted that although the odds ratio is a natural parameter, the difference seems to be of more meaning to the clinician.

## 1.3 Problems with confidence intervals for the difference

### 1.3.1 Unavailability of standard software

Confidence intervals for the difference, odds ratio and the relative risk are provided by most statistical software, like the SAS® Procedure FREQ (SAS/STAT Software: Changes and Enhancements through Release 6.12, Chapter 9, 1997) and CIA (Gardner,1989). Unfortunately, accurate confidence intervals for the difference are not available in these software packages.

Thus one of the problems with confidence intervals for the difference of proportions has been that the methods needed to calculate such confidence intervals accurately are not readily available in statistical textbooks and software. Standard textbooks and software usually provide only the Asymptotic intervals (see section 2.1.1) which in general are satisfactory only for large samples and proportions close to 0.5, [Fleiss(1981), Gardner and Altman(1989), Zar(1984), Rosner(1990), Altman(1992), Snedecor and Cochran(1980)].

As will be discussed below, the conventional Asymptotic methods are prone to give intervals that do not make sense, in medical and statistical terms. Brenner and Quan(1990) noted that it has long been known that the conventional Asymptotic interval methods are problematic. As most standard statistical software packages have nothing better to offer to the user than the conventional Asymptotic interval method, this method, despite its shortcomings, is continued to be used, mainly because of its ease of calculation.

### 1.3.2 Shortcomings of the conventional Asymptotic interval method.

Problems that can occur when using the conventional Asymptotic interval method are the following:

#### 1.3.2.1 Intervals of zero length

Consider the situation when we want to calculate the confidence interval for the difference between the proportion of patients receiving two different types of drug who develop an illness. Suppose for illustration that nobody develops an illness in either of the two groups; the resulting Asymptotic confidence interval for the difference would then be [ 0 ; 0 ]. The interval is inappropriately of zero length, and equal to the observed difference, i.e. 0. Newcombe(1995:b) calls this type of violation bilateral Maximum Likelihood Estimation(MLE) tethering. MLE tethering occurs when either the upper confidence limit $U$, or the lower confidence limit $L$ of the confidence interval [ $L$ ; $U$ ] is equal to $\hat{b}$, the estimated difference between the two proportions. This is an infringement of the principle that the confidence interval [ $L$ ; $U$ ] should represent some "margin of error", on either side of $\hat{b}$ at least when $\hat{b} \neq -1$ and $\hat{b} \neq 1$. Bilateral MLE tethering, $L = \hat{b} = U$, constitutes a degenerate or zero-width interval(ZWI) and is always inappropriate.

### 1.3.2.2 Violation of the definition interval

In the same situation as in 1.3.2.1, suppose that all patients on drug A develop the illness and no patient on drug B develops the illness. Suppose that the number of patients on drug A is 20 and on drug B is 10. The Asymptotic confidence interval(see (2.2.2)) is then [ 0.93 ; 1.08 ], when using the continuity correction. The interval inappropriately extends over the definition interval [ $-1$ ; 1 ]. This is called overt overshoot(Newcombe(1995:b)).Overt overshoot occurs when $U > 1$ or $L < -1$. $U = 1$ is not counted as a violation when $\hat{b} = 1$, and the same applies for $L = -1$ when $\hat{b} = -1$.

### 1.3.2.3 Unsatisfactory coverage

Confidence intervals calculated using the conventional Asymptotic method may have actual coverage considerably different from the specified(nominal) coverage.

## 1.4 Objectives of the study

In this study the coverage and length of confidence intervals of different confidence interval methods for the difference between two Binomial proportions will be investigated. Recommendations are made as to which interval methods should be used with respect to different sample sizes. Suitable interval methods are recommended after regard to the ease and speed of computation involved with each method. These methods are made available as easy-to-use programs.

# CHAPTER 2
# CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

## 2.1 Methods to calculate confidence intervals in general

### 2.1.1 Definition of a confidence interval

A confidence interval can be defined as follows(A.M. Mood, et al (1974)): Let $X_1,\ldots,X_n$ be a random sample from the density $f(x,\theta)$, and $\tau(\theta)$ be some function of $\theta$. Furthermore, let $\mathrm{T}_1 = T_1(X_1,\ldots,X_n)$ and $\mathrm{T}_2 = T_2(X_1,\ldots,X_n)$ be two statistics(i.e. functions of $X_1,\ldots,X_n$) satisfying $T_1 \le T_2$ for which $\mathrm{P}_\theta[T_1 < \tau(\theta) < T_2] \equiv \gamma$ holds, where the probability $\gamma$ does not depend on $\theta$. If $t_1$ and $t_2$ are respectively values of $T_1$ and $T_2$ obtained from an observed random sample $x_1,\ldots,x_n$, then the interval $[t_1,t_2]$ is called a $100\gamma$ percent interval for $\tau(\theta)$; $\gamma$ is called the confidence coefficient. We will denote a $100\gamma\%$ confidence interval for $\tau(\theta)$ as $\mathrm{CONF}(\tau(\theta))_\gamma = [t_1,t_2]$. The value $t_1$ is called the lower confidence limit(LCL) and is also denoted as $L$. The value $t_2$ is called the upper confidence limit(UCL) and denoted as $U$.

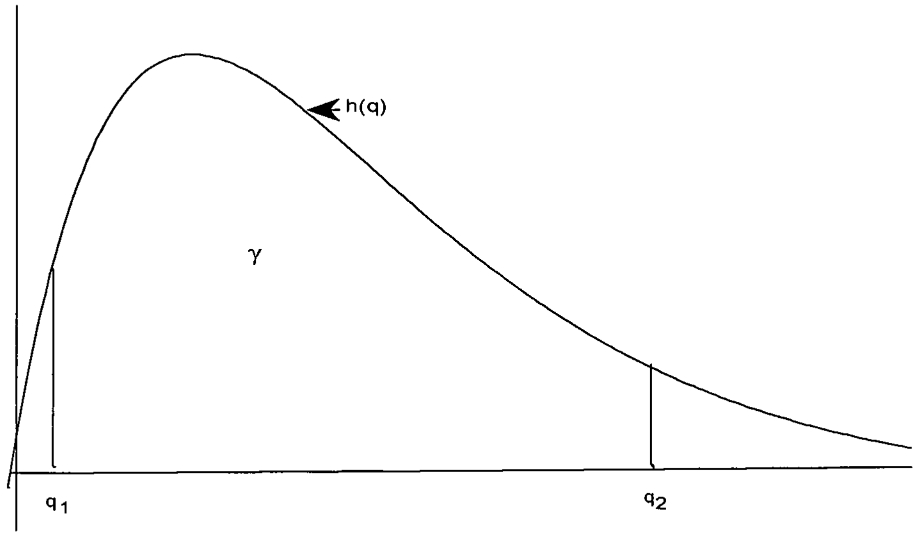
One or the other, but not both, of the two statistics $T_1 = T_1(X_1,\ldots,X_n)$ or $T_2 = T_2(X_1,\ldots,X_n)$ may be constant; that is, one of the two end points of the interval $[t_1,t_2]$ may be constant in which case the interval $[t_1,t_2]$ will be called a one-sided confidence interval. Confidence intervals can be obtained by several methods. The most frequently used methods are the pivotal quantity method and the statistical method. We first consider the pivotal quantity method.

### 2.1.2 The pivotal quantity method.

A pivotal quantity can be defined as follows(Mood, et al (1974)):
Let $X_1,\ldots,X_n$ be a random sample from the density $f(x,\theta)$ and $Q = q(X_1,\ldots,X_n;\theta)$ be a function of $X_1,\ldots,X_n$ and of the parameter $\theta$. If $Q$ has a distribution that does not depend on $\theta$, then $Q$ is said to be a pivotal quantity.

The pivotal quantity method(Mood, et al (1974)), works as follows: If $Q = q(X_1,\ldots,X_n;\theta)$ is a pivotal quantity and has a probability density function $h(q)$, then for any fixed $\gamma$ where $0<\gamma<1$, there exist values $q_1$ and $q_2$ depending on $\gamma$ such that $P[q_1 \le Q \le q_2] = \gamma$.

**Figure 2.1.2.1 Density function $h(q)$ of $Q$.**




Now, for each possible sample value $(x_1,\dots,x_n)$ it follows that: $q_1 \le Q \le q_2$ if and only if two statistics $T_1$ and $T_2$ exist such that $T_1(x_1,\dots,x_n) \le \tau(\theta) \le T_2(x_1,\dots,x_n)$. If $t_1$ and $t_2$ are the values of these statistics $T_1$ and $T_2$ then $[t_1,t_2]$ is a $100\gamma\,\%$ confidence interval for $\tau(\theta)$, since $q_1 \le Q \le q_2$ if and only if $T_1 \le \tau(\theta) \le T_2$. But $P[q_1 \le Q \le q_2] = \gamma$ so $\mathrm{P}_\theta[T_1 \le \tau(\theta) \le T_2] = \gamma$ and according to definition 2.1.1 the interval $[t_1,t_2]$ is a $100\gamma$ % confidence interval for $\tau(\theta)$.

A large sample confidence interval for a Bernoulli parameter $p$, i.e. for the probability of a success in a single Bernoulli trial, can be obtained by using the pivotal quantity method as follows. Let $Z_1,\dots,Z_n$ denote the elements(random variables) of a sample of size $n$ from a Bernoulli distribution with probability function:

$$f(z) = p^z(1-p)^{1-z}, \qquad z = 0,1$$

where $z$ denotes the number of successes in one Bernoulli trial.

Now $X = Z_1 + \ldots + Z_n$ = number of successes in $n$ trials has a Binomial distribution with probability function

$$g(x) = \binom{n}{x} p^x (1-p)^{n-x}; \qquad x = 0,1,2,\ldots,n.$$

Note that the proportion $p$ in the Bernoulli probability function and $p$ in the Binomial probability function is the same. Thus this proportion is called a Bernoulli or a Binomial proportion.

When the sample size is "large enough" it follows from the central limit theorem that the mean of $Z_1,\ldots,Z_n$ namely $\bar{z} = (Z_1 + \ldots + Z_n)/n = X/n$ is approximately normal distributed with mean $p$ and variance $p(1-p)/n$, denoted:

$$\frac{1}{n}X \sim N\left(p\,,\ \frac{p(1-p)}{n}\right).$$

The statistic $X/n$ is also the maximum likelihood estimator for the parameter $p$ and denoted as $\hat{p} = X/n$. According to Zar(1984) the normal approximation can be used if $p$ is neither very small(i.e., close to 0) nor very large(close to 1). Cochran(1977), in Snedecor and Cochran(1980), offered the following sample size recommendations for different magnitudes of $\hat{p}$:

**Table 2.1.2.1 Sample size recommendations for different magnitudes of $\hat{p}$**

| $\hat{p}$ | n |
|---|---|
| 0.5 | $\geq 30$ |
| 0.4 or 0.6 | $\geq 50$ |
| 0.3 or 0.7 | $\geq 80$ |
| 0.2 or 0.8 | $\geq 200$ |
| 0.1 or 0.9 | $\geq 600$ |
| 0.05 or 0.95 | $\geq 1400$ |
| $\sim 0$* | $\infty$ |

* When $\hat{p}$ is extremely small, $n\hat{p}$ follows the Poisson distribution.

Now

$$Q = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

is a pivotal quantity and has an approximate normal distribution N(0,1). Thus two values $q_1 = -z_{\alpha/2}$ and $q_2 = z_{\alpha/2}$ exist such that $P(q_1 \leq Q \leq q_2) = 1 - \alpha$. The inequality $q_1 \leq Q \leq q_2$ can be rewritten as:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$

Since $\sqrt{p(1-p)/n} \approx \sqrt{\hat{p}(1-\hat{p})/n}$ it follows that

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and consequently

$$\mathrm{CONF}(p)_{1-\alpha} = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \qquad (2.1.2.1)$$

The normal approximation usually does not give accurate confidence limits; according to Vollset(1993), it is especially poor when $np$ or $nq$ is less than 5, or when $p$ is near 0 or 1. Furthermore, the normal approximation gives confidence limits symmetric around $p$, which can result in the computation of a nonsensical confidence limit where the lower confidence limit(LCL) $L$ for $p$ could be less then 0 or the upper confidence limit(UCL) $U$ for $p$ could be more than 1.

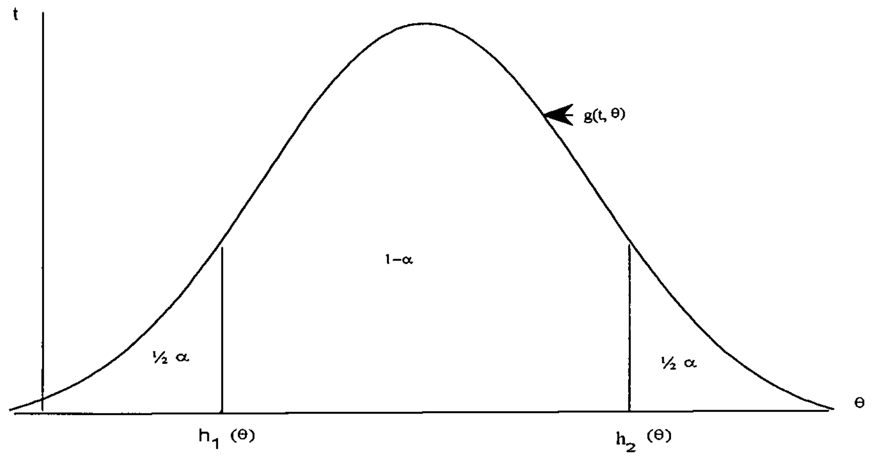
### 2.1.3 The statistical method to determine confidence intervals

The pivotal quantity method for calculating a confidence interval for $p$, described above, is only valid for large sample sizes. When the sample size $n$ is so small that the central limit theorem cannot be used, one can use the "Statistical Method" to construct a confidence interval for $p$.

Suppose $X_1,\ldots,X_n$ are the elements of a random sample from $f(x,\theta)$ and suppose $T = T(X_1,\ldots,X_n)$ is a statistic. Assume that $T$ is continuous(a similar method exists for discrete statistics) with probability density function $g(t,\theta)$. Now two functions $h_1(\theta)$ and $h_2(\theta)$ can be defined such that $P(T \le h_1(\theta)) = \alpha/2$ and $P(T \ge h_2(\theta)) = \alpha/2$ where $0 < \alpha < 1$ is a very small real number and where $h_1(\theta) < h_2(\theta)$ for all $\theta$.

This is shown graphically in Figure 2.1.3.1.

**Figure 2.1.3.1 Distribution of $T$.**




As in Figure 2.1.3.2, sketch $h_1(\theta)$ and $h_2(\theta)$ as functions of $\theta$ and let $t^*$ denote the value of $T = T(X_1,\ldots,X_n)$, obtained from a single random sample e.g. $t^* = T(x_1,\ldots,x_n)$. Draw a horizontal line at $t = t^*$ to intersect $h_1(\theta)$ and $h_2(\theta)$ respectively at $\theta$-values $v_1$ and $v_2$. The interval $[v_1, v_2]$ is then a (1-α)% confidence interval for θ.

To prove this statement let $\theta^*$ be the true value of $\theta$. Then from Figure 2.1.3.2 we see that

$$h_1(\theta^*) \le t^* \le h_2(\theta^*)$$

if and only if $v_1 \le \theta^* \le v_2$ for any observed $t^*$. But according to the definition of $h_1$ and $h_2$:

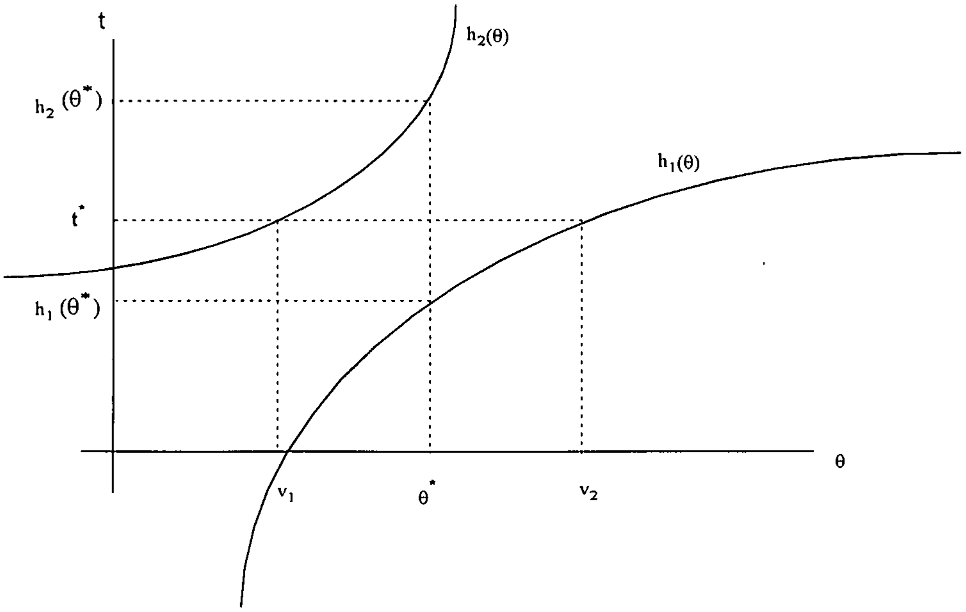
$$P\left(h_1(\theta^*) \le T \le h_2(\theta^*)\right) = 1 - \alpha,$$

so $$P(v_1 \le \theta \le v_2) = 1 - \alpha$$

and consequently according to definition 2.1.1

$$CONF(\theta)_{1-\alpha} = [v_1, v_2].$$

**Figure 2.1.3.2** $h_i(\theta)$ **as functions of** $\theta$**;** $i = 1,2$.




It is not really necessary to sketch the functions $h_1(\theta)$ and $h_2(\theta)$ to find the confidence interval. Note that $v_2$ is the value of $\theta$ where the line $t = t^*$ and the curve $t = h_1(\theta)$ intersect. Thus we solve for $\theta$ from:

$$\frac{1}{2}\alpha = P(T \le t^*)$$

$$= \int_{-\infty}^{\infty} g(t,\theta)\, dt \text{ for continuous } t \in (-\infty, \infty)$$

$$= \sum_{t=a}^{t^*} g(t,\theta) \text{ for discrete } t \text{ defined over [a,b].}$$

Similarly, to find $v_1$ we solve for $\theta$ from

$$\frac{1}{2}\alpha = P(T \ge t^*)$$

$$= \int_{t^*}^{\infty} g(t,\theta)\, dt \text{ for continuous } t \in (-\infty, \infty)$$

$$= \sum_{t=t^*}^{b} g(t,\theta) \text{ for discrete } t \text{ defined over [a,b].}$$

We illustrate the method by finding a confidence interval for a Bernoulli parameter $p$ based on a small sample.

Suppose a Bernoulli experiment is conducted $n = 4$ times and that there is $x = 1$ success. We wish to find a 95% confidence interval for $p$. Note that since $n$ is very small we cannot use the interval(2.1.2.1) which is valid only for large $n$. Consider $T = X$ as our statistic which has a Binomial Bin(4, $p$) distribution where $p$ is the probability of a success. Note that $t^* = 1$ indicates only one success. To get $v_2$ we solve for $p$ from the probability equation $P(T \le t^*) = 0.025$

$$\text{i.e.} \quad P(T \le 1) = 0.025$$

$$\binom{4}{0}(1-p)^4 + \binom{4}{1} p(1-p)^3 = 0.025$$

$$(1-p)^3(1+3p) = 0.025.$$

Let $f(p) = (1-p)^3(1+3p) = 0.025$ then if $p = 0.805$ we get $f(0.805) = 0.0253 \approx 0.025$. Thus $v_2 = 0.805$.

To find $v_1$ solve for $p$ from the probability equation $P(T \ge 1) = 0.025$.

$$\text{Thus} \quad P(T < 1) = 0.975$$

$$P(T = 0) = 0.975$$

$$\binom{4}{0} p^0 (1-p)^4 = 0.975$$

$$p = 0.0063 = \nu_1 .$$

Thus the 95% confidence interval for $p$ is: $\text{CONF}(p)_{0.95} = [\ 0.0063\ ;\ 0.805\ ]$.

## 2.2 Methods to calculate confidence intervals for the difference between two proportions

### 2.2.1 Introduction

We are interested in finding confidence intervals for the difference between two Bernoulli(or Binomial) proportions $p_1$ and $p_2$ from two independent Bernoulli distributions with probability functions $f_j(z) = p_j^z (1-p_j)^{1-z}$ where $z = 0,1$ and j=1,2. If a random sample $Z_{1j}, \dots, Z_{nj}$ of size $n_j$ is taken from the Bernoulli distribution $j$ then the sum $X_j = \sum_{i=1}^{n_j} Z_{ij}$ of these random elements $Z_{ij}$ has a Binomial distribution $B(n_j, p_j)$ with probability function:

$$g(x_i) = \binom{n_j}{x_i} p_j^{x_j} (1-p_j)^{n_j - x_j} \text{ , for } i,j = 1,2.$$

If $n_j p_j$ is "large"(usually $n_j p_j > 5$) then the Binomial distribution $B(n_i, p_i)$ can be approximated by the normal distribution with mean $n_i p_i$ and variance $n_i p_i (1-p_i)$ e.g. $X_i \sim N(n_i p_i ; n_i p_i (1-p_i))$. Approximate confidence intervals can be obtained for each $p_i$( or $q_i = 1 - p_i$) using the normal approximation to the binomial distribution as in section 1.2, in combination with the pivotal quantity method(see section 2.2.2).

### 2.2.2 The conventional approximate confidence interval based on the pivotal quantity method

An approximate confidence interval for differences between two proportions can be derived as follows. If $\hat{p}_1 = X_1 / n_1$ is the estimator of $p_1$ from the first sample, and $\hat{p}_2 = X_2 / n_2$ the estimator of $p_2$ from the second, where $X_1 \sim \text{Bin}(n_1, p_1)$ and $X_2 \sim \text{Bin}(n_2, p_2)$ independently, then approximately according to the Central Limit Theorem

$$\hat{p}_1 = \frac{X_1}{n_1} \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

independently from

$$\hat{p}_2 = \frac{X_2}{n_2} \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right).$$

Thus approximately

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2; \left(\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)\right)$$

and

$$Q = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim \mathrm{N}(0,1).$$

$Q$ is a pivotal quantity for $p_1 - p_2$. Consequently two values $q_1 = -z_{\alpha/2}$ and $q_2 = z_{\alpha/2}$ exist such that $P(q_1 \le Q \le q_2) = \gamma = 1-\alpha$. But $q_1 \le Q \le q_2$ can be rewritten in the form:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Since $p_1$ and $p_2$ are unknown, $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ is replaced by $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ and we get the $100(1-\alpha)\%$ large sample confidence interval for $b = p_1 - p_2$ as:

$$\mathrm{CONF}(p_1 - p_2)_{1-\alpha} = (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}. \quad (2.2.2.1)$$

This confidence interval corresponds to the acceptance region of the $\chi^2$ test for a 2x2 contingency table without correcting for continuity. This interval is also called the "conventional confidence interval" for the difference $p_1 - p_2$. This interval method is easy to calculate with a calculator and is used in the software SAS® Procedure FREQ(SAS/STAT Software: Changes and Enhancements through Release 6.12, Chapter 9, 1997) and CIA(Gardner, 1989) as the interval method to calculate the confidence interval for the difference $p_1 - p_2$.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$ then the 95% confidence interval is given by $\mathrm{CONF}(p_1 - p_2)_{0.95}$ = [ 0.0575 ; 0.3425 ].

### 2.2.3 Approximate confidence interval with continuity correction

Yates(1934) proposed a modification of this confidence interval, using a continuity correction, yielding the confidence interval

$\mathrm{CONF}(p_1 - p_2)_{1-\alpha} =$

$$(\hat{p}_1 - \hat{p}_2) \pm \left[ z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} + \left( \frac{1}{2n_1} + \frac{1}{2n_2} \right) \right]. \qquad (2.2.3.1)$$

This interval method is easy to calculate with a calculator.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$ then the 95% Yates confidence interval is given by $\mathrm{CONF}(p_1 - p_2)_{0.95}$=[ 0.0441 ; 0.3559 ].

The correction takes account of the fact that a continuous distribution(the chi square and normal, respectively) is being used to represent the discrete distribution of sample frequencies. Studies of the effects of the continuity correction have been made by Vollset(1993), who compared 13 methods for computing binomial intervals. He strongly discourages the "standard textbook" method and its "continuity corrected" version. This is also an opinion held by several other authors, such as Ghosh(1979), Blyth and Still(1983), Storer and Kim(1990), Edwardes(1994) and Böhning(1994). Other authors, like Fleiss(1981) and Gardner and Altman(1989), however, still recommend that the correction should be used because the incorporation of the correction for continuity brings probabilities associated with $\chi^2$ and Z into closer agreement with the exact probabilities than when it is not applied. Altman(1992) mentions that it is advisable to use a continuity correction when comparing two proportions, especially when the samples are small. The effect is to reduce slightly the observed difference between the two proportions.

### 2.2.4 The Hauck and Anderson(1986) interval

Hauck and Anderson(1986) proposed another modification of the conventional approximate confidence interval(2.2.2). They performed a simulation study to compare seven confidence interval methods for the difference of two binomial probabilities based on the normal approximation. Their approach was directed toward identifying and comparing methods for widening the uncorrected intervals, treating the Yates correction as giving the maximum widening. They considered different choices of standard errors and several continuity corrections. They recommended the use of the continuity correction $\frac{1}{2\min(n_1, n_2)}$ combined with the use of $(n_i - 1)$ rather than $n_i$ in the estimate of the standard error. This yields the following confidence interval:

$$\mathrm{CONF}(p_1 - p_2)_{1-\alpha} = (\hat{p}_1 - \hat{p}_2) \pm \left[ z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 - 1}} + \frac{1}{2\min(n_1, n_2)} \right] \tag{2.2.4.1}$$

This interval method is easy to calculate with a calculator.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% Hauck and Anderson confidence interval is given by

$$\mathrm{CONF}(p_1 - p_2)_{0.95} = [\ 0.0435\ ;\ 0.3565\ ].$$

Berry(1990) comments that the Hauck and Anderson confidence interval offers better performance than the conventional confidence interval. He also suggests that when both observed proportions are equal to zero, or both are equal to one, the lower confidence limit could be taken as $-1+(\alpha/2)^{1/n_2}$ and the upper limit as $1-(\alpha/2)^{1/n_1}$.

### 2.2.5 Improvements to the conventional approximate interval

#### 2.2.5.1 Unified method for constructing confidence intervals

The following unified method can be used to improve the conventional approximate confidence interval for $p_1 - p_2$. Let $X_1 \sim \mathrm{Bin}(n_1, p_1)$ independent of $X_2 \sim \mathrm{Bin}(n_2, p_2)$ and denote $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. The following reparametrization is useful: $a = p_1 + p_2$; $b = p_1 - p_2$; $u = (1/4)(1/n_1 + 1/n_2)$; and $v = (1/4)(1/n_1 - 1/n_2)$. Let $\hat{a}$, $\hat{b}$, $\hat{p}_1$ and $\hat{p}_2$ be maximum likelihood estimators of $a$, $b$, $p_1$ and $p_2$.

Let

$$\begin{aligned} V(a,b;u,v) &= Var(\hat{b}) = Var(\hat{p}_1 - \hat{p}_2) \\ &= Var(\hat{p}_1) + Var(\hat{p}_2) - 2Cov(\hat{p}_1, \hat{p}_2) \end{aligned}$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$= \frac{1}{n_1}\left(p_1 - p_1^2\right) + \frac{1}{n_2}\left(p_2 - p_2^2\right)$$

$$= u\left[(2-a)a - b^2\right] + 2v(1-a)b.$$

The approach to find confidence intervals for $b = p_1 - p_2$ (see Beal(1987)) is based upon solving for $b$ from the equation:

$$\left(b - \hat{b}\right)^2 = c\mathrm{Var}(\hat{b})$$

$$= c\mathrm{V}(\hat{a}, \hat{b}; u, v), \qquad (2.2.5.1.1)$$

where $c = \chi^2_{1,\alpha}$, and where $\hat{a}$ and $\hat{b}$ are expressions for $a$ and $b$ which do not necessarily contain the variable $b$. Different intervals can now be calculated by solving $b$ from (2.2.5.1.1) by choosing different values for $\hat{a}$ and $\hat{b}$.

### 2.2.5.2 Bayesian approach

A Bayesian approach to the estimation of $\hat{a}$ and $\hat{b}$ can be used. Briefly, assuming independence of the prior distributions of the two proportions, i.e. an implicit prior density function proportional to

$$\left(p_1 \cdot p_2 \cdot q_1 \cdot q_2\right)^\alpha \qquad (2.2.5.2.1)$$

as suggested by Jeffreys(1961) and Perks(1947)(see Good(1965) p.18), the posterior mean of $a = p_1 + p_2$ is given by (Beal(1987) p.943):

$$\hat{a}(\alpha) = \frac{n_1}{n_1 + 2(\alpha+1)}\hat{p}_1 + \frac{\alpha+1}{n_1 + 2(\alpha+1)} + \frac{n_2}{n_2 + 2(\alpha+1)}\hat{p}_2 + \frac{\alpha+1}{n_2 + 2(\alpha+1)} \text{ for } \alpha \geq -1.$$

If $\alpha=-1$ then $\hat{a}(-1)=\hat{a}=\hat{p}_1+\hat{p}_2$. If $\alpha$ increases, the prior for the difference puts less weight on the extremes of the unit square(i.e. the points $(-1,-1)$, $(-1,1)$, $(1,-1)$ and $(1,1)$). Then $\hat{a}(\alpha)$ tends more towards 1. Each value of $\alpha$ defines a distinct confidence interval $\mathrm{CONF}(p_1-p_2)$ for $p_1-p_2$. Let us denote this interval briefly here as $I(\alpha)$. Now $I(\alpha)$ has confidence levels larger than those of $I(-1)$. Several values of $\alpha$ were tried by Beal(1987) but $\alpha=-1/2$ seems to be a good overall choice.

The interval $I(-1/2)$ will now be referred to as the Jeffreys-Perks interval since the prior(2.2.5.2.1) with $\alpha=-1/2$ arises naturally from the invariance theories of Jeffreys and Perks(Good(1965) pp. 18, 28, 29). We discuss this interval in section 2.2.5.5. The interval $I(-1)$ will be referred to as the Haldane interval, since the prior $\alpha=-1$ is the result of Haldane priors on $p_1$ and $p_2$ (Haldane(1945) p.223). This interval is discussed in section 2.2.5.6.

### 2.2.5.3 The conventional approximate interval

By letting $\tilde{a}=\hat{a}$ and $\tilde{b}=\hat{b}$, $\tilde{a}$ and $\tilde{b}$ denote hypothetical values of $a$ and $b$, and $\mathrm{V}(\hat{a},\hat{b};u,v)=\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$, equation (2.2.5.1.1) is then

$$\left(b-\hat{b}\right)^2=\chi^2_{1-\alpha}\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)=k^2(\text{say})$$

which is a quadratic equation in $b$ with two roots $b_l=\hat{b}-k$ and $b_u=\hat{b}+k$, where

$$k = z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\,.$$

Let $L = \max(-1, b_l)$ and $U = \min(b_u, 1)$, then the confidence interval is given by

$$\mathrm{CONF}(b)_{1-\alpha} = \mathrm{CONF}(p_1 - p_2)_{1-\alpha} = [\,L\,, U\,]$$

$$= b \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \qquad (2.2.5.3.1)$$

which is the conventional interval as derived in (2.2.2). Thus the unified method for constructing confidence intervals for $p_1 - p_2$ yields the conventional approximate interval when $\tilde{a}$ and $\tilde{b}$ are chosen as $\tilde{a} = \hat{a}$ and $\tilde{b} = \hat{b}$.

### 2.2.5.4 The Anbar(1983) interval

This interval is constructed like the conventional approximate interval except that $\tilde{a}$ and $\tilde{b}$ are chosen as $\tilde{a} = 2\hat{p}_1 - b$ and $\tilde{b} = b$, which after eliminating $\tilde{a}$ gives a quadratic expression in $b$, namely:

$$\left(b-\hat{b}\right)^2 = c\left(u\left[(2-2\hat{p}_1+b)(2\hat{p}_1-b)-b^2\right]+2v(1-2\hat{p}_1+b)b\right)$$

$$= c\left(u\left(4\hat{p}_1-2b-4\hat{p}_1^2+2\hat{p}_1b+2\hat{p}_1b-2b^2\right)+2v\left(b-2\hat{p}_1b+b^2\right)\right)$$

$$\therefore\ b^2-2b\hat{b}+\hat{b}^2 = b^2(2v-2u)c+b(2v-2u+4\hat{p}_1u-4\hat{p}_1v)c+\left(4\hat{p}_1-4\hat{p}_1^2\right)uc$$

$$\therefore\ b^2(1-2vc+2uc)+2b\left(-\hat{b}-vc+uc-2\hat{p}_1uc+2\hat{p}_1vc\right)+\left(\hat{b}^2-4\hat{p}_1uc+4\hat{p}_1^2uc\right)=0\,.$$

This equation is of the form $Ab^2 + Bb + C = 0$, from which we can solve for $b$ yielding:

$$b_l = \frac{B - \sqrt{B^2 - 4AC}}{2A} \text{ and } b_u = \frac{B + \sqrt{B^2 - 4AC}}{2A}.$$

Here $c = z_{\alpha/2}$; $A = 1 - 2vc + 2uc$; $B = -\hat{b} - vc + uc - 2\hat{p}_1 uc + 2\hat{p}_1 vc$ and $C = \hat{b}^2 - 4\hat{p}_1 uc + 4\hat{p}_1^2 uc$.

Let $L = \max(-1, b_l)$ and $U = \min(b_u, 1)$ then

$$\mathrm{CONF}(b) = \mathrm{CONF}(p_1 - p_2)_{1-\alpha} = [\ L\ ;\ U\ ]$$

where $b$ is calculated from, (2.2.5.1.1).

Mee(1984) noted that we can use equation (2.2.5.1.1) also with $\tilde{a} = 2\hat{p}_2 + b$, $\tilde{b} = b$ and that the results obtained from these two methods differ. To solve this problem Mee suggested a more theoretically satisfying interval which we will discuss in section 2.2.5.7. The Anbar(1983) interval is only discussed and not used in the simulation study.

### 2.2.5.5 The Jeffreys-Perks(1987) interval

As discussed in section 2.2.5.2 the interval $I(\alpha)$ with $\alpha = -1/2$ is known as the Jeffreys-Perks interval. This interval is calculated as follows:

Set $\tilde{a} = \hat{a}$ and $\tilde{b} = b$ in equation (2.2.5.1.1) and solve for $b$.

This gives:

$$\left(b - \hat{b}\right)^2 = c\left(u\left[(2 - \hat{a})\hat{a} - b^2\right] + 2v(1 - \hat{a})b\right)$$

$$\therefore b^2 - 2\hat{b}b + \hat{b}^2 = -cub^2 + cu(2 - \hat{a})\hat{a} + 2cv(1 - \hat{a})b$$

$$\therefore b^2(1+cu)+b\left(-2\hat{b}-2vc(1-\hat{a})\right)+\left(\hat{b}^2-cu(2-\hat{a})\hat{a}\right)=0\,.$$

This equation has roots:

$$b=\frac{2\hat{b}+2cv(1-\hat{a})\pm\sqrt{4\left(\hat{b}+cv(1-\hat{a})\right)^2-4(1+cu)\left(\hat{b}^2-cu(2-\hat{a})\hat{a}\right)}}{2(1+cu)}$$

$$=\frac{\hat{b}+cv(1-\hat{a})\pm\sqrt{\left(\hat{b}+cv(1-\hat{a})\right)^2-(1+cu)\left(\hat{b}^2-cu(2-\hat{a})\hat{a}\right)}}{(1+cu)}$$

$$b-\frac{\hat{b}+cv(1-\hat{a})}{1+cu}=\pm\frac{\sqrt{\left(\hat{b}+cv(1-\hat{a})\right)^2-(1+cu)\left(\hat{b}^2-cu(2-\hat{a})\hat{a}\right)}}{1+cu}$$

$$=\pm\frac{\sqrt{c\left\{\mathrm{V}\left(\hat{a},\hat{b};u,v\right)+cu^2(2-\hat{a})\hat{a}+cv^2(1-\hat{a})^2\right\}}}{1+cu}$$

$$=\pm z_{\alpha/2}\frac{\sqrt{V\left(\hat{a},\hat{b};u,v\right)+cu^2(2-\hat{a})\hat{a}+cv^2(1-\hat{a})^2}}{1+cu}$$

where $c=z_{\alpha/2}^2$.

Thus

$\mathrm{CONF}(p_1-p_2)_{1-\alpha}=$

$$b-\frac{\hat{b}+cv(1-\hat{a})}{1+cu}\pm z_{\alpha/2}\frac{\sqrt{V\left(\hat{a},\hat{b};u,v\right)+cu^2(2-\hat{a})\hat{a}+cv^2(1-\hat{a})^2}}{1+cu} \qquad (2.2.5.5.1)$$

If $n_1=n_2=n$, this interval simplifies to

$$\mathrm{CONF}(p_1-p_2)_{1-\alpha}=\frac{\hat{b}}{1+d^2}\pm d\frac{\sqrt{(2-\hat{a})\hat{a}\left(1+d^2\right)-\hat{b}^2}}{1+d^2} \qquad (2.2.5.5.2)$$

where $d=c/\sqrt{2n}$.

Beal(1987) noted that this interval is a definite improvement on the conventional interval. There are however, some values of $p_1$ and $p_2$ for which this interval is too narrow, while the conventional approximate interval is not too narrow for these choices of $p_1$ and $p_2$.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the confidence interval is given by $\text{CONF}(p_1 - p_2)_{0.95} = [0.053\ ;\ 0.3355]$.

### 2.2.5.6 The Haldane(1987) interval

The Haldane interval is calculated in the same manner as the Jeffreys-Perks interval but with $\tilde{a} = \frac{x_1}{n_1} + \frac{x_2}{n_2}$.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the confidence interval is given by $\text{CONF}(p_1 - p_2)_{0.95} = [0.053\ ;\ 0.3377]$.

The Jeffreys-Perks and Haldane methods are not easy to calculate with a handcalculator, though this is still possible. However, the methods are easier implemented through a computer program. Newcombe(1998:b) noted that Jeffreys-Perks' and Haldane's intervals as described in Beal(1987) are prone to certain novel anomalies such as overt overshoot which occurs when the upper confidence limit is larger than 1 or the lower confidence limit is less than $-1$.

## 2.2.5.7 The Mee(1984) interval

Mee suggested that equation (2.2.5.1.1) be solved with $\hat{b}$ chosen as $\hat{b}=b$ and $\hat{a}=a^{*}(b)$ or $\hat{a}=a^{*}$, the maximum likelihood estimator of $a=p_1+p_2$ when $b=p_1-p_2$ has a given value. Since $\hat{a}=a^{*}$ is now a function of $b$, equation (2.2.5.1.1) with these values is not quadratic and cannot be solved in closed form like in cases 2.2.4.1 and 2.2.4.3. Instead, equation (2.2.5.1.1) has to be solved numerically to obtain the two roots $b_l$ and $b_u$. Since $a^{*}$ is defined only in the interval [ 0 ; 2 ] we seek the smallest value where $b_l \in [\,-1\,;\,1\,]$ and the largest value $b_u \in [\,-1\,;\,1\,]$ such that $\left(b-\hat{b}\right)^2 \le c\mathrm{V}\left(a^{*},\hat{b};u,v\right)$ for $b_l \le b \le b_u$. Then $L=b_l$ and $U=b_u$.

The following iterative method can be used to calculate the upper limit $b_u$ of the Mee interval:

(1) Use the Jeffreys-Perks method to determine the upper limit $b_u$. Let $b_1=b_u$ as a first estimate of the upper limit.

(2) Determine the maximum likelihood estimate of $p_1$, namely $p_1^{*}=p_1^{*}(b_1)$, for given $b_1$. This $p_1^{*}$ is the unique solution in the interval $(b_1,1)$ of the maximum likelihood equation:

$$Nx^3+Kx^2+Lx+M=0, \qquad (2.2.5.7.1)$$

where

$$N=n_1+n_2,$$

$$K=(n_1+2n_2)b_1-N-\hat{p}_1n_1-\hat{p}_2n_2,$$

$$L=(n_2b_1-N-2\hat{p}_2n_2)b_1+(\hat{p}_1n_1+\hat{p}_2n_2),$$

$$M = \hat{p}_2 n_2 b_1 (1 - b_1).$$

Now calculate in the following order:

$$g = \frac{K^3}{(3N)^3} - \frac{KL}{6N^2} + \frac{M}{2N},$$

$$h = \operatorname{sgn}(g) \sqrt{\frac{K^2}{(3N)^2} - \frac{L}{3N}},$$

and

$$f = \frac{1}{3}\left(\pi + \cos^{-1}\left(\frac{g}{h^3}\right)\right).$$

Then the solution of the maximum likelihood equation (2.2.5.7.1) is:

$$p_1^* = 2h\cos(f) - \frac{K}{3N}.$$

(3) As with Anbar's(1983) interval let

$$a^* = \max(2p_1^* - b_1; 1).$$

Then let, $\tilde{b} = b_1$ and $\tilde{a} = a^*$ in equation (2.2.5.1.1) to get a Jeffreys-Perks type interval with limits $(b_l, b_u)$. Let $b_2 = b_u$.

(4) Is $\frac{|b_2 - b_1|}{|b_1|} \leq 10^{-4}$?; if not set $b_1 = b_2$ and repeat steps (2) to (4); if yes then stop.

Note: If $b_1 = 1$ then $p_1^* = 1$ and step 2 is then skipped.

The lower limit $b_l$ of the Mee interval is determined likewise as in steps (1) to (4) by starting with $b_1 = b_l$, the lower limit of the Jeffreys-Perks interval. If

$b_1 = -1$ then $p_1^* = 0$ and step 2 is skipped. Santner and Yamagami(1993) note that the iteratively computed Mee intervals are not conservative. This interval method is computer intensive due to the iterative procedure involved.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% confidence interval is given by $\text{CONF}(p_1 - p_2)_{0.95} = [0.0533\ ;\ 0.3377]$.

### 2.2.5.8 The Miettinen and Nurminen(1985) interval

The interval of Miettinen and Nurminen(1985) is calculated like the Mee(1984) interval except that the constant $c$ in equation (2.2.5.1.1) is chosen as $cN/(N-1)$ so that equation (2.2.5.1.1) becomes:

$$\left(b - \hat{b}\right) = cN/(N-1)\text{Var}\left(\hat{a}, \hat{b}; u, v\right),$$

where $N = n_1 + n_2$.

Like the Mee interval method this method requires intensive computation.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% confidence interval is given by $\text{CONF}(p_1 - p_2)_{0.95} = [0.0528\ ;\ 0.3382]$.

### 2.2.5.9 The Wallenstein(1997) interval method

Since the conventional approximate interval is often too short, methods have been suggested to produce more appropriate, that is, wider confidence intervals.

The conventional Asymptotic interval:

$$\begin{aligned}\mathrm{CONF}(p_1 - p_2)_{1-\alpha} &= \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\hat{V}(p_1, p_2)} \\ &= [d' \; ; \; d],\end{aligned}$$

where

$$\hat{V}(p_1, p_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \quad .$$

The lower confidence limit is $d'(x_1, x_2)$ and the upper confidence limit is $d(x_1, x_2)$ or simply $[d' \; ; \; d]$. Let $b = p_1 - p_2$. Suppose $b = d$ the UCL, then $d$ is the solution of the implicit equation:

$$d = \hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\hat{\mathrm{V}}(p_1, p_2)} = [\text{ L ; U }]. \qquad (2.2.5.9.1)$$

Denote $\hat{p}_{i|d}$ as the estimate of $p_i$ based on assuming $b = d$. Beal(1987) sets $\hat{p}_{1|d} + \hat{p}_{2|d} = \hat{p}_1 + \hat{p}_2$, where $p_i = x_i / n_i$ and then solves (2.2.5.9.1) explicitly. Mee(1984) solves (2.2.5.9.1) by using a doubly iterative procedure. The first step fixes $b = d$ and finds $\hat{p}_{1|d}$ and $\hat{p}_{2|d}$ that maximise the likelihood, while the second step updates $d$ to satisfy (2.2.5.9.1) more closely. Miettinen and Nurminen(1985) pointed out that we need not perform iteration to find the MLE's, since they are solutions of a cubic and thus we only need a single iterative procedure. Wallenstein(1997) proposes a slight modification of Mee's procedure by using least squares estimates(LSE's) instead of MLE's for

$\hat{p}_{1|d}$ and $\hat{p}_{2|d}$, and allows for a continuity correction $\varepsilon \geq 0$. Wallenstein proposed replacing $\hat{p}_1 - \hat{p}_2$ in (2.2.5.9.1) by $\hat{p}_1 - \hat{p}_2 + \varepsilon$ to get

$$d = \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right) + z_{\alpha/2}\sqrt{V\left(\hat{p}_{1|d}, \hat{p}_{2|d}\right)} \tag{2.2.5.9.2}$$

The LSE's of $p_i$ subject to the constraint $\hat{p}_{1|d} - \hat{p}_{2|d} = d$ are

$$\hat{p}_{1|d} = \widetilde{p} + d\frac{n_2}{N}$$

and

$$\hat{p}_{2|d} = \widetilde{p} - d\frac{n_1}{N} \tag{2.2.5.9.3}$$

where $\widetilde{p} = n_1\hat{p}_1 + n_2\hat{p}_2 / N$ and $N = n_1 + n_2$.

Now substituting (2.2.5.9.3) into (2.2.5.9.2) we obtain:

$$\frac{\left[d - \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right)\right]^2}{z^2_{\alpha/2}} = \frac{\hat{p}_{1|d}\left(1 - \hat{p}_{1|d}\right)}{n_1} + \frac{\hat{p}_{2|d}\left(1 - \hat{p}_{2|d}\right)}{n_2}$$

$$= \frac{\left(\widetilde{p} + \frac{dn_2}{N}\right)\left(1 - \widetilde{p} - \frac{dn_2}{N}\right)}{n_1} + \frac{\left(\widetilde{p} - \frac{dn_1}{N}\right)\left(1 - \widetilde{p} + \frac{dn_1}{N}\right)}{n_2}$$

$$= \frac{n_2\left(\widetilde{p} + \frac{dn_2}{N}\right)\left(1 - \widetilde{p} - \frac{dn_2}{N}\right) + n_1\left(\widetilde{p} - \frac{dn_1}{N}\right)\left(1 - \widetilde{p} + \frac{dn_1}{N}\right)}{n_1 n_2}$$

This is equivalent to

$$\frac{n_1 n_2}{z^2_{\alpha/2}}\left[d - \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right)\right]^2 = n_2\left(\widetilde{p} + \frac{dn_2}{N}\right)\left(1 - \widetilde{p} - \frac{dn_2}{N}\right) + n_1\left(\widetilde{p} - \frac{dn_1}{N}\right)\left(1 - \widetilde{p} + \frac{dn_1}{N}\right)$$

This is a quadratic equation in $d$ of the form:

$$Ad^2 + Bd + C = 0, \tag{2.2.5.9.4}$$

where

$$A = 1 + N^{-1} z_{\alpha/2}^2 \left(1 + \frac{(n_1 - n_2)^2}{(n_1 n_2)}\right),$$

$$B = -2(\hat{p}_1 - \hat{p}_2 + \varepsilon) + \frac{z_{\alpha/2}^2}{n_1 n_2}(1 - 2\tilde{p})(n_1 - n_2),$$

$$C = (\hat{p}_1 - \hat{p}_2 + \varepsilon)^2 - \frac{z_{\alpha/2}^2 N\tilde{p}(1 - \tilde{p})}{n_1 n_2}. \tag{2.2.5.9.5}$$

We solve for $d$ from (2.2.5.9.4) by choosing the larger solution

$$d = \frac{-B + \sqrt{B^2 - 4AC}}{2A}.$$

If for a chosen continuity correction $\varepsilon$, the term $\hat{p}_1 - \hat{p}_2 + \varepsilon$ exceeds 1, then $\hat{p}_1 - \hat{p}_2 + \varepsilon$ should be replaced in $B$ and $C$ by 1. To compute the lower bound $d'$, we replace $\varepsilon$ by $-\varepsilon$ in $B$ and $C$ in equation (2.2.5.9.5) and solve equation (2.2.5.9.4) by choosing the smaller solution of (2.2.5.9.4) namely:

$$d' = \frac{-B - \sqrt{B^2 - 4AC}}{2A}.$$

If $\varepsilon = 0$ then the upper and lower bounds $d$ and $d'$ are the solutions of a single quadratic equation $Ad^2 + Bd + C = 0$ where $A, B$ and $C$ are given as in (2.2.5.9.5) with $\varepsilon = 0$.

If $n_1 = n_2 = m$ then the preliminary confidence interval on $b$, $(d', d)$ is

$$\frac{(\hat{p}_1 - \hat{p}_2) \pm \left(z_{\alpha/2} / \sqrt{m}\right)\left\{\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2) + z_{\alpha/2}^2 \tilde{p}(1 - \tilde{p}) / m\right\}^{1/2}}{1 + z_{\alpha/2}^2 / 2m} \tag{2.2.5.9.6}$$

in agreement with Beal(1987)(see section 2.2.4.3).

The procedure described above is valid only if the solution $d$ of the quadratic equation (2.2.5.9.4) when substituted into (2.2.5.9.3) yields estimates $\hat{p}_{1|d}$ and $\hat{p}_{2|d}$ in the interval $[\,0\,;\,1\,]$. Similarly when the solution $d'$ from (2.2.5.9.4) is substituted in (2.2.5.9.3) we should get estimates $\hat{p}_{1|d'}$ and $\hat{p}_{2|d'}$ in the interval $[\,0\,;\,1\,]$. If this constraint is violated we replace $p_{i|d}$ outside $[\,0\,;\,1\,]$ with 1 or 0. For example if $\hat{p}_1 \geq \hat{p}_2$ and $p_{2|d} < 0$ force $\hat{p}_{2|d} = 0$ so $\hat{p}_{1|d} = d$ and (2.2.5.9.2) becomes:

$$V\left(\hat{p}_{1|d}, \hat{p}_{2|d}\right) = \frac{\hat{p}_{1|d}\left(1-\hat{p}_{1|d}\right)}{n_1} + \frac{\hat{p}_{2|d}\left(1-\hat{p}_{2|d}\right)}{n_2}$$

$$= \frac{d(1-d)}{n_1} + 0$$

So equation (2.2.5.9.2) becomes

$$d = \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right) = z_{\alpha/2}\sqrt{\frac{d(1-d)}{n_1}}$$

$$= \frac{z_{\alpha/2}}{\sqrt{n_1}}\sqrt{d(1-d)}\,.$$

Square both sides:

$$\left[d - \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right)\right]^2 = \frac{z_{\alpha/2}^2}{n_1}\left(d - d^2\right)$$

$$d^2 - 2d\left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right) + \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right)^2 - d\frac{z_{\alpha/2}^2}{n_1} + d^2\frac{z_{\alpha/2}^2}{n_1} = 0\,.$$

i.e. $$d^2\left[1 + \frac{z_{\alpha/2}^2}{n_1}\right] - d\left[2\left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right) + \frac{z_{\alpha/2}^2}{n_1}\right] + \left(\hat{p}_1 - \hat{p}_2 + \varepsilon\right)^2 = 0\,.$$

The larger solution for $d$ is

$$d=\frac{2\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)+\frac{z_{\alpha/2}^2}{n_1}+\sqrt{4\left[\hat{p}_1-\hat{p}_2+\varepsilon+\frac{z_{\alpha/2}^2}{2n_1}\right]^2-4\left[1+\frac{z_{\alpha/2}^2}{n_1}\right]\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)^2}}{2\left(1+z_{\alpha/2}^2/n_1\right)}$$

$$=\frac{\hat{p}_1-\hat{p}_2+\varepsilon+z_{\alpha/2}^2/2n_1+\left(z_{\alpha/2}/\sqrt{n_1}\right)\left\{\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)\left[1-\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)\right]+z_{\alpha/2}^2/4n_1\right\}^{1/2}}{1+z_{\alpha/2}^2/n_1}$$

If $\hat{p}_1 \geq \hat{p}_2$ and $\hat{p}_{1|d} > 1$ set $\hat{p}_{1|d} = 1$ and $\hat{p}_{2|d} = 1-d$ then:

$$d=\frac{\hat{p}_1-\hat{p}_2+\varepsilon+\frac{z_{\alpha/2}^2}{2n_2}+\frac{z_{\alpha/2}}{\sqrt{n_2}}\left\{\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)\left[1-\left(\hat{p}_1-\hat{p}_2+\varepsilon\right)\right]+\frac{z_{\alpha/2}^2}{4n_2}\right\}^{1/2}}{1+z_{\alpha/2}^2/n_2}. \quad (2.2.5.9.7)$$

If $\hat{p}_2 > \hat{p}_1$ the simplest thing to do is to switch indices, i.e. we write $\hat{p}_1$ as $\hat{p}_2$ and $\hat{p}_2$ as $\hat{p}_1$. Then we solve $d$ again as in (2.2.5.9.6) and (2.2.5.9.7).

If $\hat{p}_1 \geq \hat{p}_2$ and $\hat{p}_{1|d'} < 0$ then $d' = \hat{p}_{1|d'} - \hat{p}_{2|d'} = 0 - \hat{p}_{2|d'} < 0$. Then we set $\hat{p}_{1|d'} = 0$ and $\hat{p}_{2|d'} = -d'$

$$\mathrm{Var}\left(\hat{p}_{1|d'}, \hat{p}_{2|d'}\right)=\frac{0(1-0)}{n_1}+\frac{-d'(1+d')}{n_2}.$$

So equation (2.2.5.9.2) becomes:

$$d-\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)=z_{\alpha/2}\sqrt{\frac{-d'(1+d')}{n_2}}$$

so:

$$\left(d'-\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)\right)^2=-\frac{z_{\alpha/2}^2}{n_2}d'(1+d')$$

$$d'^2-2d'\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)+\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)^2=-\frac{z_{\alpha/2}^2}{n_2}d'-\frac{z_{\alpha/2}^2}{n_2}d'^2$$

$$d'^2\left[1+\frac{z_{\alpha/2}^2}{n_2}\right]-d'\left[2\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)^2+\frac{z_{\alpha/2}^2}{n_2}\right]+\left(\hat{p}_1-\hat{p}_2-\varepsilon\right)=0.$$

So $d'$ is the smaller root namely:

$$d' = \frac{2(\hat{p}_1 - \hat{p}_2 - \varepsilon) + \dfrac{z_{\alpha/2}^2}{n_2} - \sqrt{\left[2(\hat{p}_1 - \hat{p}_2 - \varepsilon) + \dfrac{z_{\alpha/2}^2}{n_2}\right]^2 - 4\left[1 + \dfrac{z_{\alpha/2}^2}{n_2}\right][\hat{p}_1 - \hat{p}_2 - \varepsilon]^2}}{2\left(1 + z_{\alpha/2}^2 / n_2\right)}$$

$$= \frac{\hat{p}_1 - \hat{p}_2 - \varepsilon + \dfrac{z_{\alpha/2}^2}{2n_2} - \sqrt{\left[(\hat{p}_1 - \hat{p}_2 - \varepsilon) + \dfrac{z_{\alpha/2}^2}{2n_2}\right]^2 - \left[1 + \dfrac{z_{\alpha/2}}{n_2}\right][\hat{p}_1 - \hat{p}_2 - \varepsilon]}}{1 + z_{\alpha/2}^2 / n_2}$$

$$= \frac{\hat{p}_1 - \hat{p}_2 - \varepsilon + \dfrac{z_{\alpha/2}^2}{2n_2} - \sqrt{(\hat{p}_1 - \hat{p}_2 - \varepsilon)^2 + \dfrac{z_{\alpha/2}^2}{n_2}(\hat{p}_1 - \hat{p}_2 - \varepsilon) + \dfrac{z_{\alpha/2}^4}{4n_2^2} - \left[1 + \dfrac{z_{\alpha/2}^2}{n_2}\right](\hat{p}_1 - \hat{p}_2 - \varepsilon)}}{1 + z_{\alpha/2}^2 / n_2}$$

$$= \frac{\hat{p}_1 - \hat{p}_2 - \varepsilon + \dfrac{z_{\alpha/2}^2}{2n_2} - \sqrt{\dfrac{z_{\alpha/2}^2}{n_2}(\hat{p}_1 - \hat{p}_2 - \varepsilon)(1 - (\hat{p}_1 - \hat{p}_2 - \varepsilon)) + \dfrac{z_{\alpha/2}^4}{4n_2}}}{1 + z_{\alpha/2}^2 / n_2}$$

$$= \frac{\hat{p}_1 - \hat{p}_2 - \varepsilon + \dfrac{z_{\alpha/2}^2}{2n_2} - \dfrac{z_{\alpha/2}}{\sqrt{n_2}}\sqrt{(\hat{p}_1 - \hat{p}_2 - \varepsilon)(1 - (\hat{p}_1 - \hat{p}_2 - \varepsilon)) + \dfrac{z_{\alpha/2}^2}{4n_2}}}{1 + z_{\alpha/2}^2 / n_2}.$$

If $\hat{p}_1 < \hat{p}_2$ we again reverse indices. The case when $x_1 = 0$ and $x_2 = 0$ was not described in detail in Wallenstein's paper. Wallenstein, in personal correspondence, suggested the use of an ad hoc solution to this case by invoking the equation (2.2.5.9.6) and the resulting upper confidence limit is then

$$d = \frac{z^2 / 2n_1 + z / \sqrt{n_1}\left(\sqrt{z^2 / 4n_1}\right)}{1 + z^2 / n_1},$$

where $z$ denotes $z_{\alpha/2}$. The lower confidence limit is calculated as the negative of the upper bound.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% Wallenstein confidence interval without continuity correction is given by $\mathrm{CONF}(p_1 - p_2)_{0.95}$ = $[0.0528\ ;\ 0.3344]$. The 95% Wallenstein confidence interval with correction is given by $\mathrm{CONF}(p_1 - p_2)_{0.95} = [0.0392\ ;\ 0.3469]$.

### 2.2.6 Intervals based on the Wilson(1927) score interval

The following score intervals for $p_1 - p_2$ are based on Wilson's score interval for the single proportion, and were proposed by Newcombe(1998: b). We will first discuss the score interval method for the single proportion with and without continuity correction, and then the score interval method for the difference $p_1 - p_2$ with and without continuity correction.

The efficient score of a parameter $\theta$ in a distribution of a random variable $Y$ with density function parametrized by $\theta$, i.e. $f_Y(y,\theta)$, is defined by Cox and Hinkley(1974, p.107) as:

$$U(\theta) = \frac{\partial \log f_Y(y,\theta)}{\partial \theta}.$$

There is a close connection between the position of the maximum likelihood estimate(MLE) $\hat{\theta}$ and the efficient score namely: $U(\hat{\theta}) = 0$. The variance of the score $U(\theta)$ is given by the Fisher information:

$$Var(U(\theta)) = E(U^2) = i(\theta)$$

where

$$i(\theta) = E\left(-\frac{\partial^2 \log f_Y y, \theta}{\partial^2 \theta}\right).$$

For large sample sizes, the test statistic $W_U = \dfrac{U^2(\theta)}{i(\theta)}$ can be used to test the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ (Cox and Hinkley(1974, p.339) yielding a critical region

$$|z| = \frac{|U(\theta_0)|}{\sqrt{i(\theta_0)}} > z_{\alpha/2}.$$

Cox and Hinkley(1974, p.343) state that: "Confidence regions based directly on the efficient score, i.e. using the $W_U$ test statistic, have the advantage that, at least when there are no nuisance parameters, moments of the test statistic are obtained directly. Thus, for a one-dimensional parameter $\theta$, confidence regions can be obtained from the inequality

$$\frac{|U(\theta)|}{\sqrt{i(\theta)}} \leq z_{\alpha/2}.$$

These will usually be intervals. The procedure is invariant under transformation of the parameter. The test statistic has exactly mean zero and unit variance and corrections based on its higher moments are easily introduced."

#### 2.2.6.1 The score interval for a single proportion without continuity correction

When $p$ is the underlying proportion, the sample proportion $\hat{p}$ is approximately normally distributed with mean $p$ and standard error $\sqrt{pq/n}$.

In order to test the hypothesis that $p$ is equal to a prespecified $p_0$ against the alternative hypothesis that $p \neq p_0$, one may calculate the ratio

$$|z| = \frac{|p - p_0|}{\sqrt{\frac{p_0 q_0}{n}}},$$

where $q_0 = 1 - p_0$, and reject the hypothesis if $z$ exceeds the critical value of the normal curve for the desired two-tailed significance level. The test is based on the ratio and if $z_{\alpha/2}$ denotes the value cutting off the area $\alpha / 2$ in the upper tail of the standard normal distribution, an approximate $100(1-\alpha)\%$ confidence interval consists of all those values of $p$ satisfying

$$\frac{|p - \hat{p}|}{\sqrt{\hat{p}\hat{q} / n}} \leq z_{\alpha/2}.$$

Note that for confidence intervals $p_0$ is replaced by $\hat{p}$ since in confidence intervals there is no hypothesised value of $p$. Thus the score interval without correction is:

$$\text{CONF}(p)_{1-\alpha} = [\ L\ ;\ U\ ] = \frac{\left(2n\hat{p} + z_{\alpha/2}^2 \pm z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4n\hat{p}\hat{q}}\right)}{2\left(n + z_{\alpha/2}^2\right)}. \qquad (2.2.6.1.1)$$

### 2.2.6.2 The score interval for a single proportion with continuity correction

The score interval method for the single proportion with continuity correction is the same as the score interval method without continuity correction except the quantity $1/n$ subtracted in the numerator is a correction for continuity. The score interval with continuity correction is calculated as follows:

$$L = \frac{2n\hat{p} + z_{\alpha/2}^2 - 1 - z_{\alpha/2}\sqrt{z_{\alpha/2}^2 - 2 - 1/n + 4\hat{p}(n\hat{q} + 1)}}{2\left(n + z_{\alpha/2}^2\right)},$$

$$U = \frac{2n\hat{p} + z_{\alpha/2}^2 + 1 + z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 2 - 1/n + 4\hat{p}(n\hat{q} - 1)}}{2\left(n + z_{\alpha/2}^2\right)},$$

which yields the confidence interval

$$\text{CONF}(p)_{1-\alpha} = [\ L\ ;\ U\ ]. \qquad (2.2.6.2.1)$$

According to Fleiss(1981) the continuity correction should be applied only when it is numerically smaller than $|p - p_0|$. The score interval method either with or without continuity correction, is preferred to the conventional approximate interval when $p$ is near zero or one. Newcombe(1998:a) noted that if $p = 0$ then $L$ must be taken as 0 and if $p = 1$, then $U$ is taken as 1. Newcombe also noted that these interval methods result in a good degree and symmetry of coverage as well as avoidance of aberrations.

## 2.2.6.3 The score interval without continuity correction for the difference between two proportions

This method combines Wilson's score intervals for each single proportion in much the same way as the conventional approximate method combines simple intervals. To obtain the Wilson score interval, first calculate:

$\hat{b} = \hat{p}_1 - \hat{p}_2$ and then

$$L_i = \left(\left(2n_i\hat{p}_i + z^2_{\alpha/2}\right) - z_{\alpha/2}\sqrt{z^2_{\alpha/2} + 4n_i\hat{p}_i\hat{q}_i}\right) / 2\left(n_i + z^2_{\alpha/2}\right),$$

and $$U_i = \left(\left(2n_i\hat{p}_i + z^2_{\alpha/2}\right) + z_{\alpha/2}\sqrt{z^2_{\alpha/2} + 4n_i\hat{p}_i\hat{q}_i}\right) / 2\left(n_i + z^2_{\alpha/2}\right) \text{ for } i = 1,2.$$

Substitute $L_i$ and $U_i$ in $$\varepsilon_1 = z_{\alpha/2}\sqrt{\frac{L_1(1-L_1)}{n_1} + \frac{U_2(1-U_2)}{n_2}}$$

and $$\varepsilon_2 = z_{\alpha/2}\sqrt{\frac{U_1(1-U_1)}{n_1} + \frac{L_2(1-L_2)}{n_2}}.$$

Now let L=$\hat{b} - \varepsilon_1$ and U=$\hat{b} + \varepsilon_2$, then the $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$ is given by

$$\text{CONF}(p_1 - p_2)_{1-\alpha} = [\ L\ ;\ U\ ]. \tag{2.2.6.3.1}$$

This interval method is not so easy to calculate with a calculator though it can easily be implemented with a computer program.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% confidence interval is given by $\text{CONF}(p_1 - p_2)_{0.95}$ = [ 0.0524 ; 0.3339 ].

## 2.2.6.4 The score interval with continuity correction for the difference between two proportions

This interval is calculated similarly to the score interval without continuity correction. Compute the lower and upper limit as follows:

Firstly compute $\hat{b} = \hat{p}_1 - \hat{p}_2$ and then also compute

$$L_i = 2n_i\hat{p}_i + z_{\alpha/2}^2 - 1 - z_{\alpha/2}\sqrt{z_{\alpha/2}^2 - 2 - \frac{1}{n_i} + 4\hat{p}_i\,\frac{n_i\hat{q}_i + 1}{2\left(n_i + z_{\alpha/2}^2\right)}}$$

and

$$U_i = 2n_i\hat{p}_i + z_{\alpha/2}^2 + 1 + z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 2 - \frac{1}{n_i} + 4\hat{p}_i\,\frac{n_i\hat{q}_i - 1}{2\left(n_i + z_{\alpha/2}^2\right)}} \qquad \text{for } i = 1,2.$$

Substitute these values into:

$$\varepsilon_1 = \sqrt{\left(\frac{x_1}{n_1} - L_1\right)^2 + \left(U_2 - \frac{x_2}{n_2}\right)^2}$$

and

$$\varepsilon_2 = \sqrt{\left(U_1 - \frac{x_1}{n_1}\right)^2 + \left(\frac{x_2}{n_2} - L_2\right)^2}\,.$$

Let L=$\hat{b} - \varepsilon_1$ and U=$\hat{b} + \varepsilon_2$ which yield the confidence interval

$$\text{CONF}(p_1 - p_2)_{1-\alpha} = [\ L\ ;\ U\ ]. \qquad (2.2.6.4.1)$$

This interval method can easily be implemented with a computer program.

Example: If $x_1 = 56$, $x_2 = 48$, $n_1 = 70$ and $n_2 = 80$, then the 95% confidence interval is [ 0.0428 ; 0.3422 ].

# CHAPTER 3
# COMPARISON OF THE DIFFERENT INTERVAL METHODS

## 3.1 Introduction

The various interval methods were compared by a simulation study. The simulation study was done using the following algorithm:

1. Specify the parameter combination of proportions $p_1$ and $p_2$ and sample sizes $n_1$ and $n_2$ to be investigated.

2. Simulated pairs of 50 000 samples $x_1$ and $x_2$ given the specified parameter combinations of $p_1$, $p_2$, $n_1$ and $n_2$ $\left(\text{where } x_i \sim \text{Bin}(n_i, p_i) \text{ for } i = 1,2\right)$.

3. Calculate confidence intervals for $p_1 - p_2$ for each of the 50 000 simulated samples using the different interval methods.

4. Determine the interval length, the occurrence of coverage, violation of the definition interval and zero width intervals of the different interval methods for each of the 50 000 samples.

The average length was determined as the sum of the differences between the upper bound and the lower bound for each simulation divided by the total number of simulations, whereas the coverage, violation of the definition interval and the occurrence of zero width intervals are expressed as a percentage out of 50 000 simulations.

The simulation study was done for each combination of proportions and all sample size combinations, as given in Tables 3.1 and 3.2. The combination of proportions $p_1 = 0.6$ and $p_2 = 0.4$ was used as an example where both proportions are close to 0.5, where conventional interval methods might perform satisfactorily. The combination $p_1 = 0.95$ and $p_2 = 0.85$ was used as an example when both proportions are close to 1, and proportions of this magnitude are typical in clinical trials for cure rates of antibiotics. The combination $p_1 = 0.95$ and $p_2 = 0.15$ was used to examine a situation when the difference between the two proportions is large. Combination $p_1 = 0.98$ and $p_2 = 0.9$ was used to examine a situation where the difference between the two proportions is quite small. Combination $p_1 = 0.98$ and $p_2 = 0.1$ gives the corresponding large difference. Combination $p_1 = 0.98$ and $p_2 = 0.98$ was used to examine a situation where the difference between the two proportions is zero and the proportions are close to 1. Combination $p_1 = 0.98$ and $p_2 = 0.02$ was used when the difference between the two proportions is large, and the individual proportions are close to 0 and 1.

The sample sizes $n_1 = 100$ and $n_2 = 90$ were viewed as large. The sample size combination $n_1 = 60$ and $n_2 = 50$ was viewed as moderately large whereas the combination $n_1 = 30$ and $n_2 = 20$ was viewed as a relatively small sample size. The sample size combination $n_1 = 10$ and $n_2 = 10$ was examined as a small but sometimes imposed sample size.

### Table 3.1 Proportion combinations used in simulation study

| Proportions | |
|---|---|
| $p_1 = 0.6$ | $p_2 = 0.4$ |
| $p_1 = 0.95$ | $p_2 = 0.85$ |
| $p_1 = 0.95$ | $p_2 = 0.15$ |
| $p_1 = 0.98$ | $p_2 = 0.90$ |
| $p_1 = 0.98$ | $p_2 = 0.10$ |
| $p_1 = 0.98$ | $p_2 = 0.98$ |
| $p_1 = 0.98$ | $p_2 = 0.02$ |

### Table 3.2 Sample size combinations used in simulation study

| Sample sizes | |
|---|---|
| $n_1 = 100$ | $n_2 = 90$ |
| $n_1 = 60$ | $n_2 = 50$ |
| $n_1 = 30$ | $n_2 = 20$ |
| $n_1 = 10$ | $n_2 = 10$ |

Table 3.3 gives the observed coverages and the average lengths of the various confidence intervals for $\hat{p}_1 = 0.6$ and $\hat{p}_2 = 0.4$ for different sample sizes. Table 3.4 presents the deviations from the definition intervals and zero width intervals. All values are given as percentages in Table 3.4. Tables 3.5 to 3.16 in pairs provide the same information for the other parameter combinations. The results of these simulated intervals are discussed in detail in section 3.2.

The results for the different interval methods are presented in the tables in sequence of computer intensiveness from the least to the most computer intensive method. The least computer intensive method is the usual Asymptotic interval method with or without continuity correction, and the most computational intensity methods are the iterative methods of Mee, and of Miettinen and Nurminen. The different methods are also discussed in this sequence in section 3.2.

**Table 3.3 Observed coverage and average length of nominal 95% confidence intervals ($p_1 = 0.6$, $p_2 = 0.4$)**

| Sample size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 94.7 | 96.2 | 95.7 | 95.1 | 95.1 | 94.9 | 96.0 | 95.1 | 96.5 | 95.1 | 95.1 |
| | Average length | .278 | .299 | .290 | .275 | .275 | .272 | .287 | .275 | .296 | .275 | .276 |
| 60, 50 | Observed coverage(%) | 94.4 | 96.3 | 95.7 | 94.9 | 94.9 | 94.7 | 96.1 | 94.9 | 96.6 | 94.9 | 94.9 |
| | Average length | .364 | .401 | .388 | .358 | .358 | .353 | .377 | .358 | .394 | .358 | .360 |
| 30, 20 | Observed coverage(%) | 93.7 | 96.6 | 95.9 | 94.7 | 94.7 | 94.3 | 96.6 | 94.7 | 97.6 | 94.7 | 94.7 |
| | Average length | .542 | .625 | .604 | .523 | .523 | .506 | .557 | .521 | .597 | .523 | .528 |
| 10, 10 | Observed coverage(%) | 92.1 | 97.1 | 95.9 | 95.3 | 95.3 | 95.3 | 97.2 | 95.3 | 98.8 | 95.6 | 95.6 |
| | Average length | .813 | 1.01 | .957 | .751 | .755 | .713 | .813 | .759 | .929 | .759 | .776 |

As Asymptotic without correction

AsC Asymptotic with correction

HA Asymptotic with Hauck and Anderson's correction

Hal Haldane interval

JP Jeffreys-Perks interval

Sc Score interval without correction

ScC Score interval with correction

W Wallenstein's interval without correction

WC Wallenstein's interval with correction

Mee Mee's interval

MN Miettinen and Nurminen interval

**Table 3.4 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.6$, $p_2 = 0.4$)**

| Sample Size | Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 1.35 | 5.16 | 4.86 | 0 | 0 | 0 | 0 | 1.24 | 2.79 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.5 Observed coverage and average length of nominal 95% confidence intervals ($p_1 = 0.95$, $p_2 = 0.85$)**

| Sample size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 94.2 | 96.9 | 95.9 | 94.5 | 94.9 | 95.3 | 97.2 | 94.5 | 97.1 | 95.1 | 95.1 |
| | Average length | .169 | .189 | .181 | .167 | .170 | .177 | .191 | .171 | .193 | .176 | .177 |
| 60, 50 | Observed coverage(%) | 92.9 | 97.2 | 95.6 | 94.4 | 95.6 | 96.1 | 97.8 | 94.7 | 97.6 | 94.3 | 94.3 |
| | Average length | .223 | .259 | .245 | .219 | .226 | .239 | .265 | .232 | .271 | .238 | .239 |
| 30, 20 | Observed coverage(%) | 92.8 | 95.6 | 94.7 | 94.7 | 94.7 | 96.7 | 99.1 | 93.9 | 98.6 | 92.3 | 92.3 |
| | Average length | .333 | .416 | .391 | .321 | .345 | .379 | .435 | .378 | .470 | .377 | .382 |
| 10, 10 | Observed coverage(%) | 79.7 | 99.7 | 87.6 | 79.7 | 97.7 | 99.2 | 99.9 | 79.7 | 99.9 | 99.2 | 99.3 |
| | Average length | .451 | .651 | .575 | .416 | .519 | .619 | .737 | .585 | .902 | .627 | .646 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.6 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.95$, $p_2 = 0.85$)**

| Sample size | Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | .86 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 0 | .07 | .01 | 0 | 0 | 0 | 0 | .07 | .58 | 0 | 0 |
| | Zero width interval(%) | 11.7 | 0 | 0 | 11.7 | 0 | 0 | 0 | 11.7 | 0 | 0 | 0 |

As Asymptotic without correction

AsC Asymptotic with correction

HA Asymptotic with Hauck and Anderson's correction

Hal Haldane interval

JP Jeffreys-Perks interval

Sc Score interval without correction

ScC Score interval with correction

W Wallenstein's interval without correction

WC Wallenstein's interval with correction

Mee Mee's interval

MN Miettinen and Nurminen interval

**Table 3.7 Observed coverage and average length of nominal 95% confidence intervals ($p_1 = 0.95$, $p_2 = 0.15$)**

| Sample size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 93.9 | 95.7 | 95.4 | 94.8 | 94.8 | 94.4 | 96.2 | 95.2 | 97.3 | 95.0 | 95.2 |
| | Average length | .169 | .190 | .181 | .170 | .170 | .172 | .186 | .177 | .200 | .169 | .170 |
| 60, 50 | Observed coverage(%) | 93.9 | 95.9 | 95.2 | 95.3 | 95.3 | 95.1 | 96.7 | 95.8 | 98.2 | 95.3 | 95.4 |
| | Average length | .223 | .259 | .245 | .226 | .226 | .229 | .253 | .248 | .288 | .224 | .225 |
| 30, 20 | Observed coverage(%) | 88.4 | 93.7 | 93.5 | 95.6 | 95.6 | 94.2 | 97.1 | 97.0 | 99.0 | 95.6 | 95.6 |
| | Average length | .333 | .417 | .391 | .344 | .344 | .351 | .401 | .416 | .503 | .337 | .341 |
| 10, 10 | Observed coverage(%) | 87.4 | 88.1 | 87.9 | 93.6 | 95.9 | 95.9 | 95.9 | 98.1 | 99.4 | 95.9 | 95.9 |
| | Average length | .451 | .651 | .576 | .511 | .514 | .539 | .641 | .664 | .850 | .501 | .516 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.8 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.95$, $p_2 = 0.15$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1, n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 0 | .01 | 0 | 0 | 0 | 0 | 0 | .05 | .22 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | .46 | 4.01 | 1.53 | 0 | 0 | 0 | 0 | 4.69 | 9.99 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 31.6 | 62.4 | 52.6 | 0 | .87 | 0 | .87 | 41.5 | 53.2 | 0 | 0 |
| | Zero width interval(%) | .87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 67.3 | 95.9 | 93.6 | 0 | 0 | 0 | 11.8 | 51.6 | 62.1 | 0 | 0 |
| | Zero width interval(%) | 11.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.9 Observed coverage and average length of nominal 95% confidence intervals** $(p_1 = 0.98,\ p_2 = 0.9)$

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 93.9 | 96.7 | 96.0 | 94.3 | 94.3 | 96.5 | 97.9 | 95.2 | 98.3 | 95.2 | 95.3 |
| | Average length | .134 | .155 | .145 | .132 | .137 | .148 | .163 | .144 | .168 | .144 | .144 |
| 60, 50 | Observed coverage(%) | 92.8 | 95.9 | 95.2 | 92.8 | 94.7 | 97.1 | 98.8 | 96.4 | 98.2 | 94.2 | 94.2 |
| | Average length | .176 | .213 | .198 | .173 | .184 | .205 | .232 | .199 | .241 | .199 | .200 |
| 30, 20 | Observed coverage(%) | 87.1 | 87.7 | 87.7 | 87.2 | 97.8 | 99.1 | 99.9 | 93.7 | 99.8 | 95.9 | 95.9 |
| | Average length | .258 | .341 | .315 | .251 | .289 | .343 | .402 | .342 | .437 | .337 | .341 |
| 10, 10 | Observed coverage(%) | 70.2 | 99.9 | 71.7 | 71.2 | 99.8 | 99.8 | 99.9 | 71.2 | 99.9 | 99.8 | 99.8 |
| | Average length | .325 | .525 | .442 | .299 | .457 | .597 | .719 | .453 | .889 | .599 | .619 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.10 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.98$, $p_2 = 0.9$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | .17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 6.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 0 | .01 | 0 | 0 | 0 | 0 | 0 | .01 | .13 | 0 | 0 |
| | Zero width interval(%) | 28.2 | 0 | 0 | 28.2 | 0 | 0 | 0 | 28.2 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.11 Observed coverage and average length of nominal 95% confidence intervals** ($p_1 = 0.98$, $p_2 = 0.1$)

| Sample Size |  | Method |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ |  | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 92.1 | 96.5 | 96.0 | 95.5 | 95.5 | 94.2 | 97.2 | 97.4 | 98.9 | 95.9 | 95.9 |
|  | Average length | .134 | .155 | .145 | .137 | .137 | .141 | .155 | .162 | .185 | .136 | .136 |
| 60, 50 | Observed coverage(%) | 90.9 | 95.0 | 94.9 | 95.4 | 95.4 | 95.3 | 97.3 | 97.8 | 99.3 | 95.3 | 95.3 |
|  | Average length | .176 | .213 | .198 | .183 | .184 | .191 | .215 | .231 | .269 | .181 | .182 |
| 30, 20 | Observed coverage(%) | 88.7 | 93.3 | 93.3 | 96.4 | 96.4 | 95.9 | 96.4 | 98.5 | 99.9 | 96.4 | 96.4 |
|  | Average length | .258 | .341 | .314 | .287 | .288 | .305 | .355 | .393 | .468 | .281 | .284 |
| 10, 10 | Observed coverage(%) | 70.8 | 71.3 | 71.2 | 92.5 | 92.5 | 93.5 | 97.4 | 99.5 | 99.9 | 97.4 | 97.4 |
|  | Average length | .322 | .522 | .439 | .447 | .449 | .492 | .594 | .607 | .762 | .435 | .450 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
|  |  | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.12 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.98$, $p_2 = 0.1$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | .35 | 3.12 | 1.19 | 0 | 0 | 0 | 0 | 9.29 | 18.4 | 0 | 0 |
| | Zero width interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 12.3 | 40.6 | 24.8 | 0 | .15 | .15 | .15 | 44.0 | 56.9 | 0 | 0 |
| | Zero width interval(%) | .15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 67.7 | 91.1 | 88.7 | 0 | 6.5 | 0 | 6.5 | 66.8 | 69.4 | 0 | 0 |
| | Zero width interval(%) | 6.5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 63.9 | 99.5 | 98.6 | 0 | 0 | 0 | 28.6 | 53.9 | 57.5 | 0 | 0 |
| | Zero width interval(%) | 28.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.13 Observed coverage and average length of nominal 95% confidence intervals ($p_1 = 0.98$, $p_2 = 0.98$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 96.0 | 99.6 | 98.7 | 96.0 | 98.7 | 98.7 | 99.8 | 97.5 | 99.7 | 97.5 | 97.5 |
| | Average length | .076 | .097 | .088 | .075 | .086 | .107 | .123 | .09 | .117 | .107 | .107 |
| 60, 50 | Observed coverage(%) | 98.3 | 99.9 | 99.6 | 99.2 | 99.6 | 99.4 | 99.9 | 99.2 | 99.9 | 99.2 | 99.2 |
| | Average length | .095 | .132 | .116 | .094 | .119 | .160 | .189 | .139 | .187 | .161 | .162 |
| 30, 20 | Observed coverage(%) | 99.8 | 99.9 | 99.9 | 99.6 | 99.9 | 99.6 | 99.9 | 99.6 | 99.9 | 99.6 | 99.6 |
| | Average length | .119 | .202 | .172 | .127 | .205 | .299 | .363 | .304 | .420 | .303 | .308 |
| 10, 10 | Observed coverage(%) | 99.8 | 100 | 99.9 | 99.9 | 99.9 | 99.9 | 100 | 99.9 | 100 | 99.9 | 99.9 |
| | Average length | .133 | .333 | .240 | .122 | .380 | .570 | .699 | .197 | .920 | .572 | .592 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.14 Percentage of confidence intervals with deviation and zero width ($p_1 = 0.98$, $p_2 = 0.98$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 2.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 10.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 36.4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Zero width interval(%) | 66.7 | 0 | 0 | 66.7 | 0 | 0 | 0 | 66.7 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.15 Observed coverage and average length of nominal 95% confidence intervals ($p_1 = 0.98$, $p_2 = 0.02$)**

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Observed coverage(%) | 88.9 | 97.9 | 89.3 | 93.9 | 93.9 | 92.5 | 96.1 | 94.3 | 98.5 | 96.1 | 96.1 |
| | Average length | .076 | .097 | .088 | .086 | .086 | .094 | .108 | .096 | .116 | .085 | .085 |
| 60, 50 | Observed coverage(%) | 88.6 | 88.6 | 88.7 | 97.0 | 97.0 | 93.1 | 94.2 | 97.8 | 99.1 | 95.7 | 95.7 |
| | Average length | .095 | .131 | .115 | .118 | .118 | .135 | .158 | .139 | .171 | .117 | .118 |
| 30, 20 | Observed coverage(%) | 63.0 | 99.9 | 63.1 | 92.2 | 94.8 | 93.4 | 93.7 | 95.7 | 99.1 | 96.1 | 96.1 |
| | Average length | .118 | .201 | .171 | .199 | .202 | .239 | .290 | .237 | .295 | .206 | .209 |
| 10, 10 | Observed coverage(%) | 33.4 | 100 | 99.9 | 94.1 | 94.1 | 94.1 | 96.6 | 95.4 | 33.1 | 94.1 | 94.1 |
| | Average length | .133 | .333 | .241 | .369 | .370 | .430 | .535 | .416 | .519 | .364 | .379 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

**Table 3.16 Percentage of confidence intervals with deviation and zero width** ($p_1 = 0.98$, $p_2 = 0.02$)

| Sample Size | | Method | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_1$, $n_2$ | | As | AsC | HA | Hal | JP | Sc | ScC | W | WC | Mee | MN |
| 100, 90 | Violation of definition interval(%) | 45.2 | 81.7 | 66.9 | 0 | 2.1 | 0 | 2.1 | 25.7 | 26.7 | 0 | 0 |
| | Zero width interval(%) | 2.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60, 50 | Violation of definition interval(%) | 71.1 | 97.7 | 92.9 | 0 | 11.2 | 11.2 | 11.2 | 28.2 | 28.3 | 0 | 0 |
| | Zero width interval(%) | 11.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30, 20 | Violation of definition interval(%) | 61.2 | 99.9 | 99.7 | 0 | 36.9 | 0 | 36.9 | 19.9 | 19.9 | 0 | 0 |
| | Zero width interval(%) | 36.9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10, 10 | Violation of definition interval(%) | 33.3 | 100 | 99.9 | 0 | 0 | 0 | 66.6 | 15.1 | 15.3 | 0 | 0 |
| | Zero width interval(%) | 66.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | |
|---|---|---|---|
| As | Asymptotic without correction | Sc | Score interval without correction |
| AsC | Asymptotic with correction | ScC | Score interval with correction |
| HA | Asymptotic with Hauck and Anderson's correction | W | Wallenstein's interval without correction |
| | | WC | Wallenstein's interval with correction |
| Hal | Haldane interval | Mee | Mee's interval |
| JP | Jeffreys-Perks interval | MN | Miettinen and Nurminen interval |

## 3.2 Discussion of simulation study results

The ideal interval method should result in an interval that is short but has actual coverage close to the nominal coverage, does not deviate from the definition interval, and has no occurrence of zero width intervals.

### 3.2.1 Case $p_1 = 0.6$ and $p_2 = 0.4$ (Table 3.3 and 3.4)

All methods have relative good coverage except the Asymptotic method without correction (As). The methods with continuity correction exceed the nominal coverage of the interval. When the sample size is small ($n_1 = n_2 = 10$) some of the methods deviate from the definition interval.

Methods recommended for large($n \geq 50$) and small samples are the Haldane (Hal), Jeffreys-Perks (JP), Score interval without correction (Sc), Wallenstein's interval without correction (W), Mee (Mee) and Miettinen and Nurminen (MN) intervals. Easy calculable methods are the Haldane interval, Jeffreys-Perks interval and Score interval without correction.

### 3.2.2 Case $p_1 = 0.95$ and $p_2 = 0.85$ (Table 3.5 and 3.6)

Most of the methods have a good coverage when the sample size is at least moderately large. When the sample is smaller the methods either over or under cover. Violation of the definition interval and zero width intervals also occur when the sample size is small. Continuity corrected methods tend to exceed the nominal coverage of the interval.

Methods recommended for large samples are the Jeffreys-Perks (JP), Score interval without correction (Sc), Wallenstein's interval without correction (W), Mee (Mee) and Miettinen and Nurminen (MN) intervals.

The recommended method for small samples is the Jeffreys-Perks (JP). The Mee (Mee) and Miettinen and Nurminen (MN) intervals do not attain nominal coverage for sample size combination (30,20). The Jeffreys-Perks (JP) method is easy calculable.

### 3.2.3 Case $p_1 = 0.95$ and $p_2 = 0.15$ (Table 3.7 and 3.8)

Most of the methods have a good coverage for large sample sizes. As the sample size gets smaller most of the methods either exceed or attain coverage less than the nominal coverage of the interval. Violation of the definition interval is common when the sample size is less than 100.

Methods recommended for large samples are the Asymptotic with Hauck and Anderson's correction (HA), Haldane (Hal), Jeffreys-Perks (JP), Mee (Mee) and Miettinen and Nurminen (MN) intervals.

Recommended methods for small samples are the Jeffreys-Perks (JP), Score interval without correction (Sc), Mee (Mee) and Miettinen and Nurminen (MN) intervals. Easy calculable methods are the Jeffreys-Perks and Score interval without correction.

### 3.2.4 Case $p_1 = 0.98$ and $p_2 = 0.9$ (Table 3.9 and 3.10)

Most of the methods exceed or attain coverage less than the nominal coverage of the interval for large sample sizes (100,90). As the sample size decreases the tendency for exceeding or never attaining the nominal coverage of the interval grows in all the methods. Methods with continuity correction tend to exceed the nominal coverage.

Methods recommended for large samples are the Jeffreys-Perks (JP), Wallenstein's interval without correction (W), Mee (Mee) and Miettinen and Nurminen (MN) intervals.

Recommended methods for small samples are the Jeffreys-Perks (JP), Score interval without correction (Sc), Mee (Mee) and Miettinen and Nurminen (MN) intervals are recommended. It must be noted that for sample size combination $n_1 = n_2 = 10$ all recommended methods have coverage close to 100%. Easy calculable methods are the Jeffreys-Perks and Score interval without correction.

### 3.2.5 Case $p_1 = 0.98$ and $p_2 = 0.1$ (Table 3.11 and 3.12)

Methods either exceed or never attain the nominal coverage of the interval and this worsens as the sample size decreases. Violation of the definition interval also occurs in some methods for all sample size combinations.

Methods recommended for large samples are Haldane (Hal), Jeffreys-Perks (JP), Score interval without correction (Sc), Mee (Mee) and Miettinen and Nurminen (MN) intervals.

Recommended methods for small samples are the Mee (Mee) and Miettinen and Nurminen (MN) intervals, though both methods exceed the nominal coverage of the interval.

### 3.2.6 Case $p_1 = 0.98$ and $p_2 = 0.98$ (Table 3.13 and 3.14)

Most of the methods exceed the nominal coverage of the interval. This occurs more as the sample size gets smaller. Zero width intervals occur in some of the methods for the small sample size($n_1 = 10 = n_2$).

Most methods exceed the nominal coverage, but taken length, deviations and zero width intervals into account the following methods could be used. Methods recommended for large samples are the Asymptotic with Hauck and Anderson's correction (HA), Haldane (Hal), Jeffreys-Perks (JP), Score interval without correction (Sc), Wallenstein's interval without correction (W), Mee (Mee) and Miettinen and Nurminen (MN) intervals.

Recommended methods for small samples are the Jeffreys-Perks (JP), Score interval without correction (Sc), Mee (Mee) and Miettinen and Nurminen (MN) intervals. Easy calculable methods are then the Jeffreys-Perks and Score interval without correction.

### 3.2.7 Case $p_1 = 0.98$ and $p_2 = 0.02$ (Table 3.15 and 3.16)

Methods either exceed or do not attain the nominal coverage of the interval for large sample sizes. For smaller sample sizes this tendency is continued.

Violation of the definition interval occurs in most of the methods for all sample size combinations.

It is difficult to recommend a method in this case as they either exceed or never attain the nominal coverage of the interval. Here the methods of Mee (Mee) and Miettinen and Nurminen (MN) may be used for large and small sample sizes as both methods do not suffer from any deviations or occurrence of zero width intervals. Both methods tend to exceed the nominal coverage of the interval for large samples and never attain the nominal coverage of the interval for the smallest sample size combination.

### 3.2.8 An investigation of the occurrence of zero width intervals and violation of definition interval for two special cases

As mentioned earlier the conventional Asymptotic interval method suffers from the occurrence of zero width intervals and violation of the definition interval. (see section 1.3.2) We calculated examples with the different interval methods for these cases which are reported in Table 3.17 and 3.18.

**Table 3.17 95% Confidence intervals for $\hat{p}_1 = 0$ and $\hat{p}_2 = 0$.**

| | Sample sizes $(n_1, n_2)$ | | | |
|---|---|---|---|---|
| Method | (100,90) | (60,50) | (30,20) | (10,10) |
| Asymptotic without correction | [ 0 ; 0 ] | [ 0 ; 0 ] | [ 0 ; 0 ] | [ 0 ; 0 ] |
| Asymptotic with correction | [ -0.01 ; 0.01 ] | [ -0.02 ; 0.02 ] | [ -0.04 ; 0.04 ] | [ -0.10 ; 0.10 ] |
| Asymptotic with Hauck and Anderson's correction | [ -0.01 ; 0.01 ] | [ -0.01 ; 0.01 ] | [ -0.03 ; 0.03 ] | [ -0.05 ; 0.05 ] |
| Haldane interval | [ -0.002 ; 0 ] | [ -0.01 ; 0 ] | [ -0.03 ; 0 ] | [0 ; 0 ] |
| Jeffreys-Perks interval | [ -0.02 ; 0.02 ] | [ -0.04 ; 0.03 ] | [ -0.09 ; 0.06 ] | [ -0.17 ; 0.17 ] |
| Score interval without correction | [ -0.04 ; 0.04 ] | [ -0.07 ; 0.06 ] | [ -0.16 ; 0.11 ] | [ -0.28 ; 0.28 ] |
| Score interval with correction | [ -0.05 ; 0.05 ] | [ -0.09 ; 0.07 ] | [ -0.20 ; 0.14 ] | [ -0.34 ; 0.34 ] |
| Wallenstein's interval without correction | [ -0.04 ; 0.04 ] | [ -0.06 ; 0.06 ] | [ -0.11 ; 0.11 ] | [ -0.28 ; 0.28 ] |
| Wallenstein's interval with correction | [ -0.06 ; 0.06 ] | [ -0.09 ; 0.09 ] | [ -0.18 ; 0.18 ] | [ -0.40 ; 0.40 ] |
| Mee's interval | [ -0.04 ; 0.04 ] | [ -0.07 ; 0.06 ] | [ -0.16 ; 0.11 ] | [ -0.28 ; 0.28 ] |
| Miettinen and Nurminen interval | [ -0.04 ; 0.04 ] | [ -0.07 ; 0.06 ] | [ -0.16 ; 0.12 ] | [ -0.29 ; 0.29 ] |

**Table 3.18 95% Confidence intervals when $\hat{p}_1 = 1.0$ and $\hat{p}_2 = 0$.**

| | Sample sizes $(n_1, n_2)$ | | | |
|---|---|---|---|---|
| Method | (100,90) | (60,50) | (30,20) | (10,10) |
| Asymptotic without correction | [ 1.00 ; 1.00 ] | [ 1.00 ; 1.00 ] | [ 1.00 ; 1.00 ] | [ 1.00 ; 1.00 ] |
| Asymptotic with correction | [ 0.98 ; 1.01 ] | [ 0.98 ; 1.02 ] | [ 0.96 ; 1.04 ] | [ 0.90 ; 1.10 ] |
| Asymptotic with Hauck and Anderson's correction | [ 0.99 ; 1.01 ] | [ 0.99 ; 1.01 ] | [ 0.98 ; 1.03 ] | [ 0.95 ; 1.05 ] |
| Haldane interval | [ 0.96 ; 1.00 ] | [ 0.93 ; 1.00 ] | [ 0.85 ; 1.00 ] | [ 0.68 ; 1.00 ] |
| Jeffreys-Perks interval | [ 0.96 ; 1.00 ] | [ 0.93 ; 1.00 ]* | [ 0.85 ; 1.00 ]* | [ 0.68 ; 1.00 ]* |
| Score interval without correction | [ 0.94 ; 1.00 ] | [ 0.91 ; 1.00 ] | [ 0.80 ; 1.00 ] | [ 0.61 ; 1.00 ] |
| Score interval with correction | [ 0.93 ; 1.00 ]* | [ 0.88 ; 1.00 ]* | [ 0.75 ; 1.01 ] | [ 0.51 ; 1.01 ] |
| Wallenstein's interval without correction | [ 0.96 ; 1.00 ] | [ 0.93 ; 1.00 ] | [ 0.85 ; 1.00 ] | [ 0.68 ; 1.00 ] |
| Wallenstein's interval with correction | [ 0.94 ; 0.99 ] | [ 0.89 ; 0.98 ] | [ 0.78 ; 0.96 ] | [ 0.53 ; 0.92 ] |
| Mee's interval | [ 0.96 ; 1.00 ] | [ 0.93 ; 1.00 ] | [ 0.84 ; 1.00 ] | [ 0.68 ; 1.00 ] |
| Miettinen and Nurminen interval | [ 0.96 ; 1.00 ] | [ 0.93 ; 1.00 ] | [ 0.84 ; 1.00 ] | [ 0.66 ; 1.00 ] |

* Violation of the definition interval occurs at or later than the third decimal

### 3.2.8.1 95% Confidence intervals for $\hat{p}_1 = 0, \hat{p}_2 = 0$ (Table 3.17)

Zero width intervals occur with the Asymptotic without correction for all sample size combinations and Haldane interval method for the combination $n_1 = n_2 = 10$.

### 3.2.8.2 95% Confidence intervals for $\hat{p}_1 = 1.0, \hat{p}_2 = 0$ (Table 3.18)

Violation of the definition interval occur with the Asymptotic with correction and Asymptotic with Hauck and Anderson's correction for all sample size combinations. For the Score interval with correction it occurs for sample size combination $n_1 = 100, n_2 = 90$ [ 0.93 ; 1.0014 ] and $n_1 = 60, n_2 = 50$ [ 0.88 ; 1.0024 ]. For the Jeffreys-Perks interval it occurs for the sample size combination $n_1 = 60, n_2 = 50$ [ 0.93 ; 1.00003], $n_1 = 30, n_2 = 20$ [ 0.85 ; 1.00014 ] and $n_1 = n_2 = 10$ [ 0.68 ; 1.00149 ]. Zero width intervals occur with the Asymptotic without correction for all sample size combinations.

## 3.3 Recommendations

We recommend the interval methods of Miettinen and Nurminen and Mee for use across all cases including small sample size. They are, however, very computer intensive because of their iterative nature. The best of the more easy calculable methods is the Jeffreys-Perks method, which is always satisfactorily if sample size is $\geq 50$. The second best is the Score interval without correction.

Beal(1987) also came to this conclusion in his more limited comparison of methods, when he noted that if one wants to implement a more sophisticated and complicated interval using a computer program, the Mee and Miettinen and Nurminen intervals are good choices. Beal also recommended the use of the Jeffreys-Perks interval when one wants to compute a simple interval.

Newcombe(1998:b) suggests that his methods are the methods yielding shortest intervals among the intervals, which are easily calculated. We differ somewhat from this conclusion and found that the Jeffreys-Perks interval yields the best results among easily calculable intervals. Newcombe's objection that some of the Bayes estimates of parameters used in the calculation process lie outside the definition interval [ 0 ; 1 ] may be overlooked since the resulting interval has better coverage, shorter length and less violation of the definition interval than the Score methods.

The Jeffreys-Perks method has a coverage close to 95% or higher and also a relatively short length for all sample sizes $\geq 50$. Most practical situations in clinical trials have a sample size of $\geq 50$. The Jeffreys-Perks method is thus a good method to use.

## APPENDIX A
## COMPUTER PROGRAM TO CALCULATE THE CONFIDENCE INTERVALS DESCRIBED

```
c
c
c   This program produces thirteen types of confidence
c   intervals for theta=p1-p2 for a 2x2 contingency table:
c
c    1. Simple asymptotic, without continuity correction
c    2. Simple asymptotic, with continuity correction
c    3. Beal's Haldane interval.
c    4. Beal's Jeffreys-Perks interval.
c    5. Mee interval.
c    6. Miettinen-Nurminen interval.
c    7. Asymptotic with Hauck and Anderson correction
c    8. Score interval without continuity correction
c    9. Score interval with continuity correction
c   10. Wallenstein's interval without continuity correction
c   11. Wallenstein's interval with continuity correction
c
c
c
c   Limitation: group totals m and n not to exceed 200.
c
c
c   Number of positive outcomes in first sample, x1.
c   Total number in first sample, n1.
c   Number of positive outcomes in second sample, x2.
c   Total number in second sample, n2.
c
c
      implicit double precision (a-h,l-z)
      implicit integer (i-k)
        complex(8)
AAA,BBB,CCC,dA,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,CCW,l32,
     * ep,xk,mag1,mag2,c1,c2,seor,lor,cc1,
        * cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,bo,
     * onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,dda,db,dc,
     * ddd,dm,dn
```

```
      double precision d1(2),d2(2),pp(2),prr(2,101),thb(2),e(11,2),
     & x(11,2),xx(101,101,11,2),y(11,2,3),fcase(4),flo(4),
     & ee(101,101,11,2),xt(101,101,2),et(101,101,2)
      integer ilo(101,101,4)     ! ,ilt(101,101)
      data thb/-1.0d+0,1.0d+0/
      common /a1/ a,aa,b,bb,c,cc,conf,d,deriv1,deriv2,diff,dA,
     &
d1,d2,e,ee,et,AAA,BBB,CCC,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,
CCW,ep,
     & f,fcase,flo,four,half,hth,l1,l2,m,n,nm,nn,np11,np22,nx,one,oldps,
     & p,pbar,pi,pp,pr,prr,ps,psihat,psimin,psimax,psi0,p1,p2,l32,
     & q,q1,q2,r,rw,ss,sss,se,sum,tail,
     & th,thb,theta0,thhat,three,ths,th1,th2,tol7,tol8,tol10,tol12,two,
     & u,u1,u2,v,w,x,x1,x2,xx1,xx2,xt,xx,y,z,zero,
     & i,ia,iamax,ib,ic,icase,icmax,id,ierr,ihcf,ilo,ih,
     & im,im0,in0,in,inp,intabs,inx,iopt,ip,
     & ir,ir0,is,itab,iter,itest,it0,i1,i2,j,k,
     & xk,mag1,mag2,c1,c2,seor,lor,cc1,
         & cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,
     & bo,onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,
         & dda,db,dc,ddd,dm,dn

c
c
c  initialise constants
c
      open(11,file='c:\riette\dif.out')
      tol7  =1.0d-7
      tol8  =1.0d-8
      tol10 =1.0d-10
      tol12 =1.0d-12
      zero  =0.0d+0
      half  =0.5d+0
      one   =1.0d+0
      two   =2.0d+0
      three =3.0d+0
      four  =4.0d+0
      conf  =9.5d+1
      tail  =half*(one-conf*1.0d-2)
      z     =1.959963985d+0
      intabs=1
```

```
1000 print*,'Number of positive outcomes in first sample: ia'
   print*, 'Total number in first sample: im'
   print*, 'Number of positive outcomes in second sample: ib'
   print*, 'Total number in second sample: in'
   print*, 'THE LARGEST SAMPLE SIZE FOR THE MOMENT IS 200'
   read (5,*) ia,im,ib,in
   if  (ia .ge. 0 .and. ib .ge. 0
  & .and. ia .le. im .and. ib .le. in
  & .and. im .gt. 0 .and. im .le. 200
  & .and. in .gt. 0 .and. in .le. 200) goto 1200
   write (11,9710) ia,im,ib,in
   goto 1300
1200 call sixmethods
   write(11,9299)ia,im,ib,in
9299 format(/'ia=',i5 ' im=',i5,' ib=',i5,' in=',i5)
9301 format('Asymptotic interval without continuity correction'/
  * 2F15.10)
   write(11,9301) (x(1,k),k=1,2)
9302 format(/'Asymptotic interval with continuity correction'/2F15.10)
   write (11,9302) (x(2,k),k=1,2)
9303 format(/'Haldane interval'/2F15.10)
   write (11,9303) (x(3,k),k=1,2)
9304 format(/'Jeffrey-Perks interval'/2F15.10)
   write (11,9304) (x(4,k),k=1,2)
9305 format(/'Mee''s interval'/2F15.10)
   write (11,9305) (x(5,k),k=1,2)
9306 format(/'Miettinen and Nurminen interval'/2F15.10)
   write (11,9306) (x(6,k),k=1,2)
9307 format(/'Asymptotic interval with Hauck and Anderson''s'
  *,' correction'/2F15.10)
   write (11,9307) (x(7,k),k=1,2)
9308 format(/'Score interval without continuity correction'/2F15.10)
   write (11,9308) (x(8,k),k=1,2)
9309 format(/'Score interval with continuity correction'/2F15.10)
   write (11,9309) (x(9,k),k=1,2)
7304 format(/'Wallenstein interval without continuity correction'
  */2F15.10)
   write (11,7304) (x(10,k),k=1,2)
7305 format(/'Wallenstein interval with continuity correction'/2F15.10)
   write (11,7305) (x(11,k),k=1,2)
```

```
 1300 continue
      close(11,status='keep')
 9710 format('Error: ',i3,' / ',i3,' - ',i3,' / ',i3)
      end

      subroutine sixmethods
c
c
      implicit double precision (a-h,l-z)
      implicit integer (i-k)
         complex(8)
AAA,BBB,CCC,dA,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,CCW,l32,
     * ep,xk,mag1,mag2,c1,c2,seor,lor,cc1,
         * cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,bo,
     * onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,dda,db,dc,
     * ddd,dm,dn
      double precision d1(2),d2(2),pp(2),prr(2,101),thb(2),e(11,2),
     & x(11,2),xx(101,101,11,2),y(11,2,3),fcase(4),flo(4),
     & ee(101,101,11,2),xt(101,101,2),et(101,101,2)
      integer ilo(101,101,4)    !  ,ilt(101,101)
      common /a1/ a,aa,b,bb,c,cc,conf,d,deriv1,deriv2,diff,dA,
     &
d1,d2,e,ee,et,AAA,BBB,CCC,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,
CCW,ep,
     & f,fcase,flo,four,half,hth,l1,l2,m,n,nm,nn,np11,np22,nx,one,oldps,
     & p,pbar,pi,pp,pr,prr,ps,psihat,psimin,psimax,psi0,p1,p2,l32,
     & q,q1,q2,r,rw,ss,sss,se,sum,tail,
     & th,thb,theta0,thhat,three,ths,th1,th2,tol7,tol8,tol10,tol12,two,
     & u,u1,u2,v,w,x,x1,x2,xx1,xx2,xt,xx,y,z,zero,
     & i,ia,iamax,ib,ic,icase,icmax,id,ierr,ihcf,ilo,ih,
     & im,im0,in0,in,inp,intabs,inx,iopt,ip,
     & ir,ir0,is,itab,iter,itest,it0,i1,i2,j,k,
     & xk,mag1,mag2,c1,c2,seor,lor,cc1,
         & cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,
     & bo,onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,
         & dda,db,dc,ddd,dm,dn
c
c
      if (im .gt. 0 .and. in .gt. 0) goto 2000
      write (11,9900) im,in
      return
```

```
 2000 ic=im-ia
      id=in-ib
      it0=in*ia-im*ib
      m=dble(float(im))
      n=dble(float(in))
      u=(one/m+one/n)/four
      v=(one/m-one/n)/four
      a=dble(float(ia))
      b=dble(float(ib))
      c=dble(float(ic))
      d=dble(float(id))
      thhat=a/m-b/n
      psihat=half*(a/m+b/n)
c
c     Usual Asymptotic interval without and with correction (1 & 2)
c
      se=dsqrt(a*c/m**3+b*d/n**3)
      x(1,1)=thhat-z*se
      x(2,1)=x(1,1)-two*u
      x(1,2)=thhat+z*se
      x(2,2)=x(1,2)+two*u
c
c Approximate general interval with Hauck and Anderson correction (7)
c
      p11=a/m
      p22=b/n

      q11=1-p11
      q22=1-p22
      hacor=one/(two*(min(m,n)))
      var=((p11*q11)/(m-1))+((p22*q22)/(n-1))
      hamaal=(z*(dsqrt(var)))+hacor
      x(7,1)=thhat-hamaal
      x(7,2)=thhat+hamaal
c
c Newcombe's score interval without continuity correction (8)
c
      theta1=(two*m*p11)+z**2
      vierk1=dsqrt(z**2+(four*m*p11*q11))
      deel1=two*(m+z**2)
      l31=(theta1-(z*vierk1))/deel1
```

```
      u31=(theta1+(z*vierk1))/deel1
      theta2=(two*n*p22)+z**2
      vierk2=dsqrt(z**2+(four*n*p22*q22))
      deel2=two*(n+z**2)
      l32=(theta2-(z*vierk2))/deel2
      u32=(theta2+(z*vierk2))/deel2
      ep=z*(sqrt((u31*(one-u31)/m)+(l32*(one-l32)/n)))
      sn=z*(sqrt((l31*(one-l31)/m)+(u32*(one-u32)/n)))
      x(8,1)=thhat-sn
      x(8,2)=thhat+ep
c
c Newcombe's score interval with continuity correction (9)
c
      theta11=(two*m*p11)+z**2-one
      vier11=z**2-two-(one/m)+((four*p11)*((m*q11)+one))
      vierk11=z*(dsqrt(vier11))
      dd1=two*(m+z**2)
      l11=(theta11-vierk11)/dd1
      theta12=(two*m*p11)+z**2+one
      vier12=z**2+two-(one/m)+((four*p11)*((m*q11)-one))
      vierk12=z*(dsqrt(vier12))
      u11=(theta12+vierk12)/dd1
      theta21=(two*n*p22)+z**2-one
      vier21=z**2-two-(one/n)+((four*p22)*((n*q22)+one))
      vierk21=z*(dsqrt(vier21))
      dd2=two*(n+z**2)
      l21=(theta21-vierk21)/dd2
      theta22=(two*n*p22)+z**2+one
      vier22=z**2+two-(one/n)+((four*p22)*((n*q22)-one))
      vierk22=z*(dsqrt(vier22))
      u21=(theta22+vierk22)/dd2
      tep=dsqrt(((u11-(a/m))**2+((b/n)-l21)**two))
      tsn=dsqrt((((a/m)-l11)**2+(u21-(b/n))**two))
      x(9,1)=thhat-tsn
      x(9,2)=thhat+tep
c
c       Wallenstein's interval without continuity correction  (10)
c
      eps=zero
         rw=m+n
c
```

```
c        SHOULD WE SWITCH INDICES? ONLY NECESSARY IF p11<p22.
c
         if (p11.lt.p22) then ss=one
         else ss=zero
         if (p11.lt.p22) goto 2
         goto 1

c        WHEN IT IS NOT NECESSARY TO SWITCH INDICES
c        SOLUTION FOR THE UPPER BOUND d:
1        if (a.eq.0.and.b.eq.0) goto 80
         goto 81
80       dA=(z**2/(2*m)+(z/dsqrt(m))*dsqrt(z**2/(4*m)))/(1+z**2/m)
         goto 6

81   rw=rw
         pbar=((m*p11)+(n*p22))/rw

         EPS1=p11-p22+eps
         if (EPS1.ge.one) then EPS1=one
         else  EPS1=EPS1

         AAA =  one+(z**2/rw)*(one+(m-n)**2/(m*n))
         BBB = -two*EPS1 + (z**2/(m*n))*(one-two*pbar)*(m-n)
         CCC = EPS1**2 -(z**2/(m*n)) *rw*pbar*(one-pbar)

         dA = (-BBB+sqrt(BBB**2 -four*AAA*CCC))/(two*AAA)


c        ddo=dA

         p1d=pbar+(dA*n/rw)
         p2d=pbar-(dA*m/rw)

c        [Set p2d=0   p1d=d and var(p1d,p2d) = d*(1-d)/m yielding:]
         if (p2d.lt.zero) goto 4
         if (p1d.gt.one) goto 5
         goto 6

4        dA=(EPS1+(z**2/(2*m))+(z/dsqrt(m))*dsqrt(EPS1*(1-EPS1)+
    *(z**2/(4*m))))/(1+(z**2/m))
goto 6
```

```
c       d = do
c        [Set p1d=1   p2d=1-d and var(p1d,psd) = (1-d)*d/n yielding:]

5   dA=EPS1+(z**2/(2*n))+(z/dsqrt(n))*dsqrt(EPS1*(1-EPS1)+
   * (z**2/(4*n)))/(1+(1/n)*z**2)

6       UCL =dA
        if (p11.eq.p22) goto 70
        goto 71
70      LCL=-UCL
        goto 14

c       SOLUTION FOR THE LOWER BOUND d`, NOW DENOTED AS dd:

71      EPS2=p11-p22-eps
        if (EPS2.le.zero)  then EPS2=zero
        else EPS2=EPS2


        AAD =  one + (z**2/rw)*(1+(m-n)**2/(m*n))
        BBD = -two*EPS2 + (z**2/(m*n))*(one-two*pbar)*(m-n)
        CCD = EPS2**2 - ((z**2)/(m*n)) *rw*pbar*(one-pbar)

        dd = (-BBD-sqrt(BBD**2 -four*AAD*CCD))/(two*AAD)


        p1dd=pbar+((dd*n)/rw)
        p2dd=pbar-((dd*m)/rw)

c        [Set p1dd=0 then p2dd = -dd  var(p1dd,p2dd) = -dd(1+dd)/n
c          yielding:]
        if  (p1dd.lt.zero) goto 7
        goto 20
7       dd=(EPS2+(z**2/(two*n))+z*dsqrt(one/n*(EPS2*(one-EPS2) +
   * (z*z/(four*n)))))/(one+((z*z)/n))



20      LCL = dd
```

```
      goto 14

c      WHEN IT IS NECESSARY TO SWITCH INDICES:
c      ss=1
2             nm=m
              nn=n
              xx1=a
              xx2=b
              m=nn
              n=nm
              a=xx2
              b=xx1
              p11=a/m
              p22=b/n

c     SOLUTION FOR THE UPPER BOUND d:
              rw=m+n
              pbar=(m*p11+n*p22)/rw
              EPS1=p11-p22+eps
              if (EPS1.ge.one) then  EPS1=one
              else EPS1=EPS1

              AW =  one+(z**2/rw)*(one+((m-n)**2/(m*n)))
              BW = -two*EPS1 + (z**2/(m*n))*(one-two*pbar)*(m-n)
              CW = EPS1**2 -(z**2/(m*n)) *rw*pbar*(one-pbar)

              dA = (-BW+sqrt(BW**2 -four*AW*CW))/(two*AW)

              p1d=pbar+((dA*n)/rw)
              p2d=pbar-((dA*m)/rw)

c     p2d=0   p1d=d and var(p1d,p2d) = d*(1-d)/m yielding:
              if (p2d.lt.zero) goto 8
              if (p1d.gt.1) goto 9
              goto 10
8             dA=(EPS1+(z**2/(2*m))+(z/dsqrt(m))*dsqrt(EPS1*(1-EPS1)+
     *(z**2/(4*m))))/(1+(z**2/m))


               goto 10
```

```
c      p1d=1   p2d=1-d and var(p1d,psd) = (1-d)*d/n yielding:
9             dA=EPS1+(z**2/(2*n))+(z/dsqrt(n))*dsqrt(EPS1*(1-EPS1)+
      * (z**2/(4*n)))/(1+(1/n)*z**2)


10            UCL =dA

53            if (p11.eq.p22) goto 13
              goto 15
13            LCL=-UCL
              goto 14

c      SOLUTION FOR THE LOWER BOUND d`, NOW DENOTED AS dd:
15            EPS2=p11-p22-eps
              if (EPS2.le.zero) then  EPS2=zero
              else EPS2=EPS2


              AAW =  one + (z**2/rw)*(one+(m-n)**2/(m*n))
              BBW = -two*EPS2 + (z**2/(m*n))*(one-two*pbar)*(m-n)
              CCW = EPS2**2 - (z**2/(m*n)) *rw*pbar*(one-pbar)

              dd = (-BBW-sqrt(BBW**2 -four*AAW*CCW))/(two*AAW)

       p1dd=pbar+((dd*n)/rw)
       p2dd=pbar-((dd*m)/rw)

c      p2dd = -dd and var(p1dd,p2dd) = -dd(1+dd)/n yielding
       if  (p1dd.eq.zero) goto 12
       goto 21
12     dd=(EPS2+(z**2/(two*n))+z*dsqrt(one/n*(EPS2*(one-EPS2) +
     * ((z*z)/(four*n)))))/(one+((z*z)/n))


21     LCL = dd

    vn=m
       vm=n
       m=vm
       n=vn
       vb=a
```

```
      va=b
      a=va
      b=vb

c     'SO THE CORRECT WALLENSTEIN CONFIDENCE INTERVAL IS
c     GIVEN BY:')

   nUCL=-UCL
      nLCL=-LCL
      LCL=nUCL
      UCL=nLCL


14    x(10,1)=LCL
      x(10,2)=UCL
c
c     Wallenstein's interval with continuity correction  (11)
c
      eps=1/(2*m)+1/(2*n)
      rw=m+n
      p11=a/m
      p22=b/n
c
c      SHOULD WE SWITCH INDICES? ONLY NECESSARY IF p11<p22.
c
      if (p11.lt.p22) then ss=one
      else ss=zero
      if (p11.lt.p22) goto 32
      goto 31

c      WHEN IT IS NOT NECESSARY TO SWITCH INDICES
c     SOLUTION FOR THE UPPER BOUND d:
31    rw=rw
      pbar=((m*p11)+(n*p22))/rw

      EPS1=p11-p22+eps
      if (EPS1.ge.one) then EPS1=one
      else  EPS1=EPS1

      AAA =  one+(z**2/rw)*(one+(m-n)**2/(m*n))
      BBB = -two*EPS1 + (z**2/(m*n))*(one-two*pbar)*(m-n)
```

```
      CCC = EPS1**2 -(z**2/(m*n)) *rw*pbar*(one-pbar)

      dA = (-BBB+sqrt(BBB**2 -four*AAA*CCC))/(two*AAA)



      p1d=pbar+(dA*n/rw)
      p2d=pbar-(dA*m/rw)

c      [Set p2d=0   p1d=d and var(p1d,p2d) = d*(1-d)/m yielding:]
      if (p2d.lt.zero) goto 34
      if (p1d.gt.one) goto 35
      goto 36

34    dA=(EPS1+(z**2/(2*m))+(z/dsqrt(m))*dsqrt(EPS1*(1-EPS1)+
     *(z**2/(4*m))))/(1+(z**2/m))

goto 36


c      [Set p1d=1   p2d=1-d and var(p1d,psd) = (1-d)*d/n yielding:]

35    dA=EPS1+(z**2/(2*n))+(z/dsqrt(n))*dsqrt(EPS1*(1-EPS1)+
     * (z**2/(4*n)))/(1+(1/n)*z**2)

36     UCL =dA

      if (p11.eq.p22) goto 60
      goto 61
60    LCL=-UCL
      goto 33
c     SOLUTION FOR THE LOWER BOUND d`, NOW DENOTED AS dd:

61    EPS2=p11-p22-eps

      if (EPS2.le.zero)  then EPS2=zero
      else EPS2=EPS2


            AAD =  one + (z**2/rw)*(1+(m-n)**2/(m*n))
```

```
          BBD = -two*EPS2 + (z**2/(m*n))*(one-two*pbar)*(m-n)
          CCD = EPS2**2 - ((z**2)/(m*n)) *rw*pbar*(one-pbar)

          dd = (-BBD-sqrt(BBD**2 -four*AAD*CCD))/(two*AAD)


          p1dd=pbar+((dd*n)/rw)
          p2dd=pbar-((dd*m)/rw)
c      [Set p1dd=0 then p2dd = -dd  var(p1dd,p2dd) = -dd(1+dd)/n
c        yielding:]
      if  (p1dd.lt.zero) goto 37
      goto 40
37    dd=(eps2-(z**2/(2*n)))+((z/sqrt(n))*
   & sqrt((z**2/(4*n))-(eps2*(1+eps2))))/(1+(z**2/n))



c     dd1=dd


40    LCL = dd

      goto 33

c      WHEN IT IS NECESSARY TO SWITCH INDICES:
c      ss=1
32    nm=m
      nn=n
      xx1=a
      xx2=b
      m=nn
      n=nm
      a=xx2
      b=xx1
      p11=a/m
      p22=b/n

c     SOLUTION FOR THE UPPER BOUND d:
      rw=m+n
      pbar=(m*p11+n*p22)/rw
      EPS1=p11-p22+eps
```

```
      if (EPS1.ge.one) then  EPS1=one
      else EPS1=EPS1

      AW =  one+(z**2/rw)*(one+((m-n)**2/(m*n)))
      BW = -two*EPS1 + (z**2/(m*n))*(one-two*pbar)*(m-n)
      CW = EPS1**2 -(z**2/(m*n)) *rw*pbar*(one-pbar)

      dA = (-BW+sqrt(BW**2 -four*AW*CW))/(two*AW)


      p1d=pbar+((dA*n)/rw)
      p2d=pbar-((dA*m)/rw)

c     p2d=0   p1d=d and var(p1d,p2d) = d*(1-d)/m yielding:
      if (p2d.lt.zero) goto 38
      if (p1d.gt.1) goto 39
      goto 50
38    dA=(EPS1+(z**2/(2*m))+(z/dsqrt(m))*dsqrt(EPS1*(1-EPS1)+
     *(z**2/(4*m))))/(1+(z**2/m))

c     d1=dA

     goto 50

c     p1d=1   p2d=1-d and var(p1d,psd) = (1-d)*d/n yielding:
39    dA=EPS1+(z**2/(2*n))+(z/dsqrt(n))*dsqrt(EPS1*(1-EPS1)+
     * (z**2/(4*n)))/(1+(1/n)*z**2)


50    UCL =dA
      if (p11.eq.p22) goto 46
      goto 45
46    LCL=-UCL
      goto 33

c     SOLUTION FOR THE LOWER BOUND d`, NOW DENOTED AS dd:
45    EPS2=p11-p22-eps
      if (EPS2.le.zero) then  EPS2=zero
      else EPS2=EPS2
```

```
      AAW =  one + (z**2/rw)*(one+(m-n)**2/(m*n))
      BBW = -two*EPS2 + (z**2/(m*n))*(one-two*pbar)*(m-n)
      CCW = EPS2**2 - (z**2/(m*n)) *rw*pbar*(one-pbar)

      dd = (-BBW-sqrt(BBW**2 -four*AAW*CCW))/(two*AAW)

      p1dd=pbar+((dd*n)/rw)
      p2dd=pbar-((dd*m)/rw)

c     p2dd = -dd and var(p1dd,p2dd) = -dd(1+dd)/n yielding
      if  (p1dd.eq.zero) goto 42
      goto 41
42   dd=(eps2-(z**2/(2*n)))+((z/sqrt(n))*
    & sqrt((z**2/(4*n))-(eps2*(1+eps2))))/(1+(z**2/n))



41    LCL = dd


c     'SO THE CORRECT WALLENSTEIN CONFIDENCE INTERVAL IS
GIVEN BY:')

    nUCL=-UCL
      nLCL=-LCL
      LCL=nUCL
      UCL=nLCL

33   x(11,1)=LCL
      x(11,2)=UCL
c
c   Beal's Haldane & Jeffreys-Perks methods 3 & 4
c
    do 2500 i=3,4                  ! method
    ilo(ia+1,ib+1,i)=0
    ps=(a/m+b/n)/two
    if (i .eq. 4) ps=((a+half)/(m+one)+(b+half)/(n+one))/two
    w=u*(four*ps*(one-ps)-thhat**2)
    w=w+two*v*(one-two*ps)*thhat
    w=w+four*(one-ps)*ps*(z*u)**2
```

```
      w=w+(z*v*(one-two*ps))**2
      w=z*dsqrt(w)/(one+z*z*u)
      th=(thhat+z*z*v*(one-two*ps))/(one+z*z*u)
      x(i,1)=th-w
      x(i,2)=th+w
      do 2200 k=1,2                 ! lower/upper
      do 2200 j=1,2                 ! 1st/2nd sample
      pi=ps-half*x(i,k)*(-one)**j
      if (pi .le. -tol12 .or. pi .ge. one+tol12) goto 2300
 2200 continue
      goto 2500
 2300 ilo(ia+1,ib+1,i)=1
 2500 continue
c
c     Mee's interval and Miettinen and Nurminen's interval (5 & 6)
c
 3000 do 4000 i=5,6                 ! method
      do 4000 k=1,2                 ! lower/upper
      if  ((k .eq. 1 .and. ia .eq. 0 .and. id .eq. 0)
     & .or. (k .eq. 2 .and. ib .eq. 0 .and. ic .eq. 0)) goto 3900
      th1=thhat
      th2=thb(k)
      th=(th1+th2)*half
      do 3200 iter=1,40             ! for theta within 1e-12
      call profile
      pp(1)=ps+hth
      pp(2)=ps-hth
      f=((th-thhat)/z)**2
      if (i .eq. 6) f=f*(one-one/(m+n))
      f=pp(1)*(one-pp(1))/m+pp(2)*(one-pp(2))/n-f
      if (f .lt. zero) goto 3100
      th1=th
      goto 3200
 3100 th2=th
 3200 th=(th1+th2)*half
      goto 3950
 3900 th=thb(k)
      ps=half
 3950 e(i,k)=ps
 4000 x(i,k)=th
      if (x(5,1) .lt. x(6,1) .or. x(5,2) .gt. x(6,2))
```

```
   & write (11,9800) ia,im,ib,in,((x(i,k),k=1,2),i=5,6)
    return
 9800 format('Warning: for',i5,' / ',i5,' - ',i5,' / ',i5/
   & 'Mee interval',f15.10,' to ',f15.10,' not a subset of'/
   & 'M-N interval',f15.10,' to ',f15.10)
 9900 format('Error: im=',i6,' in=',i6)
    end

    subroutine profile
c
c   Given a, m, b and n, this subroutine calculates the psi value, ps,
c   that maximises the likelihood for any hypothetical value of theta,
c   viz. th, between -1 and +1.
c   Four cases are distinguished according to the pattern of empty cells.
c
c
    implicit double precision (a-h,l-z)
    implicit integer (i-k)
        complex(8)
AAA,BBB,CCC,dA,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,CCW,l32,
   * ep,xk,mag1,mag2,c1,c2,seor,lor,cc1,
        * cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,bo,
   * onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,dda,db,dc,
   * ddd,dm,dn
    double precision d1(2),d2(2),pp(2),prr(2,101),thb(2),e(11,2),
   & x(11,2),xx(101,101,11,2),y(11,2,3),fcase(4),flo(4),
   & ee(101,101,11,2),xt(101,101,2),et(101,101,2)
    integer ilo(101,101,4)    !  ,ilt(101,101)
    common /a1/ a,aa,b,bb,c,cc,conf,d,deriv1,deriv2,diff,dA,
   &
d1,d2,e,ee,et,AAA,BBB,CCC,AAD,BBD,CCD,dd,AW,BW,CW,AAW,BBW,
CCW,ep,
   & f,fcase,flo,four,half,hth,l1,l2,m,n,nm,nn,np11,np22,nx,one,oldps,
   & p,pbar,pi,pp,pr,prr,ps,psihat,psimin,psimax,psi0,p1,p2,l32,
   & q,q1,q2,r,rw,ss,sss,se,sum,tail,
   & th,thb,theta0,thhat,three,ths,th1,th2,tol7,tol8,tol10,tol12,two,
   & u,u1,u2,v,w,x,x1,x2,xx1,xx2,xt,xx,y,z,zero,
   & i,ia,iamax,ib,ic,icase,icmax,id,ierr,ihcf,ilo,ih,
   & im,im0,in0,in,inp,intabs,inx,iopt,ip,
   & ir,ir0,is,itab,iter,itest,it0,i1,i2,j,k,
   & xk,mag1,mag2,c1,c2,seor,lor,cc1,
```

```
     & cc2,rbo,rlow,dc1,dc2,ee11,ee12,ee21,ee22,yy11,yy12,yy21,yy22,
     & bo,onder,maal,nbo1,nbo2,non1,non2,na,nb,nc,nd,add,cbb,or,
     & dda,db,dc,ddd,dm,dn

c
      ps=half
      hth=half*th
      if (dabs(th) .le. one+tol12) goto 500
      write (11,9900) th
      goto 5100
  500 if (dabs(th) .gt. one-tol12) goto 5000        ! then ps=half
      if ((ib .eq. 0  .and. ic .eq. 0) .or.
     &    (ia .eq. 0  .and. id .eq. 0)) goto 4000
      if ((ia .eq. 0  .and. ib .eq. 0) .or.
     &    (ic .eq. 0  .and. id .eq. 0)) goto 3000
      if  (ia .eq. 0  .or.  ib .eq. 0 .or.
     &     ic .eq. 0  .or.  id .eq. 0)  goto 2000
c
c     Case (i): no empty cells
c
 1000 psimin=dabs(hth)
      psimax=one-psimin
 1100 p1=ps+hth
      p2=ps-hth
      q1=one-p1
      q2=one-p2
      if (p1 .lt. tol12 .or. p2 .lt. tol12 .or.
     &    q1 .lt. tol12 .or. q2 .lt. tol12) goto 1900
      deriv1=a/p1+b/p2-c/q1-d/q2
      deriv2=-a/p1**2-b/p2**2-c/q1**2-d/q2**2
      oldps=ps
      ps=ps-deriv1/deriv2
      if (ps .le. psimin+tol10) ps=half*(psimin+oldps)
      if (ps .gt. psimax-tol10) ps=half*(psimax+oldps)
      if (dabs(ps-oldps) .gt. tol12) goto 1100
      goto 5000
 1900 write (11,9910) ia,im,ib,in,i,th,p1,p2
      goto 5000
c
c     Case (ii): one empty cell
c
```

```
 2000 aa=m+n
    if (ib .eq. 0) goto 2200
    if (ic .eq. 0) goto 2300
    if (id .eq. 0) goto 2400
 2100 ths=-(aa-dsqrt(aa**2-four*b*c))*half/c
    ps=-hth
    if (th .le. ths) goto 5000
    bb=c*(one-th)+d
    cc=-hth*((c-d)*(one-hth)+b*hth)
    goto 2500
 2200 ths=(aa-dsqrt(aa**2-four*a*d))*half/d
    ps=hth
    if (th .ge. ths) goto 5000
    bb=d*(one+th)+c
    cc=hth*((d-c)*(one+hth)-a*hth)
    goto 2500
 2300 ths=(aa-dsqrt(aa**2-four*a*d))*half/a
    ps=one-hth
    if (th .ge. ths) goto 5000
    bb=a*(one+th)+b
    cc=hth*((a-b)*(one+hth)-d*hth)
    goto 2500
 2400 ths=-(aa-dsqrt(aa**2-four*b*c))*half/b
    ps=one+hth
    if (th .lt. ths) goto 5000
    bb=b*(one-th)+a
    cc=-hth*((b-a)*(one-hth)+c*hth)
 2500 ps=(bb+dsqrt(bb**2-four*aa*cc))*half/aa
    if (ia .eq. 0 .or. ib .eq. 0) ps=one-ps
    goto 5000
c
c    Case (iii): two empty cells, same row.
c
 3000 ps=dabs(hth)
    if (ic .eq. 0) ps=one-ps
    goto 5000
c
c    Case (iv): two empty cells, same diagonal.
c
 4000 if (im .eq. in) goto 5000       ! then ps=half
    if (ia .eq. 0) goto 4500
```

```
    if (im .lt. in) goto 4300
    if (m*th .lt. n) goto 4400
4100 ps=one-hth
    goto 5000
4300 if (n*th .lt. m) goto 4400
    ps=hth
    goto 5000
4400 ps=(m+(m-n)*hth)/(m+n)
    goto 5000
4500 if (in .lt. im) goto 4800
    if (n*(-th) .lt. m) goto 4900
4600 ps=one+hth
    goto 5000
4800 if (m*(-th) .lt. n) goto 4900
    ps=-hth
    goto 5000
4900 ps=(n+(m-n)*hth)/(m+n)
5000 continue
5100 continue
9900 format('Error: theta=',f15.12,' out of range')
9910 format('Error: for',i5,' / ',i5,' - ',i5,' / ',i5,' method',i2
   & /'theta=',f15.12,' p1=',f16.12,' p2=',f16.12)
    return
    end
```

## REFERENCES


Altman DG. (1992) *Practical statistics for medical research.* London: Chapman & Hall

Anbar D. (1983) On estimating the difference between two probabilities, with speical reference to clinical trials. *Biometrics*, 39: 257-262

Beal SL. (1987) Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics*, 43: 941-950

Berry CC. (1990) A tutorial on confidence intervals for proportions in diagnostic radiology. *American Journal of Roëntology*, 154: 477-480

Böhning D. (1994) Better approximate confidence intervals for a binomial parameter. *The Canadian Journal of Statistics*, 22: 207-218

Blyth CR, Still HA. (1983) Binomial confidence intervals. *Journal of the American Statistical Association*, 78: 108-116

Brenner DJ, Quan H. (1990) Exact confidence limits for binomial proportions-Pearson and Hartley revisited. *The Statistician*, 39: 391-397

Cox DR, Hinkley DV. (1974) *Theoretical Statistics*. London: Chapman and Hall Ltd.

Edwardes M. (1994) Confidence intervals for a binomial proportion. *Statistics in Medicine*, 13: 1693-1698

Fleiss JL. (1981) *Statistical Methods for Rates and Proportions.* Second edition. New York: Wiley

Gardner MJ, Altman DG. (1989) *Statistics with Confidence - Confidence intervals and statistical guidelines.* London: British Medical Journal

Gardner MJ. (1989) *Statistics with Confidence - Confidence intervals and statistical guidelines.* London: British Medical Journal

Ghosh BK. (1979) A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, 74: 894-900

Good IJ. (1965) *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge, Massachusetts: The M.I.T. Press

Haldane J.B.S. (1945) On a method of estimating frequencies. *Biometrika*, 33: 222-225

Hauck WW, Anderson S. (1986) A comparison of large sample confidence intervals for p1-p2 and p1/p2 in 2x2 contingency tables. *Journal of the American Statistical Association*, 75: 386-394

Jeffreys H. (1961) *Theory of probability.* Oxford: Claredon Press

Mee RW. (1984) Confidence bounds for the difference between two probabilities (letter). *Biometrics*, 40: 1175-1176

Miettinen OS, Nurminen M. (1985) Comparative analysis of two rates. *Statistics in Medicine*, 4: 213-226

Mood AM, Graybill FA, Boes DC. (1974) *Introduction to the theory of statistics.* Tokyo: Mac Graw Hill

Newcombe RG. (1998:a) Two-sided confidence intervals for the single proportion. A comparative evaluation of seven methods. *Statistics in Medicine*, 17: 857-872

Newcombe RG. (1998:b) Interval estimation for the difference between independent proportions. A comparative evaluation of eleven methods. *Statistics in Medicine*, 17: 873-890

Perks W. (1947) Some observations on inverse probability including a new indifference rule. *Journal of Institute of Actuaries*, 73:285-312

Rosner B. (1990) *Fundamentals of Biostatistics.* Boston: PWS-Kent

Santner TJ, Yamagami S. (1993) Invariant small sample confidence intervals for the difference of two success probabilities. *Communication in Statistical Simulation*, 22 (1): 33-59

SAS Institute Inc., *SAS/STAT® Software: Changes and Enhancements through Release 6.12,* Cary, NC: SAS Institute., 1997. 1167pp

Snedecor GW, Cochran WG. (1980) *Statistical methods.* Ames: Iowa State University Press

Storer BE, Kim C. (1990) Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, 85: 146-155

Vollset SE. (1993) Confidence intervals for a binomial proportion. *Statistics in Medicine*, 12: 809-824

Wallenstein S. (1997) A Non-iterative accurate asymptotic confidence interval for the difference between two proportions. *Statistics in Medicine*,16: 1329-1336

Wilson EB. (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22: 209-212

Yates F. (1934) Contingency tables involving small numbers and the $\chi^2$ test. *Journal of the Royal Statistical Society Suppl.*,1: 217-235

Zar JH. (1984) *Biostatistical Analysis.* 2nd ed. New Jersey: Prentice-Hall