University of Granada

Department of Statistics and Operations Research

Doctoral Thesis

Doctoral Program in Mathematical and Applied Statistics

Inferences on the Weighted Kappa Coefficient of Binary Diagnostic Tests

Raid Amro

Granada

May, 2017

Editor: Universidad de Granada. Tesis Doctorales
Autor: Raid Amro
ISBN: 978-84-9163-628-1
URI: http://hdl.handle.net/10481/48613

University of Granada

Department of Statistics and Operations Research

Inferences on the Weighted Kappa Coefficient of Binary Diagnostic Tests

Report on the DOCTORAL THESIS written under the supervision of José Antonio Roldán Nofuentes, Associate professor of the Department of Statistics and Operations Research of the University of Granada, presented by the Graduate Raid Amro to obtain a PhD in Statistics

Signed: Raid Amro

Approval by the Supervisor

Signed: José Antonio Roldán Nofuentes

The doctoral candidate Raid Amro and the thesis supervisor José Antonio Roldán Nofuentes

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Place and date:

Granada. May, 2017

Thesis supervisor:

Doctoral candidate:

Signed: José Antonio Roldán Nofuentes

Signed: Raid Amro

This doctoral thesis is protected by the manuscript:

Roldán-Nofuentes, J.A., Amro, R. (2017). Approximate confidence intervals for the weighted kappa coefficient of a binary diagnostic test subject to a case-control design. Journal of Statistical Computation and Simulation, 87: 530-545. DOI: 10.1080/00949655.2016.1218881.

Acknowledgements

To the memory of my late Father Mohammad.

I would like to place on record my deepest gratitude to my family, especially my mother, my wife Ikhlas and my children Mohammad, Ashraf, Gwan and Mays, who tolerated my absence while away from them in Spain for several months. Indeed, their moral support is highly appreciated.

I also would like to express sincere thanks to my advisor Professor Jose Antonio Roldán Nofuentes, for his constant and continuous support throughout my entire study for the Ph.D. His immense knowledge, guidance and inspiration have been invaluable. I could not have imagined having a better mentor.

Index

Summary

Diagnostic Methods are fundamental in Clinical Medicine and in Epidemiology. Therefore, part of the discipline of Statistics has focused on the development of new methods to solve the problems that have been posed in this field, leading to what are known as Statistical Methods for Diagnosis in Medicine. This doctoral thesis seeks to contribute to research into new methods of estimation of parameters of binary diagnostic tests. It focuses on the study of binary diagnostic tests, whose assessment in relation to a gold standard gives rise to a 2×2 table when there is a single diagnostic test, or a 2×4 table when there are two binary diagnostic tests. In all the situations analysed in this Thesis, it is assumed that the disease status of all the individuals in the sample, or samples, is known. This doctoral thesis is structured in three Chapters.

In Chapter 1, the main parameters of a binary diagnostic test are defined and studied: sensitivity and specificity, likelihood ratios, predictive values and the weighted kappa coefficient.

Chapter 2 studies the estimations of the parameters presented in Chapter 1 when the study is cross-sectional and when it is case-control. The cross-sectional study consists of the application of the binary diagnostic test and the gold standard to all the individuals in a random sample; and the case-control study consists of applying the binary diagnostic test to all of the individuals in two samples, one of individuals with the disease (case sample) and another of individuals without the disease (control sample). The contribution made by this Chapter is the estimation of the weighted kappa coefficient subject to case-control sampling. Several confidence intervals are studied for this parameter, Monte Carlo simulation experiments are carried out to study the

asymptotic coverage of these intervals and a method is proposed to calculate the size of each sample. The results obtained are applied to real example.

Chapter 3 studies two different problems: the comparison of parameters of two binary diagnostic tests subject to a paired design and the combination of parameters of two binary diagnostic tests. On the one hand, we present the hypothesis tests and confidence intervals to compare the parameters of two binary diagnostic tests, and on the other, we study the combination of parameters of two binary diagnostic tests. The contribution of this Chapter is the combination of the weighted kappa coefficients of two binary diagnostic tests in parallel testing, defining the weighted kappa coefficient of the combination of the two diagnostic tests and studying its properties. We have studied the conditions in which the combination of the two diagnostic tests produces an increase in the weighted kappa coefficient of the combination. Fieller's method is applied to obtain a confidence interval for the ratio between the weighted kappa coefficient of the combination and each weighted kappa coefficient, and Monte Carlo simulation experiments are carried out to study the asymptotic behaviour of this interval. An *R* program is written to solve the problem posed and the results were applied to a real example.

Granada

May, 2017

Chapter 1

Binary Diagnostic Test and its Parameters

1.1. Introduction

The diagnosis of diseases is fundamental in the practice of Medicine and the study of Statistical Methods for Diagnosis in Medicine is an important topic in Biostatistics (Zhou et al, 2002; Pepe, 2003). A diagnostic test (*DT*) is medical test that is applied to an individual in order to determine the presence or absence of a certain disease. The level of *PSA* for the diagnosis of prostate cancer and a stress test for the diagnosis of coronary artery disease are two examples of *DTs*. A *DT* can be helpful for screening, for diagnosis and for individual management:

a) Screening: to detect a disease in asymptomatic individuals, and therefore reduce disease morbidity and mortality through early treatment.
b) Diagnosis: establish or exclude the presence of a disease in symptomatic individuals.
c) Individual management: evaluate the severity of a disease, estimate prognosis, monitor the course of a disease (progression, stability, or resolution), detect disease recurrence, select drugs and therapy and adjust them.

The application of a *DT* for the assessment of diseases has different purposes (Zhou et al, 2002):

a) To provide reliable information about the disease status of an individual.
b) To influence the planning of the treatment of an individual.
c) To understand the mechanism and the nature of the disease through research.

The interpretation of a *DT* depends on several factors (Zhou et al, 2002):

a) The intrinsic ability of the *DT* in order to distinguish between diseased and non-diseased individuals (discriminatory accuracy).
b) The particular characteristics of each individual.
c) The environment in which the *DT* is applied.

A *DT* may make a mistake in the diagnosis of the disease status of an individual, and therefore the accuracy of a *DT* is measured in terms of probabilities (or functions of them). When the result of a *DT* is positive (indicating the presence of the disease) or negative (indicating the absence of the disease), the *DT* is called a binary diagnostic test (*BDT*) and its accuracy is measured in terms of two fundamental parameters: sensitivity

and specificity. Other parameters to assess the performance of a *BDT* are the likelihood ratios (*LRs*), the predictive values (PVs) and the weighted kappa coefficient $\left(\kappa(c)\right)$. In order to obtain an unbiased estimators of the sensitivity and the specificity of the *BDT*, and therefore for the other parameters, it is necessary to know the true disease status (present or absent) of each individual in the random sample. The medical test through which the true disease status of each individual is known is called the gold standard (*GS*). A biopsy for the diagnosis of prostate cancer and an angiography for the diagnosis of coronary disease are two examples of *GS*. Therefore, there are two methods to diagnose the disease: *DT* and *GS*. The *DT* may make a mistake in the diagnosis of the disease while the *GS* does not. Why not always use the *GS*? There are several reasons to use a *DT* instead of a *GS*:

a) The *DT* is usually less expensive than the *GS*.

b) The use of a *GS* may lead to some risk for the individual. For example, a coronary angiography may cause a stroke, thrombosis or even death of the individual.

c) There is not always a *GS*. For example, in the case of some psychiatric diseases.

The objective is to estimate the accuracy of the *BDT*, not to know if the individual has the disease or not. In the following sections of this chapter we will study each of the parameters of a *BDT*:

a) Sensitivity and specificity.

b) Likelihood ratios.

c) Predictive values.

d) Weighted kappa coefficient.

1.2. Sensitivity and specificity

Sensitivity and specificity are the fundamental measures of accuracy of a *BDT*. Let *D* be the random variable which models the result of the *GS*, so that $D=1$ when an individual does have the disease and $D=0$ when an individual does not have the disease. Let T be the random variable which models the result of the *BDT*, in such a way that $T=1$ when the result is positive (indicating the presence of the disease) and $T=0$ when it is negative (indicating the absence of the disease). The probability of a randomly chosen individual from among all of the individuals in the population having the disease, denoted as $p=P(D=1)$, is known as the disease prevalence. Sensitivity (*Se*) is the probability that the result of the *BDT* will be positive when an individual has the disease, i.e.

$$Se=P(T=1|D=1). \tag{1.1}$$

The probability $P(T=0|D=1)=1-Se$ is called the probability of a false negative. Specificity (Sp) is the probability that the result of the *BDT* will be negative when the individual does not have the disease, i.e.

$$Sp=P(T=0|D=0). \tag{1.2}$$

The probability $P(T=1|D=0)=1-Sp$ is called the probability of a false positive. Sensitivity and specificity only depend on the intrinsic ability of the *BDT* to distinguish between individuals who have the disease and those who do not, i.e. these parameters depend on the physical, chemical and biological bases upon which the *BDT* has been developed, and they are not affected by the prevalence of the disease. A *BDT* with high

Se is useful to exclude a diagnosis because a highly sensitive test will render few results that are falsely negative. A *BDT* with high *Sp* is useful to confirm a diagnosis, because a highly specific test will have few results that are falsely positive. Cicchetti et al (1995) have classified the *Se* and the *Sp* at the following intervals: $< 70\%$ "Poor", $70\% - 79\%$ "Fair", $80\% - 89\%$ "Good" and $90\% - 100\%$ "Excellent".

The parameter $Y = Se + Sp - 1$ is called Youden index (Youden, 1950). The Youden index is a summary measure of accuracy of a *BDT*. The Youden index does not depend on the prevalence of disease and it indicates the likelihood of a positive result among individuals with the disease versus those without the disease. The Youden index has the following property: if the *BDT* and the disease are independent, then *Se* and *Sp* are complementary $(Se + Sp = 1)$. This aspect is not desirable in a *BDT*, and therefore it is demanded that the Youden index of a *BDT* be bigger than zero $(Y > 0$ or $Se + Sp > 1)$. If $Y < 0$ the results of the diagnosis are interchanged, $T = 1$ should be a negative result and $T = 0$ should be a positive result, and the analysis should be limited only to the positive values of the Youden index.

1.3. Likelihood ratios

Other parameters for evaluating the performance of a *BDT* are the likelihood ratios (*LRs*). When the result of the *BDT* is positive, the *LR* (called the positive likelihood ratio, LR^+) is the quotient between the probability of a positive result for the *BDT* when the individual has the disease and the probability of a positive result for the *BDT* when the individual does not have the disease, i.e.

$$LR^{+} = \frac{Se}{1-Sp}. \tag{1.3}$$

When the result of the *BDT* is negative, the *LR* (called the negative likelihood ratio, LR^{-}) is the quotient between the probability of a negative result of the *BDT* when the individual has the disease and the probability of a negative result of the *BDT* when the individual does not have the disease, i.e.

$$LR^{-} = \frac{1-Se}{Sp}. \tag{1.4}$$

The *LRs* depend on the *Se* and *Sp* of the diagnostic test and do not depend on the disease prevalence, and their values vary between zero and infinite. When the result of the *BDT* and the disease status are independent, then $LR^{+} = LR^{-} = 1$. When the *BDT* correctly classifies all of the individuals (diseased and non-diseased), then $LR^{+} = \infty$ and $LR^{-} = 0$. A value $LR^{+} > 1$ indicates that a positive result of the BDT is more probable in a diseased individual than in a non-diseased individual, and a value $LR^{-} < 1$ indicates that a negative result of the *BDT* is more probable in a non-diseased individual than in a diseased individual. The *LRs* quantify the increase in knowledge of the presence of the disease through the application of the *BDT*. Before applying the *BDT*, the odds of a individual being diseased is

$$\text{pre-test odds} = \frac{p}{1-p},$$

where *p* is the disease prevalence. After applying the *BDT*, the odds of disease are

$$\text{post-test odds} = \frac{P\left(D=1 \middle| T=j\right)}{P\left(D=0 \middle| T=j\right)},\ j=0,1.$$

The *LRs* relate the pre-test odds and post-test odds

$$\text{Post test odds } (T=1) = LR^{+} \times \text{pre test odds}$$
$$\text{Post test odds } (T=0) = LR^{-} \times \text{pre test odds},$$

Therefore, the likelihood ratios quantify the change in the odds of disease obtained by knowledge of the application of the *BDT*.

1.4. Predictive values

Predictive values (*PVs*) are the measures of the clinical accuracy of a *BDT*. When the result of the *BDT* is positive, the *PV* (called the positive predictive value, *PPV*) is the probability of a individual being diseased when the test result is positive, i.e.

$$PPV = P(D=1|T=1) = \frac{p \times Se}{p \times Se + q \times (1-Sp)}, \tag{1.5}$$

where $q = 1-p$. When the result of the *BDT* is negative, the *PV* (called the negative predictive value, *NPV*) is the probability of a individual not being diseased when the test result is negative, i.e.

$$NPV = P(D=0|T=0) = \frac{q \times Sp}{p \times (1-Se) + q \times Sp}. \tag{1.6}$$

While sensitivity and specificity quantify how well the *BDT* reflexes the true disease status (present or absent), the *PVs* quantify the clinical value of the *BDT*, since both the clinic and the individual are more interested in knowing how probable it is to have the disease given a diagnostic test result. Cicchetti et al (1995) have classified the *PPV* and the *NPV* at the following intervals: $<70\%$ "Poor", $70\%-79\%$ "Fair", $80\%-89\%$ "Good" and $90\%-100\%$ "Excellent". In Figures 1.1 and 1.2 we show how the *PVs* varies according to the prevalence *p* for different values of *Se* and *Sp*.

Figure 1.1. Predictive values for $Se = 0.95$ and $Sp = 0.90$.

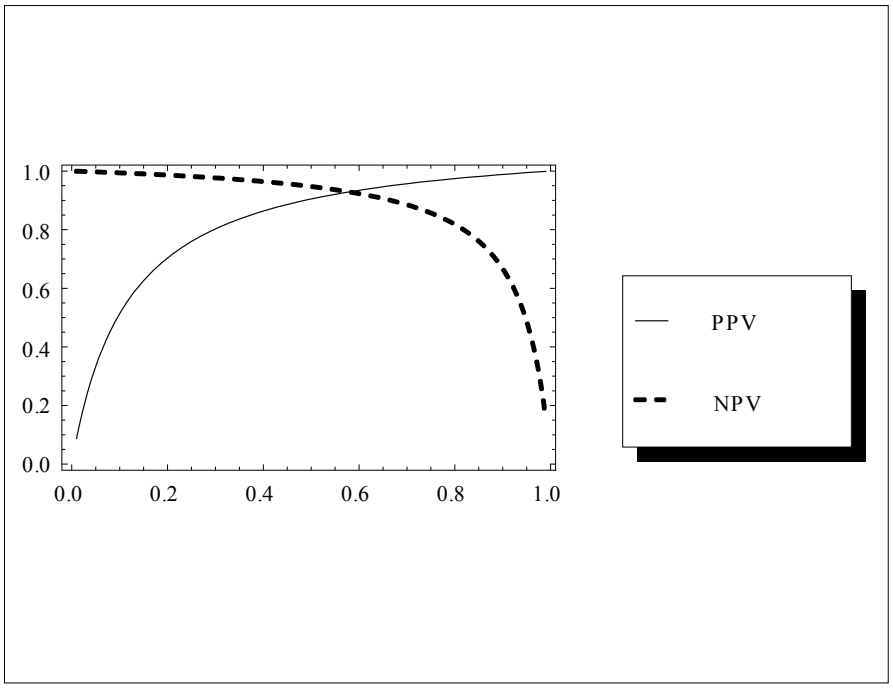
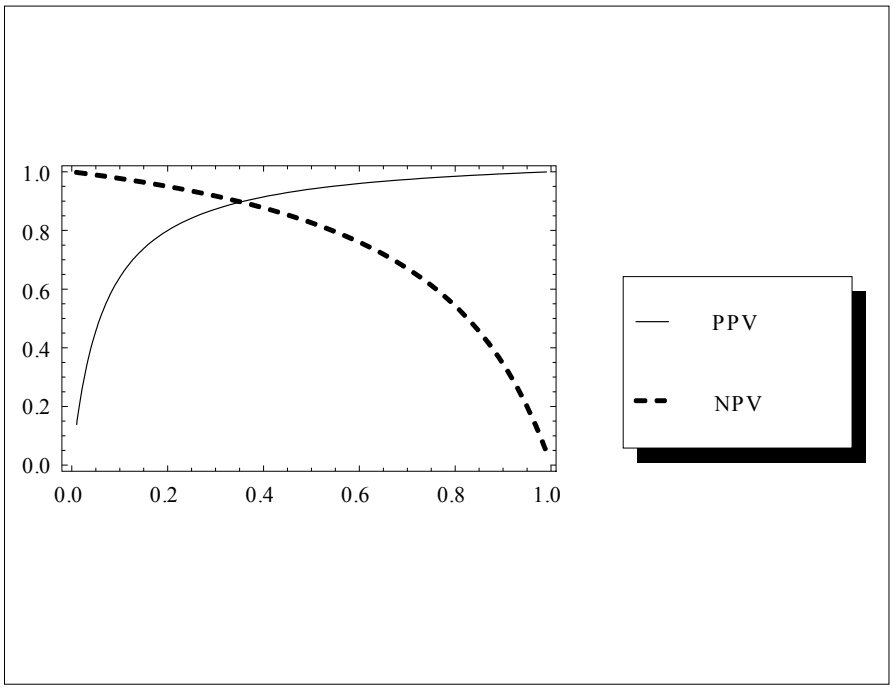


Figure 1.2. Predictive values for $Se = 0.80$ and $Sp = 0.95$.



1.5. Weighted kappa coefficient

The weighted kappa coefficient (Kraemer, 1992; Kraemer et al, 2002) of a *BDT* is defined as a measure of the beyond-chance agreement between the *BDT* and the *GS*, and it is a parameter that considers the losses associated with an erroneous classification with the *BDT*. Let L be the loss or the cost that occurs when wrongly classifying a diseased individual with the *BDT*, and L' the loss or cost that occurs when wrongly classifying a non-diseased individual with the *BDT*. It is assumed that the losses L and L' are 0 when an individual (diseased or non-diseased) is classified correctly with the *BDT*. Table 1.1 shows the probabilities and the losses associated to the assessment of the *BDT*, where the random variables T and D are defined in Section 1.2. In terms of the probabilities of Table 1.1, the expected loss is

$$p(1-Se)L+q(1-Sp)L', \tag{1.7}$$

and the random loss is

$$p\{p(1-Se)+qSp\}L+q\{pSe+q(1-Sp)\}L'. \tag{1.8}$$

The expected loss given by equation (1.7) is described by Bloch (1998) as the risk of error and is interpreted as the average loss which occurs when the *BDT* erroneously classifies an individual (diseased or non-diseased) and its values vary between zero and infinite. The random loss given by equation (1.8) is the expected loss when the *BDT* and the *GS* are independent, i.e. it is the expected loss when $P(T=i|D=j)=P(T=i)$ with $i,j=0,1$. The weighted kappa coefficient of a *BDT* is defined as

$$\kappa=\frac{\text{Random loss}-\text{Expected loss}}{\text{Random loss}} \tag{1.9}$$

Substituting in equation (1.9) each loss with its expression it holds that the weighted kappa coefficient of the *BDT* is

$$\kappa(c) = \frac{pqY}{p(1-Q)c + qQ(1-c)}, \tag{1.10}$$

where $Q = pSe + (1-p)(1-Sp)$ is the probability that the result of the *BDT* will be positive, $1-Q = p(1-Se) + qSp$ is the probability that the result of the *BDT* will be negative, and $c = L/(L+L')$ is the weighting index.

Table 1.1. Probabilities and losses.

Probabilities			
	$T=1$	$T=0$	Total
$D=1$	$p \times Se$	$p \times (1-Se)$	p
$D=0$	$q \times (1-Sp)$	$q \times Sp$	q
Total	$Q = p \times Se + q \times (1-Sp)$	$1-Q = p \times (1-Se) + q \times Sp$	1
Losses			
	$T=1$	$T=0$	Total
$D=1$	0	L	L
$D=0$	L'	0	L'
Total	L'	L	$L+L'$

When the loss L is zero then $c=0$ and the weighted kappa coefficient is

$$\kappa(0) = \frac{Sp-(1-Q)}{Q} = \frac{PPV-p}{q}, \tag{1.11}$$

and when the loss L' is zero then $c=1$ and the weighted kappa coefficient is

$$\kappa(1)=\frac{Se-Q}{1-Q}=\frac{NPV-q}{p}. \tag{1.12}$$

The weighted kappa coefficient $\kappa(c)$ can be written in terms of p, Q, $\kappa(0)$ and $\kappa(1)$ as

$$\kappa(c)=\frac{p(1-Q)c\kappa(1)+qQ(1-c)\kappa(0)}{p(1-Q)c+qQ(1-c)}, \tag{1.13}$$

with $0\le c\le 1$, and therefore the weighted kappa coefficient is a weighted average of $\kappa(0)$ and $\kappa(1)$. Index c is between 0 and 1, and it is a clinician's judgment of the relative clinical cost of false positives and false negatives. For example, let us consider a diagnosis of colon cancer using a colonography as *BDT*. If the colonography is positive for an individual who does not have the cancer (false positive), a colonoscopy (*GS*) will be performed on the individual, and this will give a final negative diagnosis. The loss L' is determined from the economic costs of the diagnosis and also of factors such as risk, stress and anxiety caused to the individual. If the colonography is negative for an individual with the cancer (false negative), the individual may be diagnosed later. In this case the cancer may have spread and the possibility of the treatment being successful will be reduced. The loss L is determined from these considerations. Hence, the losses L and L' are not only measured in economic terms but also by factors such as risk and stress, for which reason the value of these losses cannot be determined in clinical practice. This is why the relative discrepancy between the losses L and L' is replaced by the relative discrepancy between the false positives and the false negatives. The value of the weighting index $c=L/(L'+L)$ may be supposed, depending on the considerations

taken into account by the clinician about the false positives and the false negatives. If the clinician is going to use the *BDT* as a screening test, then there is more concern about false negatives and the *c* index is greater than 0.5 $\left(0.5 < c \leq 1\right)$. If the clinician is going to use the *BDT* as a first step towards an risk treatment, then there is more concern about false positives and the *c* index is less than 0.5 $\left(0 \leq c < 0.5\right)$. Therefore, the value of the *c* index will depend on what are the clinical objectives for using the *BDT*. If in equation (1.13) $L = L'$, then $c = 0.5$ and $\kappa\left(0.5\right)$ is known as Cohen's kappa coefficient; if $L > L'$ then $0.5 < c \leq 1$, and if $L' > L$ then $0 \leq c < 0.5$. For example, if the clinician considers that the false positives are three times more important than the false negatives, then $c = 1/\left(3+1\right) = 0.25$; and if the clinician considers that the false negatives are four times more important than the false positives then $c = 4/\left(1+4\right) = 0.8$. The weighted kappa coefficient has the following properties:

a) When the agreement between the *BDT* and the *GS* is perfect $\left(Se = Sp = 1\right)$, the expected loss is zero and then $\kappa\left(c\right) = 1$.

b) When *Se* and *Sp* are complementary $\left(Se = 1 - Sp\right)$, i.e. when the diagnosis of the disease is made randomly, then $\kappa\left(c\right) = 0$.

c) If the random loss is greater than expected loss then $\kappa\left(c\right) > 0$, and if the expected loss is greater than random loss then $\kappa\left(c\right) < 0$. If $\kappa\left(c\right) < 0$ the results of the diagnosis are interchanged, $T = 1$ should be a negative result and $T = 0$ should be a positive result, and the analysis should be limited only to the positive values of the weighted kappa coefficient $\left(0 \leq \kappa\left(c\right) \leq 1\right)$.

d) The weighted kappa coefficient is a function of the c index which may be increasing (if $Q > p$) or decreasing (if $Q < p$), or it can be a constant function which is equal to the Youden index $(Se + Sp - 1)$ if $Q = p$.

Figures 1.3 and 1.4 show how the weighted kappa coefficient varies according to the weighting index c and the disease prevalence p, for high values of Se and Sp. Coefficient $\kappa(c)$ is an increasing or decreasing function, depending on p, in the weighting index c.

Figure 1.3. Weighted kappa coefficients for $Se = 0.95$, $Sp = 0.90$ and different values of prevalence *p*.

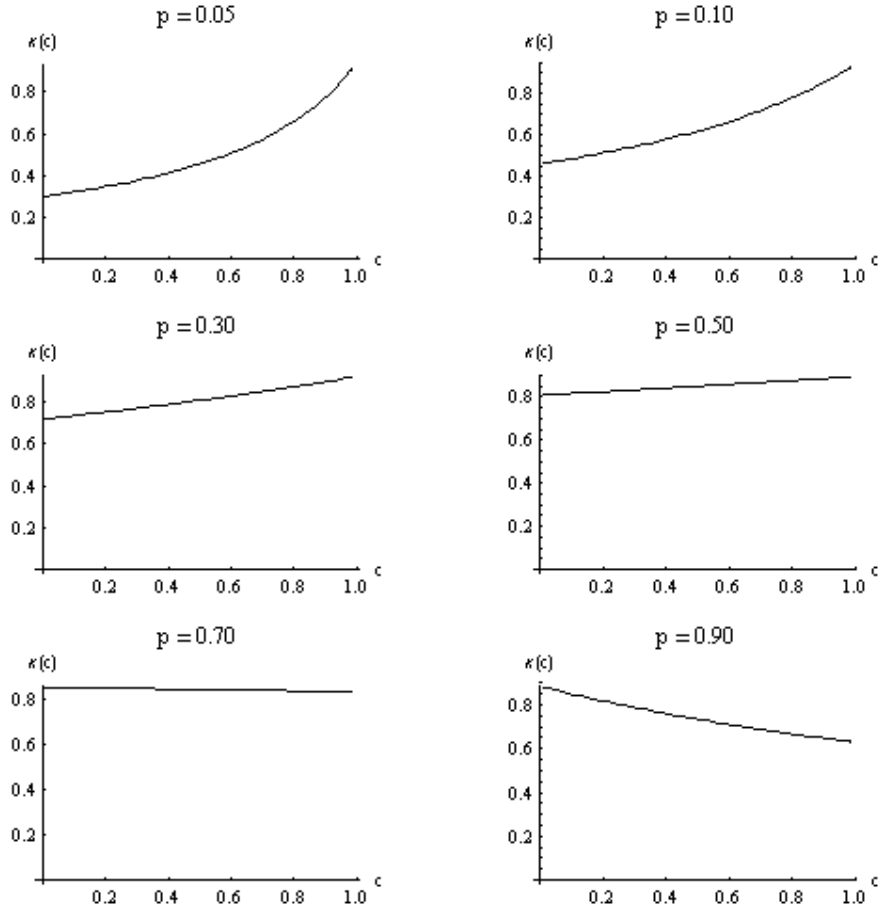
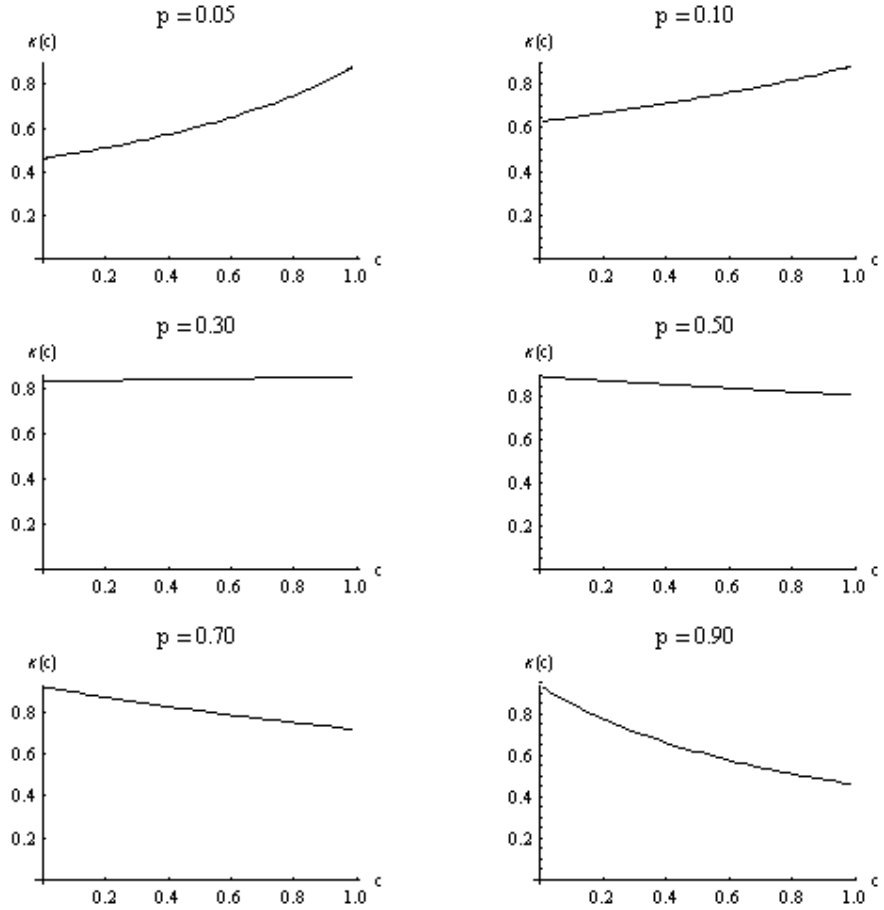


Figure 1.4. Weighted kappa coefficients for $Se = 0.90$, $Sp = 0.95$ and different values of prevalence *p*.



Chapter 2

Estimation of the Parameters of a Binary Diagnostic Test

2.1. Introduction

In this Chapter we study the estimation of the parameters of a *BDT* studied in Chapter 2 subject to two types of study: cross-sectional study and case-control study. The estimation of the weighted kappa coefficient subject to a case-control study is the first contribution to this Doctoral Thesis. For the other parameters, different methods of estimation are reviewed. The advantages and disadvantages of case-control design over the cross-sectional can be seen in the book of Pepe (2003). Summarizing, case-control study has some advantages over the cross-sectional study:

a) Case-control study is more efficient in terms of sample size requirements.

b) Case-control study allow for the exploration of subject-related characteristics on the test.

Nevertheless, the case-control study has the disadvantage that by using it we cannot estimate the prevalence of the disease *p*. Therefore, if in a case-control study we wish to estimate parameters that depend on the prevalence, it is necessary to have a value for the prevalence, e.g. an estimation obtained from health surveys, other studies, etc.

First, we study the estimation subject to a cross-sectional study, and secondly we study the estimation subject to a case-control study.

2.2. Estimation subject to a cross-sectional study

The estimation of the parameters of a *BDT* in relation to a *GS* subject to a cross-sectional study consists of applying the *BDT* and the *GS* to all of the individuals in a random sample sized *n*, giving rise to Table 2.1.

Table 2.1. Frequencies subject to a cross-sectional study.

	$T=1$	$T=0$	Total
$D=1$	s_1	s_0	s
$D=0$	r_1	r_0	r
Total	s_1+r_1	s_0+r_0	n

2.2.1. Sensitivity and specificity

Conditioning in variable *D*, the samples (s_1, s_0) and (r_1, r_0) are two independent samples, and it is verified that the observed frequency s_1 is the product of a binomial distribution $B(s, Se)$, and the frequency r_0 is the product of a binomial distribution $B(r, Sp)$. Therefore, the estimators of *Se* and *Sp* are the estimators of binomial proportions, i.e.

$$\hat{Se} = \frac{s_1}{s} \text{ and } \hat{Sp} = \frac{r_0}{r}, \tag{2.1}$$

and the estimators of their variances are

$$\hat{Var}\left(\hat{Se}\right) = \frac{\hat{Se}\left(1-\hat{Se}\right)}{s} \text{ and } \hat{Var}\left(\hat{Sp}\right) = \frac{\hat{Sp}\left(1-\hat{Sp}\right)}{r}. \tag{2.2}$$

Subject to a a cross-sectional study, the estimator of the prevalence of the disease *p* is

$$\hat{p} = \frac{s}{n}. \tag{2.3}$$

The estimation of a binomial proportion has been the object of many studies. We present five confidence intervals (*CIs*) for *Se* and *Sp* : Clopper-Pearsson, Wilson, Agresti-Coull, Yu et al, and the arcsine *CI* of Martín-Andrés and Álvarez-Hernández. The first *CI* is an exact interval and the other *CIs* are approximate intervals that have a good asymptotic behaviour.

2.2.1.1. Clopper-Pearson CI

Clopper and Pearson (1934) studied an exact *CI* for a binomial proportion. The $100(1-\alpha)\%$ Clopper-Pearson *CI* for *Se* is

$$\left(\frac{s_1}{s_1 + (s_0+1)F_{\alpha/2}\left(2(s_0+1);2s_1\right)} , \frac{(s_1+1)F_{\alpha/2}\left(2(s_1+1);2s_0\right)}{s_0 + (s_1+1)F_{\alpha/2}\left(2(s_1+1);2s_0\right)} \right), \tag{2.4}$$

and for *Sp* the *CI* is

$$\left(\frac{r_0}{r_0 + (r_1+1)F_{\alpha/2}\left(2(r_1+1),2r_0\right)} , \frac{(r_0+1)F_{\alpha/2}\left(2(r_0+1),2r_1\right)}{r_1 + (r_0+1)F_{\alpha/2}\left(2(r_0+1),2r_1\right)} \right), \tag{2.5}$$

where $F_{\alpha/2}(v_1, v_2)$ is the $\alpha/2$ quantile from an *F*-distribution with v_1 and v_2 degrees of freedom.

2.2.1.2. Wilson score CI

Wilson (1927) proposed, using the approximation to the normal distribution, an approximate *CI* for a binomial proportion. The Wilson *CI* is sometimes called the Wilson score *CI*. The $100(1-\alpha)\%$ Wilson *CI* for *Se* is

$$\frac{s}{s+z_{1-\alpha/2}^2}\left(\hat{Se} + \frac{z_{1-\alpha/2}^2}{2s} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{Se}\left(1-\hat{Se}\right)}{s} + \frac{z_{1-\alpha/2}^2}{4s^2}} \right), \tag{2.6}$$

and for *Sp* is

$$\frac{r}{r+z_{1-\alpha/2}^2}\left(\hat{Sp} + \frac{z_{1-\alpha/2}^2}{2r} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{Sp}\left(1-\hat{Sp}\right)}{r} + \frac{z_{1-\alpha/2}^2}{4r^2}} \right), \tag{2.7}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution.

2.2.1.3. Agresti-Coull CI

The *CI* of Agresti and Coull (1998) is another approximate *CI* for a binomial proportion. The $100(1-\alpha)\%$ Agresti and Coull *CI* for *Se* is

$$\widehat{Se} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{Se}\left(1-\widehat{Se}\right)}{s+z_{1-\alpha/2}^2}}, \tag{2.8}$$

where $\widehat{Se}=\dfrac{s_1+z_{1-\alpha/2}^2/2}{s+z_{1-\alpha/2}^2}$ is the adjusted estimator of *Se*. The $100(1-\alpha)\%$ Agresti and Coull *CI* for *Sp* is

$$\widehat{Sp} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{Sp}\left(1-\widehat{Sp}\right)}{r+z_{1-\alpha/2}^2}}, \tag{2.9}$$

where $\widehat{Sp}=\dfrac{r_0+z_{1-\alpha/2}^2/2}{r+z_{1-\alpha/2}^2}$ is the adjusted estimator of *Sp*.

2.2.1.4. Modified score CI

Yu et al (2014) have proposed a *CI*, called modified score interval, based on a modification of the midpoint of the Wilson *CI* .The $100(1-\alpha)\%$ modified score *CI* for *Se* is

$$0.5+\frac{s+\frac{z_{1-\alpha/2}^4}{53}}{s+z_{1-\alpha/2}^2}\left(\hat{Se}-0.5\right)\pm\frac{z_{1-\alpha/2}}{s+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{s_1s_0}{s}}\,. \tag{2.10}$$

Similarly, the $100(1-\alpha)\%$ modified score *CI* for *Sp* is

$$0.5+\frac{r+\frac{z_{1-\alpha/2}^4}{53}}{r+z_{1-\alpha/2}^2}\left(\hat{Sp}-0.5\right)\pm\frac{z_{1-\alpha/2}}{r+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{r_0r_1}{r}}\,. \tag{2.11}$$

2.2.1.5. Arcsine CI

Martín-Andrés and Álvarez-Hernández (2014a) evaluated 29 approximate *CIs* (excluding the Yu et al *CI*) for a binomial proportions, recommending using the arcsine *CI* with continuity correction. The $100(1-\alpha)\%$ arcsine *CI* with continuity correction for *Se* is

$$\sin^2\left(\sin^{-1}\sqrt{\frac{s_1+0.5}{s+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4(s+1)}}\right), \tag{2.12}$$

and for *Sp*

$$\sin^2\left(\sin^{-1}\sqrt{\frac{r_0+0.5}{r+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4(r+1)}}\right). \tag{2.13}$$

Martín-Andrés and Álvarez-Hernández (2016) have compared the performance of the 29 *CIs* studied, recommending:

a) For $s\le 80$ ($r\le 80$) and $\alpha=1\%$ or $\alpha=5\%$, to use the modified score *CI* of Yu et al (2014).

b) For $s \geq 100$ ($r \geq 100$) and $\alpha = 10\%$, to use the arcsine *CI* with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

c) In other situations, to use the *CI* of Agresti and Coull (1998).

2.2.2. Likelihood ratios

The estimators of the likelihood ratios are

$$\hat{LR}^{+} = \frac{\hat{Se}}{1-\hat{Sp}} = \frac{s_1}{r_1}\frac{r}{s} \text{ and } \hat{LR}^{-} = \frac{1-\hat{Se}}{\hat{Sp}} = \frac{s_0}{r_0}\frac{r}{s}. \quad (2.14)$$

Applying the delta method, the estimated asymptotic variances are

$$\hat{Var}\left(\hat{LR}^{+}\right) = \frac{\hat{LR}^{+}}{\left(1-\hat{Sp}\right)}\left(\frac{1-\hat{Se}}{s} + \frac{\hat{Sp}}{r}\hat{LR}^{+}\right) \text{ and } \hat{Var}\left(\hat{LR}^{-}\right) = \frac{\hat{LR}^{-}}{\hat{Sp}}\left(\frac{\hat{Se}}{s} + \frac{1-\hat{Sp}}{r}\hat{LR}^{-}\right). \quad (2.15)$$

The *LRs* are the ratio of the two independent binomial proportions, so that the *LRs* can be estimated applying methods to estimate the ratio of two independent binomial proportions. We now present several of these methods.

2.2.2.1. Gart-Nam CI

The Gart and Nam (1988) *CI* is a classic method to estimate the ratio of two independent binomial proportions. The $100(1-\alpha)\%$ *CI* of Gart and Nam for LR^{+} is obtained by solving the equation

$$\frac{s_1 - LR^+\left(1-\tilde{S}p\right)s}{\left\{1-LR^+\left(1-\tilde{S}p\right)\right\}\sqrt{\dfrac{1}{\dfrac{\tilde{S}p}{r\left(1-\tilde{S}p\right)}+\dfrac{1-LR^+\left(1-\tilde{S}p\right)}{LR^+\left(1-\tilde{S}p\right)s}}}} = \pm z_{1-\alpha/2}\,, \tag{2.16}$$

where $\tilde{S}p$ is the appropriate solution for the quadratic equation

$$LR^+\left(1-\tilde{S}p\right)^2 n - \left\{\left(r_1+s\right)LR^+ + s_1 + r\right\}\left(1-\tilde{S}p\right) + s_1 + r_1 = 0\,.$$

In a similar way, the $100(1-\alpha)\%$ *CI* for LR^- is obtained by solving the equation

$$\frac{s_0 - sLR^-\tilde{S}p}{\left(1-LR^-\tilde{S}p\right)\sqrt{\dfrac{1}{\dfrac{1-\tilde{S}p}{r\tilde{S}p}+\dfrac{1-LR^-\tilde{S}p}{sLR^-\tilde{S}p}}}}LR^- = \pm z_{1-\alpha/2}\,, \tag{2.17}$$

where $\tilde{S}p$ is the appropriate solution for the quadratic equation

$$nLR^-\tilde{S}p^2 - \left\{\left(r_0+s\right)LR^- + s_0 + r\right\}\tilde{S}p + s_0 + r_0 = 0\,.$$

2.2.2.2. Logarithmic CI

The *LRs* have a non-symmetrical distribution and therefore the napierian logarithm of the *LRs* has a more symmetrical distribution which is closer to the normal distribution. Thus, based on the asymptotic normality of the napierian logarithms of the $\hat{L}Rs$, Simel et al (1991) proposed the following *CIs* for *LRs*,

$$LR^+ \in \exp\left\{\ln\left(\hat{L}R^+\right) \pm z_{1-\alpha/2}\sqrt{\frac{s_0}{s_1 s}+\frac{r_0}{r_1 r}}\right\} \tag{2.18}$$

and

$$LR^{-} \in \exp\left\{\ln\left(\hat{L}R^{-}\right) \pm z_{1-\alpha/2}\sqrt{\frac{s_1}{s_0 s}+\frac{r_1}{r_0 r}}\right\}. \tag{2.19}$$

2.2.2.3. Martín-Andrés and Álvarez-Hernández CI

Martín-Andrés and Álvarez-Hernández (2014b) studied *CIs* for the ratio of two independent binomial proportions. The $100(1-\alpha)\%$ *CI* for the *LRs* are

$$LR^{+} \in \frac{n's_1'r_1' + \frac{z^2_{1-\alpha/2}}{2}\left(s's_1' + r'r_1' - 2s_1'r_1'\right) \pm z_{1-\alpha/2}\sqrt{n'^2 s_1'r_1'\left(s_1' + r_1' - n'\hat{p}_1'\hat{p}_2'\right) + \frac{z^2_{1-\alpha/2}}{4}\left(s's_1' - r'r_1'\right)^2}}{r_1'\left\{n's'\hat{p}_1' - z^2_{1-\alpha/2}\left(s' - r_1'\right)\right\}} \tag{2.20}$$

and

$$LR^{-} \in \frac{n's_0'r_0' + \frac{z^2_{1-\alpha/2}}{2}\left(s's_0' + r'r_0' - 2s_0'r_0'\right) \pm z_{1-\alpha/2}\sqrt{n'^2 s_0'r_0'\left(s_0' + r_0' - n'\hat{p}_3'\hat{p}_4'\right) + \frac{z^2_{1-\alpha/2}}{4}\left(s's_0' - r'r_0'\right)^2}}{r_0'\left\{n's'\hat{p}_3' - z^2_{1-\alpha/2}\left(s' - r_0'\right)\right\}} \tag{2.21}$$

where $s_i' = s_i + 0.5$, $r_i' = r_i + 0.5$, $s' = s_1' + s_0'$, $r' = r_1' + r_0'$, $n' = s' + r'$, $\hat{p}_1' = r_1'/r'$, $\hat{p}_2' = s_1'/s'$, $\hat{p}_3' = r_0'/r'$ and $\hat{p}_4' = s_0'/s'$. If the upper limit of the interval for the LR^{+} is lower than $s_1'/\left(n' - r_1'\right)$ or higher than $\hat{L}R^{+}$, then the upper limit of the interval is

$$\frac{1}{s'\left(\hat{p}_1'\right)^2 + z^2_{1-\alpha/2}}\left\{s_1'\hat{p}_1' + \frac{z^2_{1-\alpha/2}}{2} - z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4} + s_1'\left(\hat{p}_1' - \hat{p}_2'\right)}\right\},$$

and the upper limit of this interval is higher than $\left(n' - s_1'\right)/r_1'$ or lower than $\hat{L}R^{+}$, then the upper limit of the *CI* is

$$\frac{1}{r'\left(\hat{p}_1'\right)^2}\left\{r_1'\hat{p}_2'+\frac{z_{1-\alpha/2}^2}{2}+z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+r_1'\left(\hat{p}_2'-\hat{p}_1'\right)}\right\}.$$

Regarding the *CI* for LR^-, if the upper limit of this *CI* is lower than $s_0'\Big/\left(n'-r_0'\right)$ or higher than $\hat{LR}^-$, then the lower limit is

$$\frac{1}{s'\left(\hat{p}_3'\right)^2+z_{1-\alpha/2}^2}\left\{s_0'\hat{p}_3'+\frac{z_{1-\alpha/2}^2}{2}-z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+s_0'\left(\hat{p}_3'-\hat{p}_4'\right)}\right\},$$

and if the upper limit of this interval is higher than $\left(n'-s_0'\right)\Big/r_0'$ or lower than $\hat{LR}^-$, then the upper limit is

$$\frac{1}{r'\left(\hat{p}_3'\right)^2}\left\{r_0'\hat{p}_4'+\frac{z_{1-\alpha/2}^2}{2}+z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+r_0'\left(\hat{p}_4'-\hat{p}_3'\right)}\right\}.$$

The *CI* proposed by Martín-Andrés and Álvarez-Hernández is the interval that has a better asymptotic coverage.

2.2.3. Predictive Values

Conditioning in variable *T*, the samples $\left(s_1,r_1\right)$ and $\left(s_0,r_0\right)$ are two independent samples It is verified that the frequency s_1 is the product of a binomial distribution $B\left(s_1+r_1,PPV\right)$ and that the observed frequency r_0 is the realization of a binomial distribution $B\left(s_0+r_0,NPV\right)$. Therefore, the estimators of the *PV*s are the estimators of binomial proportions, i.e.

$$\hat{PPV} = \frac{s_1}{s_1 + r_1} \text{ and } \hat{NPV} = \frac{r_0}{s_0 + r_0}, \tag{2.22}$$

and the estimators of their variances are

$$\hat{Var}\left(\hat{PPV}\right) = \frac{\hat{PPV}\left(1 - \hat{PPV}\right)}{s_1 + r_1} \text{ and } \hat{Var}\left(\hat{NPV}\right) = \frac{\hat{NPV}\left(1 - \hat{NPV}\right)}{s_0 + r_0}. \tag{2.23}$$

The estimation through *CIs* of the *PVs* can be carried out by applying the same *CIs* that for a binomial proportion.

2.2.3.1. Clopper-Pearson CI

The $100(1-\alpha)\%$ Clopper-Pearson *CI* for *PPV* is

$$\left(\frac{s_1}{s_1 + (r_1 + 1) F_{\alpha/2}\left(2(r_1 + 1); 2s_1\right)}, \frac{(s_1 + 1) F_{\alpha/2}\left(2(s_1 + 1); 2r_1\right)}{r_1 + (s_1 + 1) F_{\alpha/2}\left(2(s_1 + 1); 2r_1\right)} \right), \tag{2.24}$$

and for *NPV* the *CI* is

$$\left(\frac{r_0}{r_0 + (s_0 + 1) F_{\alpha/2}\left(2(s_0 + 1), 2r_0\right)}, \frac{(r_0 + 1) F_{\alpha/2}\left(2(r_0 + 1), 2s_0\right)}{s_0 + (r_0 + 1) F_{\alpha/2}\left(2(r_0 + 1), 2s_0\right)} \right), \tag{2.25}$$

where $F_{\alpha/2}(v_1, v_2)$ is the $\alpha/2$ quantile from an *F*-distribution with v_1 and v_2 degrees of freedom.

2.2.3.2. Wilson score CI

The $100(1-\alpha)\%$ Wilson *CI* for *PPV* is

$$\frac{s_1+r_1}{s_1+r_1+z_{1-\alpha/2}^2}\left(P\hat{P}V+\frac{z_{1-\alpha/2}^2}{2(s_1+r_1)}\pm z_{1-\alpha/2}\sqrt{\frac{P\hat{P}V\left(1-P\hat{P}V\right)}{s_1+r_1}+\frac{z_{1-\alpha/2}^2}{4(s_1+r_1)^2}}\right), \quad (2.26)$$

and for *NPV* is

$$\frac{s_0+r_0}{s_0+r_0+z_{1-\alpha/2}^2}\left(N\hat{P}V+\frac{z_{1-\alpha/2}^2}{2(s_0+r_0)}\pm z_{1-\alpha/2}\sqrt{\frac{N\hat{P}V\left(1-N\hat{P}V\right)}{s_0+r_0}+\frac{z_{1-\alpha/2}^2}{4(s_0+r_0)^2}}\right). \quad (2.27)$$

2.2.3.3. Agresti-Coull CI

The $100(1-\alpha)\%$ Agresti and Coull *CI* for the *PPV*

$$P\tilde{P}V\pm z_{1-\alpha/2}\sqrt{\frac{P\tilde{P}V\left(1-P\tilde{P}V\right)}{s_1+r_1+z_{1-\alpha/2}^2}}, \quad (2.28)$$

where $P\tilde{P}V=\dfrac{s_1+z_{1-\alpha/2}^2/2}{s_1+r_1+z_{1-\alpha/2}^2}$ is the adjusted estimator of *PPV*. The $100(1-\alpha)\%$ Agresti and Coull *CI* for the *NPV*

$$N\tilde{P}V\pm z_{1-\alpha/2}\sqrt{\frac{N\tilde{P}V\left(1-N\tilde{P}V\right)}{s_0+r_0+z_{1-\alpha/2}^2}}, \quad (2.29)$$

where $NPV=\dfrac{r_0+z_{1-\alpha/2}^2/2}{s_0+r_0+z_{1-\alpha/2}^2}$ is the adjusted estimator of *NPV*.

2.2.3.4. Modified score CI

The $100(1-\alpha)\%$ modified score *CI* of Yu et al for *PPV* is

$$0.5+\frac{s_1+r_1+\frac{z_{1-\alpha/2}^4}{53}}{s_1+r_1+z_{1-\alpha/2}^2}\left(\hat{PPV}-0.5\right)\pm\frac{z_{1-\alpha/2}}{s_1+r_1+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{s_1r_1}{s_1+r_1}}\,, \tag{2.30}$$

and the $100(1-\alpha)\%$ modified score *CI* for *NPV* is

$$0.5+\frac{s_0+r_0+\frac{z_{1-\alpha/2}^4}{53}}{s_0+r_0+z_{1-\alpha/2}^2}\left(\hat{NPV}-0.5\right)\pm\frac{z_{1-\alpha/2}}{s_0+r_0+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{s_0r_0}{s_0+r_0}}\,. \tag{2.31}$$

2.2.3.5. Arcsine CI

The $100(1-\alpha)\%$ arcsine *CI* of Martín-Andrés and Álvarez-Hernández (2014a) for the *PPV* is

$$\sin^2\left(\sin^{-1}\sqrt{\frac{s_1+0.5}{s_1+r_1+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4\left(s_1+r_1+1\right)}}\right), \tag{2.32}$$

and for the *NPV*

$$\sin^2\left(\sin^{-1}\sqrt{\frac{r_0+0.5}{s_0+r_0+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4\left(s_0+r_0+1\right)}}\right). \tag{2.33}$$

Martín-Andrés and Álvarez-Hernández (2016) recommend:

a) For $s_1+r_1\le 80$ $\left(s_0+r_0\le 80\right)$ and $\alpha=1\%$ or $\alpha=5\%$, to use the modified score *CI* of Yu et al (2014).

b) For $s_1 + r_1 \geq 100$ $\left(s_0 + r_0 \geq 100\right)$ and $\alpha = 10\%$, to use the arcsine *CI* with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

c) In other situations, to use the *CI* of Agresti and Coull (1998).

2.2.4. Weighted kappa coefficient

Substituting in equation (1.10), or (1.13), the parameters *Se*, *Sp* and *p* with their estimators given by the equations (2.1) and (2.3) respectively, the estimator of $\kappa(c)$ is

$$\hat{\kappa}(c) = \frac{s_1 r_0 - s_0 r_1}{r\left(s_1 + r_1\right)(1-c) + s\left(s_0 + r_0\right)c}, \tag{2.34}$$

with $0 \leq c \leq 1$. Applying the delta method, the estimated asymptotic variance of $\hat{\kappa}(c)$ is

$$\begin{aligned}\hat{Var}\left(\hat{\kappa}(c)\right) = {} & \frac{nr}{s\left[n^2(1-c)r_1 + n\left(cr_0 - 2(1-c)r_1\right)s_0 + n\left\{r_0 - (1-c)r_1\right\}s_1 + s\left(s_0 r_1 - s_1 r_0\right)\right]^4} \times \\ & \left\{\left(s_0 r_1 - s_1 r_0\right)^2\left[2(1-c)r_1 ns - (1-c)r_1 n^2 + s\left(c\left(s_0 r_0 + 2s_0 r_1 + s_1 r_1\right) - r_1 s\right)\right]^2 + \right. \\ & \left. s_1 s_0 n r^3\left[(1-c)r_1 n + s\left(cr - r_1\right)\right]^2 + r_1 r_0 n s r^2\left[s_1 r + c\left(s^2 - s_1 n\right)\right]^2\right\}.\end{aligned} \tag{2.35}$$

Roldán-Nofuentes et al (2009) have studied different *CIs* for $\kappa(c)$: Wald *CI*, Logit *CI* and bootstrap *CI*.

2.2.4.1. Wald CI

Assuming the asymptotic normality of $\hat{\kappa}(c)$, the $100(1-\alpha)\%$ *CI* for $\kappa(c)$ is

$$\hat{\kappa}(c) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\hat{\kappa}(c)\right]}. \tag{2.36}$$

This *CI* performs well for relatively small samples $(n=100)$.

2.2.4.2. Logit CI

Assuming the asymptotic normality of $\hat{\kappa}(c)$, the logit transformation of $\hat{\kappa}(c)$, $\text{logit}\left[\hat{\kappa}(c)\right]=\log\left[\hat{\kappa}(c)/\left(1-\hat{\kappa}(c)\right)\right]$, is closer to a normal distribution with mean $\text{logit}\left[\kappa(c)\right]$. The $100(1-\alpha)\%$ *CI* for the $\text{logit}\left[\kappa(c)\right]$ is $\text{logit}\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left[\text{logit}\left(\hat{\kappa}(c)\right)\right]}$, where

$$\hat{V}ar\left[\text{logit}\left(\hat{\kappa}(c)\right)\right]=\frac{1}{\left[r_1 s-(1-c)r_1 n-c\left(s_0 r_0+2s_0 r_1+s_1 r_1\right)\right]^2}\times$$
$$\left\{\frac{s_1 s_0 r^2\left[(1-c)r_1 n+(cr-r_1)s\right]^2+r_1 r_0 sr\left[s_1 n-s_1 s+c\left(s^2-s_1 n\right)\right]^2}{s\left(s_0 r_1-s_1 r_0\right)^2}+\right. \qquad (2.37)$$
$$\left.\frac{\left[2(1-c)r_1 ns-(1-c)r_1 n^2+s\left(c\left(s_0 r_0+2s_0 r_1+s_1 r_1\right)-r_1 s\right)\right]}{nsr}\right\}.$$

Finally, the logit *CI* for $\kappa(c)$ is

$$\text{expit}\left\{\text{logit}\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left[\text{logit}\left(\hat{\kappa}(c)\right)\right]}\right\}, \qquad (2.38)$$

where $\text{expit}(\cdot)=\exp(\cdot)/\left[1+\exp(\cdot)\right]$ is the inverse of logit. This *CI* performs well for samples of 200 or more.

2.2.4.3. Bootstrap CI

The bootstrap *CI* is calculated generating *K* samples with replacement from the sample. The coefficient $\kappa(c)$ is estimate from each sample with replacement, and the bootstrap estimator of $\kappa(c)$ is estimated as the average of the *K* estimated $\hat{\kappa}_i(c)$, i.e. $\hat{\kappa}_B(c) = \sum_{i=1}^{K} \hat{\kappa}_i(c) \Big/ K$, and its variance is estimated as $\sum_{k=1}^{K} \left[\hat{\kappa}_i(c) - \hat{\kappa}_B(c)\right]^2 \Big/ (K-1)$. Then, the bias-corrected bootstrap *CI* (Efron and Tibshirani, 1993) is calculated. Let $A = \#\left(\hat{\kappa}_i(c) < \hat{\kappa}(c)\right)$ be the number of bootstrap estimators that are lower than the estimator, and let $\hat{z}_0 = \Phi^{-1}(A/K)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cumulative distribution function. Let $\alpha_1 = \Phi\left(2\hat{z}_0 - z_{1-\alpha/2}\right)$ and $\alpha_2 = \Phi\left(2\hat{z}_0 + z_{1-\alpha/2}\right)$, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)th$ percentile of the normal standard distribution, then the bias-corrected bootstrap *CI* is

$$\left(\hat{\kappa}_i^{(\alpha_1)}(c)\ ,\ \hat{\kappa}_i^{(\alpha_2)}(c)\right), \tag{2.39}$$

where $\hat{\kappa}_i^{(\alpha_j)}(c)$ is the *j*th quantile of the distribution of the *K* bootstrap estimations of $\kappa(c)$. In general, the performance of the bootstrap *CI* is similar to that of the Wald *CI*.

2.3. Estimation subject to a case-control study

The estimation of the parameters of a *BDT* subject to a case-control study consists of applying the *BDT* to two random independent samples, one of n_1 individuals who have the disease (sample of case) and another of n_0 individuals who do not have the disease

(sample of control). Let us suppose that of the n_1 individuals who have the disease, in n_{10} the *BDT* gives a negative result and in the rest $(n_{11} = n_1 - n_{10})$ the *BDT* gives a positive result. In the same way, let us suppose that of the n_0 individuals without the disease, in n_{00} the *BDT* gives a negative result and in the rest $(n_{01} = n_0 - n_{00})$ the *BDT* gives a positive result. The results are summarized in Table 2.2. The sample of individuals that have the disease is extracted from a population of individuals that have the disease (e.g. registers of diseases), and the control sample is extracted from a population of individuals who are known not to have the disease.

Table 2.2. Frequencies subject to case-control study.

	$T=1$	$T=0$	Total
Case	n_{11}	n_{10}	n_1
Control	n_{01}	n_{00}	n_0

In this situation, it is verified that the observed frequency n_{11} is the product of a binomial distribution $B(n_1, Se)$ and the observed frequency n_{00} is the product of a binomial distribution $B(n_0, Sp)$. In a case-control study, the quotient $n_1/(n_1 + n_0)$ is not an estimator of the disease prevalence because the sample sizes n_1 and n_0 are set by the researcher. Therefore, it is not possible to estimate the prevalence of the disease from a case-control study.

2.3.1. Sensitivity and specificity

The estimators of *Se* and *Sp* are the estimators of binomial proportions, i.e.

$$\hat{Se} = \frac{n_{11}}{n_1} \text{ and } \hat{Sp} = \frac{n_{00}}{n_0}, \tag{2.40}$$

and the estimators of their variances are

$$\hat{Var}\left(\hat{Se}\right) = \frac{\hat{Se}\left(1-\hat{Se}\right)}{n_1} \text{ and } \hat{Var}\left(\hat{Sp}\right) = \frac{\hat{Sp}\left(1-\hat{Sp}\right)}{n_0}. \tag{2.41}$$

The *CIs* for *Se* and *Sp* are the same that in Section 2.2.1 but replacing s_1 by n_{11}, s_0 by n_{10}, r_1 by n_{01}, r_0 by n_{00}, s by n_1 and r by n_0.

2.3.1.1. Clopper-Pearson CI

The $100(1-\alpha)\%$ Clopper-Pearson *CI* (1934) for *Se* is

$$\left(\frac{n_{11}}{n_{11} + (n_{10}+1) F_{\alpha/2}\left(2(n_{10}+1); 2n_{11}\right)}, \frac{(n_{11}+1) F_{\alpha/2}\left(2(n_{11}+1); 2n_{10}\right)}{n_{10} + (n_{11}+1) F_{\alpha/2}\left(2(n_{11}+1); 2n_{10}\right)} \right), \tag{2.42}$$

and for *Sp* the *CI* is

$$\left(\frac{n_{00}}{n_{00} + (n_{01}+1) F_{\alpha/2}\left(2(n_{01}+1), 2n_{00}\right)}, \frac{(n_{00}+1) F_{\alpha/2}\left(2(n_{00}+1), 2n_{01}\right)}{n_{01} + (n_{00}+1) F_{\alpha/2}\left(2(n_{00}+1), 2n_{01}\right)} \right). \tag{2.43}$$

2.3.1.2. Wilson score CI

The $100(1-\alpha)\%$ Wilson *CI* (1927) for *Se* is

$$\frac{n_1}{n_1+z_{1-\alpha/2}^2}\left(\hat{Se}+\frac{z_{1-\alpha/2}^2}{2n_1}\pm z_{1-\alpha/2}\sqrt{\frac{\hat{Se}\left(1-\hat{Se}\right)}{n_1}+\frac{z_{1-\alpha/2}^2}{4n_1^2}}\right), \tag{2.44}$$

and for *Sp*

$$\frac{n_0}{n_0+z_{1-\alpha/2}^2}\left(\hat{Sp}+\frac{z_{1-\alpha/2}^2}{2n_0}\pm z_{1-\alpha/2}\sqrt{\frac{\hat{Sp}\left(1-\hat{Sp}\right)}{n_0}+\frac{z_{1-\alpha/2}^2}{4n_0^2}}\right). \tag{2.45}$$

2.3.1.3. Agresti-Coull CI

The $100(1-\alpha)\%$ Agresti and Coull (1998) *CI* for *Se* is

$$\widetilde{Se}\pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{Se}\left(1-\widetilde{Se}\right)}{n_1+z_{1-\alpha/2}^2}}, \tag{2.46}$$

where $\widetilde{Se}=\dfrac{n_{11}+z_{1-\alpha/2}^2/2}{n_1+z_{1-\alpha/2}^2}$ is the adjusted estimator of *Se*. The *CI* for *Sp* is

$$\widetilde{Sp}\pm z_{1-\alpha/2}\sqrt{\frac{\widetilde{Sp}\left(1-\widetilde{Sp}\right)}{n_0+z_{1-\alpha/2}^2}}, \tag{2.47}$$

where $\widetilde{Sp}=\dfrac{n_{00}+z_{1-\alpha/2}^2/2}{n_0+z_{1-\alpha/2}^2}$ is the adjusted estimator of *Sp*.

2.3.1.4. Modified score CI

The $100(1-\alpha)\%$ modified score *CI* of Yu et al (2014) for *Se* is

$$0.5+\frac{n_1+\frac{z_{1-\alpha/2}^4}{53}}{n_1+z_{1-\alpha/2}^2}\left(\hat{Se}-0.5\right)\pm\frac{z_{1-\alpha/2}}{n_1+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{n_{11}n_{10}}{n_1}}\,, \tag{2.48}$$

and for *Sp*

$$0.5+\frac{n_0+\frac{z_{1-\alpha/2}^4}{53}}{n_0+z_{1-\alpha/2}^2}\left(\hat{Sp}-0.5\right)\pm\frac{z_{1-\alpha/2}}{n_0+z_{1-\alpha/2}^2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+\frac{n_{01}n_{00}}{n_0}}\,. \tag{2.49}$$

2.3.1.5. Arcsine CI

The $100(1-\alpha)\%$ arcsine *CI* with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a) for *Se* is

$$\sin^2\left(\sin^{-1}\sqrt{\frac{n_{11}+0.5}{n_1+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4(n_1+1)}}\right), \tag{2.50}$$

and for *Sp*

$$\sin^2\left(\sin^{-1}\sqrt{\frac{n_{00}+0.5}{n_0+1}}\pm\frac{z_{1-\alpha/2}}{\sqrt{4(n_0+1)}}\right). \tag{2.51}$$

The recommendations are similar to the previous case (Martín-Andrés and Álvarez-Hernández, 2016):

d) For $n_1\le 80$ ($n_0\le 80$) and $\alpha=1\%$ or $\alpha=5\%$, to use the modified score *CI* of Yu et al (2014).

e) For $n_1 \geq 100$ ($n_0 \geq 100$) and $\alpha = 10\%$, to use the arcsine *CI* with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

f) In other situations, to use the *CI* of Agresti and Coull (1998).

2.3.2. Likelihood ratios

The estimators of the likelihood ratios are

$$\hat{LR}^+ = \frac{\hat{Se}}{1-\hat{Sp}} = \frac{n_0 n_{11}}{n_1 n_{01}} \text{ and } \hat{LR}^- = \frac{1-\hat{Se}}{\hat{Sp}} = \frac{n_0 n_{10}}{n_1 n_{00}}, \tag{2.52}$$

and applying the delta method the estimated asymptotic variances are

$$\hat{Var}\left(\hat{LR}^+\right) = \frac{\hat{LR}^+}{\left(1-\hat{Sp}\right)}\left(\frac{1-\hat{Se}}{n_1} + \frac{\hat{Sp}}{n_0}\hat{LR}^+\right) \text{ and } \hat{Var}\left(\hat{LR}^-\right) = \frac{\hat{LR}^-}{\hat{Sp}}\left(\frac{\hat{Se}}{n_1} + \frac{1-\hat{Sp}}{n_0}\hat{LR}^-\right). \tag{2.53}$$

The *CIs* for *LRs* are the same that in Section 2.2.2.

2.3.2.1. Gart-Nam CI

The $100(1-\alpha)\%$ *CI* of Gart and Nam (1988) for LR^+ is obtained by solving the equation

$$\frac{n_{11} - LR^+\left(1-\tilde{Sp}\right)n_1}{\left\{1-LR^+\left(1-\tilde{Sp}\right)\right\}\sqrt{\dfrac{1}{\dfrac{\tilde{Sp}}{n_0\left(1-\tilde{Sp}\right)} + \dfrac{1-LR^+\left(1-\tilde{Sp}\right)}{LR^+\left(1-\tilde{Sp}\right)n_1}}}} = \pm z_{1-\alpha/2}, \tag{2.54}$$

where $\tilde{Sp}$ is the appropriate solution for the quadratic equation

$$LR^{+}\left(1-\tilde{S}p\right)^{2}\left(n_{1}+n_{0}\right)-\left\{\left(n_{01}+n_{1}\right)LR^{+}+n_{11}+n_{0}\right\}\left(1-\tilde{S}p\right)+n_{11}+n_{01}=0\,.$$

The $100(1-\alpha)\%$ *CI* for LR^{-} is obtained by solving the equation

$$\frac{n_{10}-n_{1}LR^{-}\tilde{S}p}{\left(1-LR^{-}\tilde{S}p\right)\sqrt{\dfrac{1}{\dfrac{1-\tilde{S}p}{n_{0}\tilde{S}p}+\dfrac{1-LR^{-}\tilde{S}p}{n_{1}LR^{-}\tilde{S}p}}}}=\pm z_{1-\alpha/2}\,, \tag{2.55}$$

where $\tilde{S}p$ is the appropriate solution for the quadratic equation

$$\left(n_{1}+n_{0}\right)LR^{-}\tilde{S}p^{2}-\left\{\left(n_{00}+n_{1}\right)LR^{-}+n_{10}+n_{0}\right\}\tilde{S}p+n_{10}+n_{00}=0\,.$$

2.3.2.2. Logarithmic CI

The $100(1-\alpha)\%$ logarithmic *CI* (Simel et al, 1991) for *LRs* are

$$LR^{+}\in\exp\left\{\ln\left(\hat{L}R^{+}\right)\pm z_{1-\alpha/2}\sqrt{\frac{n_{10}}{n_{11}n_{1}}+\frac{n_{00}}{n_{01}n_{0}}}\right\} \tag{2.56}$$

and

$$LR^{-}\in\exp\left\{\ln\left(\hat{L}R^{-}\right)\pm z_{1-\alpha/2}\sqrt{\frac{n_{11}}{n_{10}n_{1}}+\frac{n_{01}}{n_{00}n_{0}}}\right\}. \tag{2.57}$$

2.3.2.3. Martín-Andrés and Álvarez-Hernández CI

The $100(1-\alpha)\%$ *CI* of Martín-Andrés and Álvarez-Hernández (2014b) for the *LRs* are

$$LR^{+}\in\frac{n'n'_{11}n'_{01}+\frac{z^{2}_{1-\alpha/2}}{2}\left(n'_{1}n'_{11}+n'_{0}n'_{01}-2n'_{11}n'_{01}\right)\pm z_{1-\alpha/2}\sqrt{n'^{2}n'_{11}n'_{01}\left(n'_{11}+n'_{01}-n'\hat{p}'_{1}\hat{p}'_{2}\right)+\frac{z^{2}_{1-\alpha/2}}{4}\left(n'_{1}n'_{11}-n'_{0}n'_{01}\right)^{2}}}{n'_{01}\left\{n'n'_{1}\hat{p}'_{1}-z^{2}_{1-\alpha/2}\left(n'_{1}-n'_{01}\right)\right\}}\,' \tag{2.58}$$

and

$$LR^{-} \in \frac{n' n'_{10} n'_{00} + \frac{z^2_{1-\alpha/2}}{2}\left(n'_1 n'_{10} + n'_0 n'_{00} - 2n'_{10} n'_{00}\right) \pm z_{1-\alpha/2}\sqrt{n'^2 n'_{10} n'_{00}\left(n'_{10} + n'_{00} - n' \hat{p}'_3 \hat{p}'_4\right) + \frac{z^2_{1-\alpha/2}}{4}\left(n'_1 n'_{10} - n'_0 n'_{00}\right)^2}}{n'_{00}\left\{n' n'_1 \hat{p}'_3 - z^2_{1-\alpha/2}\left(n'_1 - n'_{00}\right)\right\}}, \quad (2.59)$$

where $n'_{1i} = n_{1i} + 0.5$, $n'_{0i} = n_{0i} + 0.5$, $n'_1 = n'_{11} + n'_{10}$, $n'_0 = n'_{01} + n'_{00}$, $n' = n'_1 + n'_0$, $\hat{p}'_1 = n'_{01}/n'_0$, $\hat{p}'_2 = n'_{11}/n'_1$, $\hat{p}'_3 = n'_{00}/n'_0$ and $\hat{p}'_4 = n'_{10}/n'_1$. If the upper limit of the interval for the LR^{+} is lower than $n'_{11}/\left(n' - n'_{01}\right)$ or higher than $\hat{L}R^{+}$, then the upper limit of the interval is

$$\frac{1}{n'_1\left(\hat{p}'_1\right)^2 + z^2_{1-\alpha/2}}\left\{n'_{11}\hat{p}'_1 + \frac{z^2_{1-\alpha/2}}{2} - z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4} + n'_{11}\left(\hat{p}'_1 - \hat{p}'_2\right)}\right\},$$

and the upper limit of this interval is higher than $\left(n' - n'_{11}\right)/n'_{01}$ or lower than $\hat{L}R^{+}$, then the upper limit of the *CI* is

$$\frac{1}{n'_0\left(\hat{p}'_1\right)^2}\left\{n'_{01}\hat{p}'_2 + \frac{z^2_{1-\alpha/2}}{2} + z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4} + n'_{01}\left(\hat{p}'_2 - \hat{p}'_1\right)}\right\}.$$

Regarding the *CI* for LR^{-}, if the upper limit of this *CI* is lower than $n'_{10}/\left(n' - n'_{00}\right)$ or higher than $\hat{L}R^{-}$, then the lower limit is

$$\frac{1}{n'_1\left(\hat{p}'_3\right)^2 + z^2_{1-\alpha/2}}\left\{n'_{10}\hat{p}'_3 + \frac{z^2_{1-\alpha/2}}{2} - z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4} + n'_{10}\left(\hat{p}'_3 - \hat{p}'_4\right)}\right\},$$

and if the upper limit of this interval is higher than $\left(n' - n'_{10}\right)/n'_{00}$ or lower than $\hat{L}R^{-}$, then the upper limit is

$$\frac{1}{n_0'\left(\hat{p}_3'\right)^2}\left\{n_{00}'\hat{p}_4'+\frac{z_{1-\alpha/2}^2}{2}+z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4}+n_{00}'\left(\hat{p}_4'-\hat{p}_3'\right)}\right\}.$$

The *CI* proposed by Martín-Andrés and Álvarez-Hernández is the interval that has a better asymptotic coverage.

2.3.3. Predictive values

The *PVs* of a *BDT* depend on the *Se*, *Sp* and disease prevalence *p*. As the prevalence cannot be estimated from a case-control study, since the quotient $n_1/(n_1+n_0)$ is not an estimator of the prevalence, the *PVs* cannot be estimated form a case-control study. Therefore, in order to estimate the *PVs* is it necessary to know an estimation of the prevalence of the disease. From now on it is assumed that we have an estimation of the prevalence of the disease. This value can be obtained from other clinical studies, health surveys, etc. If p is an estimation of the disease prevalence, then the estimators of the *PVs* are

$$\hat{PPV}=\frac{pn_{11}n_1n_0}{pn_{11}n_1n_0+qn_{00}n_1^2} \text{ and } \hat{NPV}=\frac{qn_{00}n_1n_0}{pn_{11}n_0^2+qn_{00}n_1n_0}, \tag{2.60}$$

where $q=1-p$. Different *CIs* have been studied for the *PVs* subject to a case-control study. Mercaldo et al (2007) have studied four frequentist *CIs* for the *PVs* and Stamey and Holt (2010) have studied two Bayesian *CIs*.

2.3.3.1. Study of Mercaldo et al

Mercaldo et al. (2007) have studied four *CIs* for the *PVs* subject to a case-control study, a Wald type *CI* and a logit *CI*, in both cases using the classical estimators and the adjusted estimators of *Se* and *Sp*. The $100(1-\alpha)\%$ Wald *CIs* for the *PVs* are

$$PPV \in \hat{PPV} \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\hat{PPV}\right)} \text{ and } NPV \in \hat{NPV} \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\hat{NPV}\right)}, \quad (2.61)$$

where

$$\hat{Var}\left(\hat{PPV}\right) = \frac{\left[pq\left(1-\hat{Sp}\right)\right]^2 \frac{\hat{Se}\left(1-\hat{Se}\right)}{n_1} + \left[pq\hat{Se}\right]^2 \frac{\hat{Sp}\left(1-\hat{Sp}\right)}{n_0}}{\left[q\hat{Se} + p\left(1-\hat{Sp}\right)\right]^4} \quad (2.62)$$

and

$$\hat{Var}\left(\hat{NPV}\right) = \frac{\left[pq\hat{Sp}\right]^2 \frac{\hat{Se}\left(1-\hat{Se}\right)}{n_1} + \left[pq\left(1-\hat{Se}\right)\right]^2 \frac{\hat{Sp}\left(1-\hat{Sp}\right)}{n_0}}{\left[p\left(1-\hat{Se}\right) + q\hat{Sp}\right]^4}. \quad (2.63)$$

The $100(1-\alpha)\%$ Wald *CIs* for the *PVs* with the adjusted estimates of *Se* and *Sp* are

$$PPV \in \tilde{PPV} \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\tilde{PPV}\right)} \text{ and } NPV \in \tilde{NPV} \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\tilde{NPV}\right)}, \quad (2.64)$$

where

$$\tilde{PPV} = \frac{p\tilde{Se}}{p\tilde{Se} + q\left(1-\tilde{Sp}\right)} \text{ and } \tilde{NPV} = \frac{q\tilde{Sp}}{p\left(1-\tilde{Se}\right) + q\tilde{Sp}}, \quad (2.65)$$

$$\tilde{Se} = \frac{n_1\hat{Se} + \frac{z^2_{1-\alpha/2}}{2}}{\tilde{n}_1} \text{ and } \tilde{Sp} = \frac{n_2\hat{Sp} + \frac{z^2_{1-\alpha/2}}{2}}{\tilde{n}_2}, \quad (2.66)$$

with $\tilde{n}_i = n_i + z^2_{1-\alpha/2}$, and

$$\hat{Var}\left(P\tilde{P}V\right) = \frac{\left[pq\left(1-\tilde{S}p\right)\right]^2 \frac{\tilde{S}e\left(1-\tilde{S}e\right)}{\tilde{n}_1} + \left(pq\tilde{S}e\right)^2 \frac{\tilde{S}p\left(1-\tilde{S}p\right)}{\tilde{n}_0}}{\left[p\tilde{S}e + q\left(1-\tilde{S}p\right)\right]^4} \tag{2.67}$$

and

$$\hat{Var}\left(N\tilde{P}V\right) = \frac{\left(pq\tilde{S}p\right)^2 \frac{\tilde{S}e\left(1-\tilde{S}e\right)}{\tilde{n}_1} + \left[pq\left(1-\tilde{S}e\right)\right]^2 \frac{\tilde{S}p\left(1-\tilde{S}p\right)}{\tilde{n}_0}}{\left[p\left(1-\tilde{S}e\right) + q\tilde{S}p\right]^4}. \tag{2.68}$$

On the other hand, the $100(1-\alpha)\%$ *CI* for the logit of the *PVs* are

$$\text{logit}\left(PPV\right) \in \text{logit}\left(P\hat{P}V\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(P\hat{P}V\right)\right]} \tag{2.69}$$

and

$$\text{logit}\left(NPV\right) \in \text{logit}\left(N\hat{P}V\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(N\hat{P}V\right)\right]}, \tag{2.70}$$

where

$$\text{logit}\left(PPV\right) = \log\left(\frac{PPV}{1-PPV}\right) = \log\left|\frac{pSe}{q\left(1-Sp\right)}\right|$$

and

$$\text{logit}\left(NPV\right) = \log\left(\frac{NPV}{1-NPV}\right) = \log\left|\frac{qSp}{p\left(1-Se\right)}\right|.$$

Finally, the $100(1-\alpha)\%$ logit *CI* for the *PVs* are

$$VPP \in \text{expit}\left\{\text{logit}\left(\hat{PPV}\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\hat{PPV}\right)\right]}\right\} \tag{2.71}$$

and

$$NPV \in \text{expit}\left\{\text{logit}\left(\hat{NPV}\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\hat{NPV}\right)\right]}\right\}, \tag{2.72}$$

where $\text{expit}(\cdot) = \exp(\cdot)/\{1+\exp(\cdot)\}$ is the inverse of logit. The variances, obtained applying the delta method, are

$$\hat{Var}\left[\text{logit}\left(\hat{PPV}\right)\right] = \frac{1-\hat{Se}}{n_1\hat{Se}} + \frac{\hat{Sp}}{n_0\left(1-\hat{Sp}\right)} \tag{2.73}$$

and

$$\hat{Var}\left[\text{logit}\left(\hat{NPV}\right)\right] = \frac{\hat{Se}}{n_1\left(1-\hat{Se}\right)} + \frac{1-\hat{Sp}}{n_0\hat{Sp}}. \tag{2.74}$$

With the adjusted estimates of *Se* and *Sp*, the logit *CIs* are

$$VPP \in \text{expit}\left\{\text{logit}\left(\tilde{PPV}\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\tilde{PPV}\right)\right]}\right\} \tag{2.75}$$

and

$$NPV \in \text{expit}\left\{\text{logit}\left(\tilde{NPV}\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\tilde{NPV}\right)\right]}\right\}, \tag{2.76}$$

and where the variances are

$$\hat{Var}\left[\text{logit}\left(\hat{PPV}\right)\right] = \frac{1-\tilde{Se}}{\tilde{n}_1\tilde{Se}} + \frac{\tilde{Sp}}{\tilde{n}_0\left(1-\tilde{Sp}\right)} \tag{2.77}$$

and

$$\hat{Var}\left[\text{logit}\left(\hat{NPV}\right)\right]=\frac{\tilde{Se}}{\tilde{n}_1\left(1-\tilde{Se}\right)}+\frac{1-\tilde{Sp}}{\tilde{n}_0\tilde{Sp}}. \tag{2.78}$$

Mercaldo et al have compared the asymptotic coverage of these four *CIs*, and in general terms, they recommend to apply the logit *CI* not adjusted, except when the estimator of a *PV* is equal to 1 in which case they recommend to apply the adjusted logit *CI*.

2.3.3.2. Study of Stamey and Holt

Stamey and Holt (2010) have studied two Bayesian *CIs* for the *PVs* subject to a case-control study: a *CI* without assuming distribution for the prevalence, and another *CI* assuming a probability distribution for the prevalence. Stamey and Holt have also shown that the second *CI* has a better asymptotic behaviour than the first *CI*. A summary of this second *CI* is presented below.

The observed frequencies n_{11} and n_{00} are the product of binomial distributions, i.e.

$$n_{11}\to B\left(n_1,Se\right) \text{ and } n_{00}\to B\left(n_0,Sp\right). \tag{2.79}$$

For *Se* and *Sp*, conjugate beta prior distributions are proposed, i.e.

$$\hat{Se}\to Beta\left(\alpha_1,\beta_1\right) \text{ and } \hat{Sp}\to Beta\left(\alpha_2,\beta_2\right), \tag{2.80}$$

and therefore, their posterior distributions are

$$\hat{Se}\left|d\to Beta\left(x_{11}+\alpha_1,n_1-x_{11}+\beta_1\right)\right. \text{ and } \hat{Sp}\left|d\to Beta\left(x_{00}+\alpha_2,n_0-x_{00}+\beta_2\right)\right., \tag{2.81}$$

And for the prevalence, the model is

$$p\to Beta\left(\alpha_3,\beta_3\right) \text{ and } y\to B\left(n,p\right), \tag{2.82}$$

with posterior distribution

$$p|d \to Beta\left(y+\alpha_3, n-y+\beta_3\right), \tag{2.83}$$

where $d=\left(x_{11},x_{00},y,n_1,n_0,n\right)$, and the data $\left(y,n\right)$ are obtained from a survey or other study. Then, posterior distributions for *PVs* is approximated via Monte Carlo sampling by simulating *B* values from the posteriors distributions (2.81) and (2.83). In each iteration, the generated values of *Se*, *Sp* and *p* are plugged in to the equations

$$PPV^{(j)}=\frac{p^{(j)}Se^{(j)}}{p^{(j)}Se^{(j)}+q^{(j)}\left(1-Sp^{(j)}\right)} \text{ and } NPV^{(j)}=\frac{q^{(j)}Sp^{(j)}}{p^{(j)}\left(1-Se^{(j)}\right)+q^{(j)}Sp^{(j)}}, \tag{2.84}$$

with $q^{(j)}=1-p^{(j)}$. Finally, from the *B* values of *PVs*, *CIs* based on the quantiles are calculated for *PPV* and *NPV*. Simulation experiments have shown that the Bayesian *CI* has a better asymptotic behaviour than the Mercaldo et al *CI*.

2.3.4. Weighted kappa coefficient

Subject to case-control study, Jannarone et al (1987) and Kraemer and Bloch (1990) have studied the point estimation of the weighted kappa coefficient, assuming that we have an estimation of the prevalence obtained from another study. Roldán-Nofuentes et al (2009) have studied various *CIs* (Wald, logit and bootstrap) for the weighted kappa coefficient subject to a cross-sectional study. The results obtained by Roldán-Nofuentes et al (2009), summarized in Section 2.2.4, are not valid in a case-control study, because, as has been pointed out previously, the quotient $n_1/\left(n_1+n_0\right)$ it is not an estimator of the prevalence of the disease. The studies by Jannarone et al (1987) and Kraemer and Bloch (1990) focused on studying the point estimation of the weighted kappa coefficient and

deducing the variance of the estimator. Therefore, it is necessary to explore this topic in more depth, study the estimation by *CIs* and study a method to determine the sample sizes necessary to carry out the study.

In this Section, two objectives are studied: firstly, different approximate *CIs* for the weighted kappa coefficient subject to case-control study, and secondly we propose a method to determine the sample sizes (case sample and control sample) necessary to estimate the weighted kappa coefficient. For both objectives, it is necessary to know a value of the prevalence of the disease. In Section 2.3.4.1, different *CIs* are proposed to estimate this parameter subject to a case-control study. In Section 2.3.4.2, simulation experiments are carried out to study the asymptotic coverage of these *CIs*, giving some general rules of application. In Section 2.3.4.3, we present a method to determine the (case and control) sample sizes to estimate the weighted kappa coefficient with the precision required. In Section 2.3.4.4, the results are applied to a real example, and in Section 2.3.5.5 the results obtained are discussed.

2.3.4.1. Approximate CIs

The weighted kappa coefficient of a *BDT* is

$$\kappa(c) = \frac{pq(Se + Sp - 1)}{p\left[\, p(1-Se) + qSp \,\right]c + p\left[\, pSe + q(1-Sp) \,\right](1-c)}, \tag{2.85}$$

where the only unknown parameters are the *Se* and the *Sp*, since it is assumed that we have an estimator *p* of the disease prevalence. Substituting in this equation (2.85) each parameter with its estimator, the estimator of the weighted kappa coefficient is

$$\hat{\kappa}(c)=\frac{pq\left(n_{11}n_{00}-n_{10}n_{01}\right)}{p\left(pn_{10}n_{0}+qn_{00}n_{1}\right)c+q\left(pn_{11}n_{0}+qn_{01}n_{1}\right)(1-c)}, \tag{2.86}$$

where $q=1-p$. Applying the delta method the estimation of the variance of $\hat{\kappa}(c)$ is

$$\hat{Var}\left[\hat{\kappa}(c)\right]=\left(\frac{\hat{\kappa}(c)}{pq\hat{Y}}\right)^{2}\times \left[\left[pq-\hat{\kappa}(c)\left\{p(q-c)\right\}\right]^{2}\frac{n_{10}n_{11}}{n_{1}^{3}}+\left[pq+\hat{\kappa}(c)\left\{q(c-q)\right\}\right]^{2}\frac{n_{00}n_{01}}{n_{0}^{3}}\right]. \tag{2.87}$$

The demonstration can be seen in Appendix 2.1. We will now propose several *CIs* for the weighted kappa coefficient.

2.3.4.1.1. Wald CI

The Wald CI for a parameter is well known. Assuming that we know an estimation of the prevalence of the disease and that the researcher has set a value of the weighting index *c*, the estimator $\hat{\kappa}(c)$ is a function of $\hat{Se}$ and of $\hat{Sp}$. This situation is a particular case which is analysed by Koch et al (1977). Applying the results of Koch et al (1977), the distribution of $\hat{\kappa}(c)$ is asymptotically normal, i.e. $\hat{\kappa}(c)\xrightarrow{d}N\left(\kappa(c),Var\left[\hat{\kappa}(c)\right]\right)$. Based on the asymptotic normality, the $100(1-\alpha)\%$ *CI* for $\kappa(c)$ is

$$\hat{\kappa}(c)\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\hat{\kappa}(c)\right]} \tag{2.88}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution.

This *CI* can also be calculated from the observed frequencies increased by a certain quantity, such as 0.5, 2 or $z_{1-\alpha/2}^{2}/2$. This procedure is very frequent in the analysis of

2×2 tables and in the estimation of binomial proportions (for example, Anscombe, 1956; Agresti and Coull, 1998; Martín-Andrés and Álvarez-Hernández, 2014a) or combinations of them (Martín-Andrés and Álvarez-Hernández, 2014b). These corrections aim to improve the coverages of the *CIs*, especially when the samples are small. Let us $m_{1j}=n_{1j}+h$ and $m_{0j}=n_{0j}+h$, with h equal to 0.5, 2 or $z^2_{1-\alpha/2}/2$, and $j=0,1$. In this situation, the adjusted estimators of the *Se* and the *Sp* are $\widetilde{Se}=m_{11}/m_1$ and $\widetilde{Sp}=m_{00}/m_0$, and the estimated variances are $\hat{V}ar\left(\widetilde{Se}\right)=m_{10}m_{11}/m_1^3$ and $\hat{V}ar\left(\widetilde{Sp}\right)=m_{00}m_{01}/m_0^3$, with $m_i=n_i+2h$. Substituting in equations (2.86) and (2.87) each parameter with its adjusted estimator and each frequency n_{ij} and n_i with m_{ij} and m_i we obtain the adjusted estimators of the weighted kappa coefficient $\widetilde{\kappa}(c)$ and the adjusted Wald *CIs*. For $\alpha=5\%$ it holds that $h=2$ is approximately equal to the case $h=z^2_{1-\alpha/2}/2=1.96^2/2$. For $h=0$, the expressions (2.86), (2.87) and (2.88) are obtained.

2.3.4.1.2. Logit CI

As the value of $\kappa(c)$ is between 0 and 1, the logit transformation can be applied. Assuming the asymptotic normality of the logit of $\hat{\kappa}(c)$, i.e. $\text{logit}\left[\hat{\kappa}(c)\right]\xrightarrow{d}N\left(\text{logit}\left[\kappa(c)\right],Var\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\}\right)$, the $100(1-\alpha)\%$ *CI* for $\text{logit}\left[\kappa(c)\right]$ is

$$\text{logit}\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\}} \tag{2.89}$$

Applying the delta method it is obtained that

$$\hat{Var}\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\} = \frac{1}{\left[\hat{Y}\left\{c\left(q+p\hat{Y}-\hat{S}p\right)-q\left(1-\hat{S}p\right)\right\}\right]^2}\times$$
$$\left[\frac{n_{10}n_{11}\left\{c\left(q-\hat{S}p\right)-q\left(1-\hat{S}p\right)\right\}^2}{n_1^3}+\frac{n_{00}n_{01}\left\{c\left(\hat{S}e-p\right)-q\hat{S}e\right\}^2}{n_0^3}\right], \tag{2.90}$$

and undoing the logit transformation, the logit *CI* for $\text{t}\kappa(c)$ is

$$\text{expit}\left(\text{logit}\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\}}\right), \tag{2.91}$$

where $\text{expit}(\cdot)=\exp(\cdot)/\{1+\exp(\cdot)\}$ is the inverse of logit. This *CI* can also be calculated increasing the frequencies by a quantity of *h*, for which we substitute n_{ij}, n_i, $\hat{Se}$, $\hat{Sp}$ and $\hat{\kappa}_c$ with m_{ij}, m_i, $\widehat{Se}$, $\widehat{Sp}$ and $\widehat{\kappa_c}$ respectively for each value of *h*.

2.3.4.1.3. Arcsine CI

The arcsine transformation is a transformation that has been used to estimate a binomial proportion (Brown et al, 2001; Martín-Andrés and Álvarez-Hernández, 2014a) and is a transformation that stabilizes the variance. Performing the transformation $F\left[\kappa(c)\right]=\arcsin\left\lfloor\sqrt{\kappa(c)}\right\rfloor$ and assuming the asymptotic normality of this transformation, i.e. $\arcsin\sqrt{\hat{\kappa}(c)}\xrightarrow{d}N\left(\arcsin\sqrt{\kappa(c)},Var\left[\arcsin\sqrt{\hat{\kappa}(c)}\right]\right)$, the approximate *CI* for $F\left\lfloor\kappa(c)\right\rfloor$ is

$$F\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left\{F\left[\hat{\kappa}(c)\right]\right\}}, \tag{2.92}$$

where the variance is estimated applying the delta method and whose expression is

$$\hat{V}ar\left\{F\left[\hat{\kappa}(c)\right]\right\} = \frac{-pq}{4\hat{Y}\left\{c\left(q+p\hat{Y}-\hat{S}p\right)-q\left(1-\hat{S}p\right)\right\}\left[c\left\{1-\hat{S}p+p\left(\hat{Y}-1\right)\right\}-q\left(1-\hat{S}p+p\hat{Y}\right)\right]^2}\times \left[\frac{n_{10}n_{11}\left\{q\left(1-\hat{S}p\right)-c\left(\hat{S}p-q\right)\right\}^2}{n_1^3}+\frac{n_{00}n_{01}\left\{c\left(p-\hat{S}e\right)+q\hat{S}e\right\}^2}{n_0^3}\right]. \quad (2.93)$$

As the inverse function of $\arcsin\sqrt{\kappa(c)}$ is $\sin^2$, then the *CI* for $\kappa(c)$ is

$$\sin^2\left[F\left[\hat{\kappa}(c)\right]\pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left\{F\left[\hat{\kappa}(c)\right]\right\}}\right]. \quad (2.94)$$

As in the case of the logit *CI*, the arcsine *CI* can be calculated adding to the observed frequencies the quantity *h*, and the process is similar to that of the previous case.

2.3.4.1.4. Bootstrap CI

The bootstrap *CI* is calculated generating *K* samples with replacement from the case sample and another *K* samples with replacement based on the control sample. From each case sample with replacement we estimate $\hat{S}e_i$ and from each control sample with replacement we estimate $\hat{S}p_i$; then we estimate the weighted kappa coefficient as

$$\hat{\kappa}_i(c)=\frac{pq\left(\hat{S}e_i+\hat{S}p_i-1\right)}{p\left(1-\hat{Q}_i\right)c+q\hat{Q}_i\left(1-c\right)}, \quad (2.95)$$

with $i=1,...,K$. The bootstrap estimator of the weighted kappa coefficient is estimated as the average of the *K* estimated weighted kappa coefficients, i.e.

$$\hat{\kappa}_B(c) = \frac{\sum_{i=1}^{K} \hat{\kappa}_i(c)}{K}, \tag{2.96}$$

and its variance is estimated as $\sum_{k=1}^{K} \left[\hat{\kappa}_i(c) - \hat{\kappa}_B(c)\right]^2 / (K-1)$. We then calculate the bias-corrected bootstrap *CI* (Efron and Tibshirani, 1993) as will now be shown. Let $A = \#(\hat{\kappa}_{ci} < \hat{\kappa}_c)$ be the number of bootstrap estimators that are lower than the estimator, and let $\hat{z}_0 = \Phi^{-1}(A/K)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cumulative distribution function. Let $\alpha_1 = \Phi(2\hat{z}_0 - z_{1-\alpha/2})$ and $\alpha_2 = \Phi(2\hat{z}_0 + z_{1-\alpha/2})$, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the normal standard distribution, then the bias-corrected bootstrap *CI* is

$$\left(\hat{\kappa}_i^{(\alpha_1)}(c)\ ,\ \hat{\kappa}_i^{(\alpha_2)}(c)\right), \tag{2.97}$$

where $\hat{\kappa}_i^{(\alpha_j)}(c)$ is the *j*th quantile of the distribution of the *K* bootstrap estimations of $\kappa(c)$.

2.3.4.1.5. Bayesian CI

The *CIs* proposed in the previous sections are frequentist and are based on the asymptotic normality of the maximum likelihood estimator of $\kappa(c)$ (and of its logit and arcsine transformations). The problem is now approached from a Bayesian perspective. From the case sample it is obtained that $n_{11} \rightarrow Binomial(n_1, Se)$, and from the control sample $n_{00} \rightarrow Binomial(n_0, Sp)$. For sensitivity and specificity, conjugate beta prior

distributions are proposed, which are appropriate distributions for binomial proportions (as are $\hat{Se}$ and $\hat{Sp}$), i.e.

$$\hat{Se} \to Beta\left(\alpha_{Se}, \beta_{Se}\right), \tag{2.98}$$

and

$$\hat{Sp} \to Beta\left(\alpha_{Sp}, \beta_{Sp}\right). \tag{2.99}$$

For the prevalence of the disease (which is also a binomial proportion) we propose

$$p \to Beta\left(\alpha_{p}, \beta_{p}\right) \tag{2.100}$$

and

$$t \to Binomial\left(n^{*}, p\right), \tag{2.101}$$

where n^{*} is the sample size starting from which the prevalence has been estimated, and *t* is a binomial variable that represents the number of individuals with the disease among the n^{*}. The a posteriori distributions for $\hat{Se}$ and the $\hat{Sp}$ are

$$\hat{Se}\left|\left(n_{10}, n_{1}, n_{00}, n_{0}\right) \to Beta\left(n_{11} + \alpha_{Se}, n_{10} + \beta_{Se}\right)\right. \tag{2.102}$$

and

$$\hat{Sp}\left|\left(n_{10}, n_{1}, n_{00}, n_{0}\right) \to Beta\left(n_{00} + \alpha_{Sp}, n_{01} + \beta_{Se}\right).\right. \tag{2.103}$$

Considering the a priori distribution (2.100) and the data n^{*} and *t*, the a posteriori distribution for the prevalence is

$$p\left|\left(n_{10}, n_{1}, n_{00}, n_{0}, t, n^{*}\right) \to Beta\left(t + \alpha_{p}, n^{*} - t + \beta_{p}\right).\right. \tag{2.104}$$

Once all of the distributions are defined, the estimation of the weighted kappa coefficient is carried out applying the Monte Carlo method and calculating a *CI* based on quantiles. The algorithm for the Monte Carlo method is the following:

Step 1. Set the values of the parameters of the a priori distributions of $\hat{Se}$, $\hat{Sp}$ and p, i.e., set the values of α_{Se}, β_{Se}, α_{Sp}, β_{Sp}, α_p and β_p. For example, a non-informative distribution can be used, such as the distribution $Beta(1,1)$, for each a priori distribution.

Step 2. Generate the random number of each one of distributions (2.98), (2.99) and (2.100).

Step 3. Generate a random number of distribution (2.101) from the random number generated with distribution (2.100).

Step 4. Generate a random number of distributions (2.102), (2.103) and (2.104), using for this purpose the random values obtained in Steps 2 and 3.

Step 5. Calculate $\hat{\kappa}(c)$ with equation (2.85) using the random numbers obtained in Step 4. If it is verified that $\hat{Y} = \hat{Se} + \hat{Sp} - 1 \leq 0$, then we must go back to Step 1 (since for all *BDTs* we must demand that their Youden index be greater than 0 and, therefore, that $\hat{\kappa}(c) > 0$).

Step 6. Repeat *M* times Steps 1 to 5, obtaining *M* estimations of $\kappa(c)$.

Once this algorithm has been applied, from the *M* estimations of $\kappa(c)$ a *CI* based on the quantiles is calculated, i.e. the $100\times(1-\alpha)\%$ *CI* for $\kappa(c)$ is

$$\left(q_{\alpha/2}\ ,\ q_{1-\alpha/2}\right), \tag{2.105}$$

where q_γ is the γth quantile of the distribution of the *M* estimations of $\kappa(c)$. As the estimator of $\kappa(c)$, the average of the *M* estimations is calculated.

2.3.4.2. Simulation experiments

Simulation experiments were carried out to study the asymptotic coverage of the *CIs* proposed in previous Section 2.3.4.1. For this purpose, 10000 random samples were generated, both case samples and control samples, with sizes 25, 50, 75, 100 and 200, for case and control samples. The nine possible combinations with $\{n_1, n_0\} = \{25, 50, 75\}$ and the four possible combinations with $\{n_1, n_0\} = \{100, 200\}$ have been considered, and therefore 13 pairs of sample sizes have been considered. The case (control) samples were generated from binomial distributions with parameters n_i and *Se* (*Sp*). These random samples were generated from values of weighted kappa coefficients, and not setting the values of *Se* and of *Sp*, through the following method:

1) For the weighted kappa coefficients $\kappa(0)$ and $\kappa(1)$ we set the values $\{0.01,\ 0.02,...,\ 0.98,\ 0.99\}$.

2) For the weighting index c $(0 < c < 1)$ we set the values 0.1 and 0.9.

3) As the disease prevalence we took the values 10%, 25% and 50%.

4) Once the values of $\kappa(0)$, $\kappa(1)$, c, and of p were set, the values of *Se* and of *Sp* were calculated solving (through the Newton-Raphson method) the system formed by equations (1.11) and (1.12), i.e.

$$\kappa(0)=\frac{p\times Sp-p\times(1-Se)}{p\times Se+q\times(1-Sp)} \text{ and } \kappa(1)=\frac{q\times Se-q\times(1-Sp)}{p\times(1-Se)+q\times Sp}.$$

5) Finally, the value of the weighted coefficient was calculated with equation (1.13), i.e.

$$\kappa(c)=\frac{p(1-Q)c\kappa(1)+qQ(1-c)\kappa(0)}{p(1-Q)c+qQ(1-c)}.$$

Therefore, through this method the random samples were generated setting values for weighted kappa coefficients. We then calculated all of the *CIs* proposed in Section 2.3.4.1 to 95% $(\alpha=5\%)$, calculating the coverage probability and the average length of each one of them. Regarding the bootstrap *CI*, for each one of the 10000 case and control samples 5000 samples with replacement were generated. As for the Bayesian *CI*, for each one of the 10000 samples, another 10000 random samples were generated considering several scenarios. As a priori distributions for *Se* and *Sp*, non-informative distributions were considered, i.e. $Beta(1,1)$ distributions, for both parameters. For the prevalence of the disease, two a priori distributions were considered: $Beta(1,1)$ and $Beta(pn^*, qn^*)$. The first one corresponds to a non-informative distribution, and the second distribution is a beta distribution with an average p and one which uses information provided by the sample from which the estimation of the prevalence was

obtained. As a value of n^*, we took 500 and 5000, and thus we studied the effect of this sample size on the asymptotic behaviour of this *CI*.

The comparison of the performance of the *CIs* was made following criteria similar to those of Agreti and Coull (1998), Price and Bonnett (2004), and Martín-Andrés and Álvarez-Hernández (2014a, 2014b). These criteria consists of determining if the method "fails" for a confidence at 95%, which happens if the *CI* has a probability of coverage lower than or equal to 93% (in Appendix 2.2 this method is justified). The selection of the optimum *CI* was made with the following steps:

1) Choosing the *CIs* with the fewest failures (probability of coverage $>93\%$)

2) From the *CIs* with the fewest failures, choose those with the lowest average width (more accurate).

In Tables 2.3 and 2.4 we show the probabilities of coverage and the average lengths of the *CIs* when $\kappa(c)$ is equal to 0.1 and 0.9 (which are values close to the extremes) and for some sample sizes, indicating in each Table the values of *Se*, *Sp*, $\kappa(0)$, $\kappa(1)$, c and p with those with which we have calculated the value of $\kappa(c)$. Therefore, considering the classification of the values of the weighted kappa coefficient given by Cicchetti (2001), values of the weighted kappa coefficient were considered with different levels of clinical significance (if $0<\kappa_c<0.40$ the level of clinical significance is poor; if $0.40\leq\kappa_c\leq0.59$ the level of clinical significance is fair; if $0.60\leq\kappa_c<0.74$ the level of clinical significance is good; and if $0.75\leq\kappa_c<1$ the level of clinical significance is excellent). In these Tables, we indicate the failures (coverage probability $\leq 93\%$) in bold, and we also indicate the coverage probabilities and lengths for the 13

pairs of sample sizes considered. The results obtained for $h=2$ are not shown since they are very similar to those obtained for $h=z^2_{1-\alpha/2}/2$. For the Bayesian *CI* we only show the results obtained for $n^*=500$. From the results obtained in the scenarios considered and for the sample sizes studied, the following conclusions are obtained for each *CI*:

a) Wald *CI*. For $h=0$ and $h=0.5$ this interval does not fail, even when the value of $\kappa(c)$ is near to 0 or to 1. When the samples are small $(n_i \le 75)$ the Wald *CI* $h=0.5$ shows, in general terms, a better performance than when $h=0$; and for larger sample sizes $(n_i \ge 100)$ the value of *h* (0 or 0.5) has practically no effect on the asymptotic behaviour of this *CI*. The Wald *CI* with $h=z^2_{1-\alpha/2}/2$ (or $h=2$) has failures, especially when the samples are small $(n_i \le 75)$.

b) Logit *CI*. For $h=0$ this interval fails when $\kappa(c)$ is near to 0 $(\kappa(c)=0.1)$ and does not fail when $\kappa_c \ge 0.2$; for $h=0.5$ the interval does not fail when κ_c is not near to 0 or 1 $(0.2 \le \kappa(c) \le 0.8)$ and the sample sizes are large $(n_i \ge 100)$. The logit *CI* with $h=z^2_{1-\alpha/2}/2$ (or $h=2$) has failures especially when the sample sizes are small $(n_i \le 75)$.

c) Arcsine *CI*. The behaviour of the arcsine *CI* with $h=0$ is very similar to that of the logit *CI* with $h=0$. For $h=0.5$ this interval does not fail, and for $h=z^2_{1-\alpha/2}/2$ (or $h=2$) has failures when the samples are small $(n_i \le 75)$.

d) Bootstrap *CI*. The bias-corrected bootstrap *CI* fails when the value of $\kappa(c)$ is near to the extremes ($\kappa(c) \leq 0.2$ or $\kappa(c) \geq 0.8$) and the samples are small $(n_i \leq 75)$; for the rest of the values, in general terms this *CI* has a similar behaviour to that of Wald *CI*.

e) Bayesian *CI*. When the a priori distribution of *p* is a $Beta(1,1)$ this interval fails for moderate or very high values of $\kappa(c)$. When the a priori distribution of *p* is a $Beta(pn^*, qn^*)$, the Bayesian *CI* fails when the samples are small $(n_i \leq 75)$ and the value of $\kappa(c)$ is nearer to 1. For the rest of the values of $\kappa(c)$ the behaviour of this is generally similar to that of Wald *CI*. The performance of the Bayesian *CI* is better when we use the informative beta distribution, $p \rightarrow Beta(pn^*, qn^*)$, as when we use the non-informative one, $p \rightarrow Beta(1,1)$. This result was predictable, since the informative distribution introduces into the Bayesian model the information provided by the sample from which we estimate the disease prevalence; whereas the non-informative beta distribution is flat for all possible values of *p* and this distribution has a minimal impact on the a posteriori distribution of *p*. The performance of the Bayesian *CI* when $n^* = 5000$ is very similar to the case of $n^* = 500$, so that the sample size with which we can estimate the prevalence has practically no effect on the asymptotic behaviour of the Bayesian *CI*.

Only in terms of the sample sizes, we can give the following rules for the application of the *CIs*:

a) When the sample sizes are small $(n_i \leq 75)$, use the Wald *CI* with $h = 0.5$, as it is an interval that does not fail for the values of $\kappa(c)$ considered.

b) For other sample sizes $(n_i \geq 100)$ use the Wald *CI* with $h = 0$, also because it is the only interval that does not fail.

For sample sizes $n_i \geq 100$ we can also use the logit and arcsine *CIs* (with $h = 0$), bootstrap and Bayesian with $p \to Beta(pn^*, qn^*)$, although these last two require a greater computational effort than the rest, and the Wald one is the easiest to calculate.

Table 2.3. Coverage probabilities and lengths of the *CIs* for $\kappa(c)=0.1$.

$Se=0.55$ $Sp=0.55$ $p=50\%$ $c=0.1$ $\kappa_0=0.1$ $\kappa_1=0.1$							
Wald *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	0.959	0.434	0.966	0.424	0.981	0.397
25	50	0.953	0.357	0.959	0.351	0.977	0.337
50	25	0.970	0.394	0.975	0.385	0.983	0.363
50	50	0.961	0.319	0.964	0.315	0.974	0.305
75	75	0.973	0.264	0.973	0.262	0.978	0.256
100	100	0.968	0.235	0.969	0.233	0.972	0.229
200	200	0.981	0.179	0.981	0.179	0.983	0.178
Logit *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	**0.883**	0.648	**0.892**	0.646	**0.921**	0.642
25	50	**0.886**	0.594	**0.889**	0.594	**0.919**	0.593
50	25	**0.891**	0.631	**0.906**	0.627	**0.930**	0.619
50	50	**0.892**	0.523	**0.900**	0.522	**0.919**	0.520
75	75	**0.913**	0.451	**0.918**	0.450	0.931	0.449
100	100	0.934	0.397	0.938	0.397	0.942	0.396
200	200	0.943	0.252	0.947	0.252	0.952	0.251
Arcsine *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	**0.797**	0.480	1	0.472	**0.823**	0.450
25	50	**0.764**	0.401	1	0.397	**0.792**	0.384
50	25	**0.793**	0.439	1	0.432	**0.819**	0.411
50	50	**0.873**	0.360	1	0.356	**0.881**	0.347
75	75	**0.866**	0.298	1	0.296	**0.871**	0.291
100	100	**0.895**	0.262	1	0.261	**0.900**	0.257
200	200	0.957	0.191	1	0.190	0.962	0.189
Bootstrap *CI*				Bayesian *CI*			
				$p \to Beta(1,1)$		$p \to Beta(pn^*, qn^*)$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	**0.719**	0.476	1	0.413	0.967	0.405
25	50	**0.796**	0.365	1	0.354	0.962	0.334
50	25	**0.803**	0.452	1	0.379	0.964	0.382
50	50	**0.881**	0.340	1	0.315	0.971	0.307
75	75	**0.923**	0.277	1	0.274	0.972	0.256
100	100	0.932	0.244	1	0.246	0.968	0.229
200	200	0.977	0.180	1	0.202	0.973	0.175

Cov.: coverage probability. Leng.: length.

Table 2.4. Coverage probabilities and lengths of the *CIs* for $\kappa(c)=0.9$.

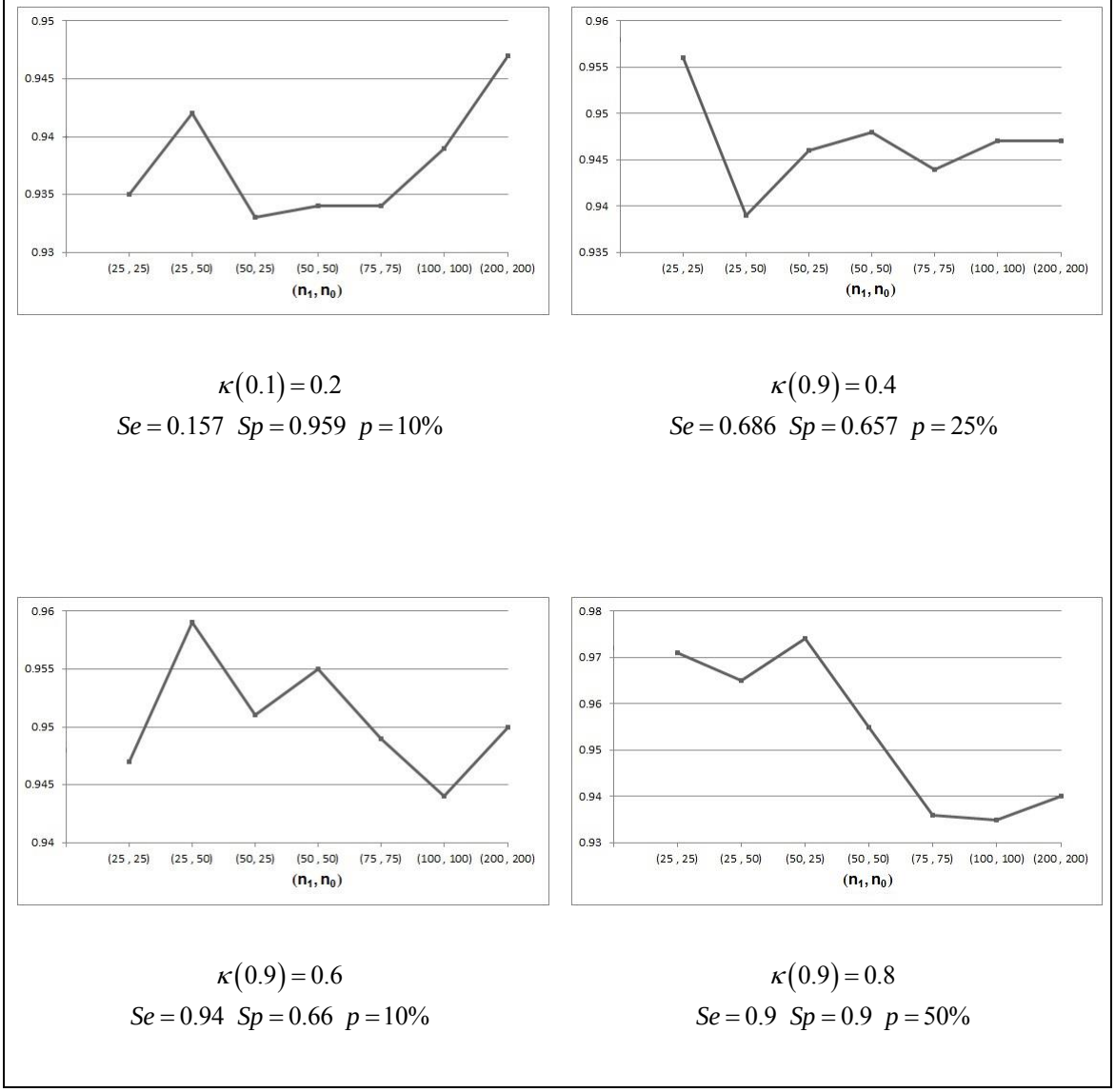
$Se=0.925$ $Sp=0.975$ $p=25\%$ $c=0.9$ $\kappa_0=0.9$ $\kappa_1=0.9$							
Wald *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	0.990	0.251	0.982	0.269	**0.868**	0.328
25	50	0.996	0.236	0.989	0.251	**0.914**	0.303
50	25	0.934	0.198	0.975	0.201	**0.881**	0.234
50	50	0.932	0.174	0.986	0.181	**0.912**	0.205
75	75	0.933	0.142	0.947	0.147	0.934	0.160
100	100	0.946	0.123	0.956	0.126	0.948	0.135
200	200	0.947	0.088	0.949	0.089	0.951	0.092
Logit *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	0.953	0.290	**0.875**	0.301	**0.578**	0.326
25	50	0.963	0.281	0.947	0.289	**0.665**	0.305
50	25	0.946	0.203	**0.886**	0.213	**0.617**	0.236
50	50	0.958	0.187	0.936	0.194	**0.798**	0.209
75	75	0.944	0.150	0.942	0.153	**0.848**	0.163
100	100	0.958	0.128	0.952	0.130	**0.886**	0.137
200	200	0.959	0.090	0.953	0.091	**0.913**	0.093
Arcsine *CI*							
		$h=0$		$h=0.5$		$h=z^2_{1-\alpha/2}/2$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	0.981	0.258	1	0.280	**0.669**	0.319
25	50	0.985	0.246	1	0.265	**0.829**	0.296
50	25	0.974	0.188	1	0.201	**0.773**	0.230
50	50	0.933	0.172	1	0.182	**0.867**	0.202
75	75	0.941	0.141	1	0.146	**0.890**	0.158
100	100	0.950	0.122	1	0.125	**0.923**	0.134
200	200	0.946	0.088	1	0.089	0.931	0.092
Bootstrap *CI*				Bayesian *CI*			
				$p \to Beta(1,1)$		$p \to Beta(pn^*, qn^*)$	
n_1	n_0	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
25	25	0.995	0.227	**0.821**	0.771	**0.867**	0.294
25	50	0.996	0.211	**0.834**	0.768	**0.923**	0.272
50	25	**0.926**	0.179	**0.777**	0.732	**0.893**	0.213
50	50	**0.904**	0.162	**0.861**	0.726	0.944	0.186
75	75	**0.907**	0.136	**0.845**	0.710	0.934	0.148
100	100	**0.924**	0.119	**0.878**	0.699	0.939	0.128
200	200	0.931	0.087	**0.867**	0.689	0.944	0.089

Cov.: coverage probability. Leng.: length.

In Figure 2.1, we show the probabilities of coverage of the Wald *CIs* (Wald with $h = 0.5$ when $n_i \leq 75$ and Wald with $h = 0$ when $n_i \geq 100$) obtained for values of $\kappa(c)$ equal to 0.2, 0.4, 0.6 and 0.8, and it can be observed that, for the sample sizes considered, these intervals do not fail.

Regarding the corrections $h = 2$ and $h = z^2_{1-\alpha/2}/2$, they do not improve the asymptotic behaviour of the *CIs*, and for small samples they even have a negative effect on the coverage probabilities of the *CIs*.

Figure 2.1. Coverage probabilities of the Wald *CIs*.



$\kappa(0.1)=0.2$
$Se=0.157 \;\; Sp=0.959 \;\; p=10\%$

$\kappa(0.9)=0.4$
$Se=0.686 \;\; Sp=0.657 \;\; p=25\%$

$\kappa(0.9)=0.6$
$Se=0.94 \;\; Sp=0.66 \;\; p=10\%$

$\kappa(0.9)=0.8$
$Se=0.9 \;\; Sp=0.9 \;\; p=50\%$

2.3.4.3. Sample sizes

A very important question in the study of statistical methods for diagnosis is the determination of the sample size necessary to be able to estimate a parameter of a *BDT* with a determined precision to a confidence $100(1-\alpha)\%$. Then, based on the Wald *CI*

(which is an interval that in general terms performs well, both with small samples and with large ones), a method is proposed to calculate sample sizes n_1 and n_0 needed to estimate κ_c with a precision δ and a confidence $100(1-\alpha)\%$. Based on the asymptotic normality of the estimator $\hat{\kappa}(c)$, it is verified that

$$\hat{\kappa}(c) \in \kappa(c) \pm z_{1-\alpha/2}\sqrt{Var\left[\hat{\kappa}(c)\right]}, \tag{2.106}$$

i.e. the probability of obtaining an estimator $\hat{\kappa}(c)$ is in interval (2.106) with a probability $100(1-\alpha)\%$. Let $\phi = n_0/n_1$. Substituting n_0 with $n_1\phi$ in equation (3.32) (see Appendix 2.1), and setting the value of ϕ and precision δ, then it is possible to calculate sample size n_1 using the following expression

$$\delta = z_{1-\alpha/2}\sqrt{Var\left[\hat{\kappa}(c)\right]}. \tag{2.107}$$

Clearing n_1 it holds that

$$n_1 = \left(\frac{z_{1-\alpha/2}\kappa(c)}{pqY\delta}\right)^2 \times \left[\left[pq-\kappa(c)\{p(q-c)\}\right]^2 Se(1-Se) + \left[pq+\kappa(c)\{q(c-q)\}\right]^2 \frac{Sp(1-Sp)}{\phi}\right]. \tag{2.108}$$

If an estimation of the disease prevalence is known and setting the value of the weighting index *c*, this method requires knowledge of the values of *Se*, *Sp* and $\kappa(c)$, i.e. it is necessary to know some estimators of these parameters (e.g. from pilot samples or other previous studies). The procedure would consist of the following steps:

Step 1. Take two pilot samples, one with n_1' individuals who have the disease and another one with n_0' individuals who do not have the disease, and based on these we

obtain $\hat{Se}$, $\hat{Sp}$, $\hat{\kappa}(c)$ and the Wald *CI*. If the Wald *CI* calculated has a precision δ, i.e. if $\frac{\text{Upper limit} - \text{Lower limit}}{2} \leq \delta$, then with the two pilot samples precision has been achieved and the process is finished; if precision is not achieved $\left(\frac{\text{Upper limit} - \text{Lower limit}}{2} > \delta\right)$ then we must go to the next step.

Step 2. From the values estimated in Step 1, calculate the new sample size n_1 with equation (2.108) (in this equation each parameter is substituted with its estimated value in Step 1) and calculate $n_0 = n_1\phi$.

Step 3. Take two samples, one composed of n_1 individuals with the disease and another composed of n_0 individuals without the disease (to samples n_i' from Step 1 individuals are added until the completion of new sample sizes n_i). From these new samples, $\hat{Se}$, $\hat{Sp}$, $\hat{\kappa}(c)$ and the Wald *CI* are calculated. If the Wald *CI* calculated has a precision δ then with the two new samples precision has been achieved and the process has finished. If the Wald *CI* does not have the desired precision, then consider these two samples to be pilot samples and go to Step 1.

The method propose to calculate the sample size is an iterative method that depends on the pilot samples and does not guarantee that with the sizes calculated $\kappa(c)$ is estimated with precision δ. Therefore, when applying this method it is necessary to check that with the sample sizes calculated the precision required is achieved.

This method can be applied considering $h = 0$ when the sizes of the pilot samples are ≥ 100 or with $h = 0.5$ when these sizes are lower than 100.

2.3.4.4. Example

The results obtained in the previous Sections were applied to the study by Patil et al (2013) on the assessment of ischaemia modified albumin in the diagnosis of acute coronary syndrome in India. This diagnostic test was applied to a sample of 102 individuals diagnosed with coronary disease and to a sample of 110 healthy control individuals. In Table 2.5 (Frequencies), the data from the study are shown (the frequencies were calculated from the results provided by the authors). In India, the prevalence of coronary heart disease has been estimated at a range of 1.6% to 7.4% in rural populations and from 1% to 13.2% in urban populations (Gupta et al, 2008). Taking into account the conclusions of the simulation experiments obtained in Section 2.3.4.2 and considering the prevalence value to be 5%, in Table 2.5 (Estimation of the weighted kappa coefficient) we can see the estimations of the weighted kappa coefficient for different values of the weighting index *c*. As for the Bayesian *CI*, two sample sizes have been supposed to estimate the prevalence: $n^* = 1000$ and $n^* = 5000$.

If the clinician considers that the false positives are more important than the false negatives (as is the case when ischaemia modified albumin is going to be used as a definitive test before a risky treatment, e.g. a surgical operation), then $0 \le c < 0.5$ and the weighted kappa coefficient has a level of clinical significance (Cicchetti, 2001) which varies mainly between 'poor' and 'fair' (in terms of the 95% CIs) depending on the *c* index. For example, if the clinician considers that $c = 0.1$, the false positives are 9 times more important than the false negatives, then the 95% Wald *CI* is $(0.21\,;\,0.55)$ and the level of clinical significance of ischaemia modified albumin is a value between 'poor' and 'fair' and, therefore, ischaemia modified albumin is not a useful test for the definitive diagnosis of coronary disease.

Table 2.5. Data from the study by Patil et al (2013) and estimations of $\kappa(c)$.

Frequencies				
	Negative test		Positive test	Total
Case	12		90	102
Control	102		8	110
Estimation of the weighted kappa coefficient				
	$\hat{\kappa}(c) \pm$ Standard error		95% Wald *CI*	
c	$h=0$	$h=0.5$	$h=0$	$h=0.5$
0.1	0.38 ± 0.087	0.37 ± 0.084	0.21 — 0.55	0.20 — 0.53
0.4	0.47 ± 0.089	0.45 ± 0.086	0.29 — 0.64	0.28 — 0.62
0.5	0.51 ± 0.087	0.49 ± 0.085	0.34 — 0.68	0.32 — 0.66
0.6	0.55 ± 0.084	0.54 ± 0.082	0.39 — 0.72	0.38 — 0.70
0.9	0.76 ± 0.050	0.75 ± 0.050	0.66 — 0.86	0.65 — 0.85
	95% Logit *CI*		95% Arcsine *CI*	
c	$h=0$	$h=0.5$	$h=0$	$h=0.5$
0.1	0.23 — 0.56	0.22 — 0.54	0.22 — 0.55	0.21 — 0.53
0.4	0.30 — 0.64	0.29 — 0.62	0.30 — 0.64	0.29 — 0.62
0.5	0.34 — 0.67	0.33 — 0.65	0.34 — 0.67	0.33 — 0.66
0.6	0.39 — 0.71	0.38 — 0.69	0.39 — 0.71	0.38 — 0.69
0.9	0.65 — 0.84	0.64 — 0.83	0.65 — 0.85	0.65 — 0.84
	95% Bayesian *CI*			
	$n^* = 1000$, $p \to Beta(pn^*, qn^*)$		$n^* = 5000$, $p \to Beta(pn^*, qn^*)$	
c	$\hat{\kappa}(c)$	95% *CI*	$\hat{\kappa}(c)$	95% *CI*
0.1	0.37	0.21 — 0.55	0.37	0.22 — 0.55
0.4	0.45	0.29 — 0.63	0.45	0.30 — 0.63
0.5	0.49	0.32 — 0.66	0.49	0.33 — 0.66
0.6	0.53	0.37 — 0.69	0.53	0.38 — 0.69
0.9	0.74	0.63 — 0.83	0.74	0.63 — 0.83

If the clinician considers that the false negatives are more important than the false positives (as is the case when ischaemia modified albumin is going to be used as a screening test), then $0.5 < c < 1$ and the level of clinical significance (Cicchetti, 2001) of the weighted kappa coefficient varies between 'fair' and 'excellent' (in terms of the 95% *CIs*) depending on the *c* index. For example, for $c = 0.9$, the false negatives are 9 times more important than the false positives, the 95% Wald *CI* is $(0.66\ ;\ 0.86)$ and the level of clinical significance of ischaemia modified albumin is a value between 'good' and 'excellent' and, therefore, this diagnostic test is very useful as a screening test.

As for the *CIs*, in this example as the sample sizes n_1 and n_0 are relatively large $(n_i > 100)$, all of them are very similar. Regarding the Bayesian *CI*, the sample size from which the prevalence is estimated does not have any effect upon the intervals obtained.

In order to illustrate the method described in Section 2.3.4.3 about the calculation of the sample sizes, let us consider that $c = 0.9$. In this situation, the 95% Wald *CI* with $h = 0$ is $(0.66\ ;\ 0.86)$ and the precision is 0.10 $\left(\dfrac{0.86 - 0.66}{2} = 0.10\right)$. As an example, let us consider that the clinician wishes to estimate the weighted kappa coefficient with a precision $\delta = 0.05$ and that the case sample and the control sample are the same size $(\phi = 1)$. As with the samples of 102 and 110 individuals, the desired precision $(\delta = 0.05 < 0.10)$ was not achieved, then using the two previous samples as pilot samples ($\hat{Se} = 0.88$, $\hat{Sp} = 0.93$ and $\hat{\kappa}(0.9) = 0.76$) applying equation (2.108) it is obtained that $n_1 = 227$ and $n_0 = \phi n_1 = 227$. Therefore, to the initial case sample it is necessary to add 125 individuals with the disease and 117 individuals without the

disease must be added to the control sample. Once these new samples have been taken, it is necessary to check that the desired precision is verified.

2.3.4.5. Discussion

The weighted kappa coefficient is a valid measure to assess and compare the performance of *BDTs*, and it depends on the sensitivity and the specificity of the test, on the prevalence of the disease and the relative importance between the false positives and the false negatives (weighting index). The sensitivity and the specificity of the test are easily estimated, both in cross-sectional studies and in case-control studies, since they are estimators of binomial proportions. Regarding the weighting index, this is set by the clinician depending on his or her knowledge of the problem in question. The estimation of the weighted kappa coefficient in case-control studies requires knowledge of an estimation of the prevalence of the disease based on another study (or from the literature, etc.), since from this design it is impossible to estimate the prevalence. In this article, we have studied the estimation of the weighted kappa coefficient of a single *BDT* subject to case-control study assuming that we have an estimation of the prevalence of the disease. Fourteen approximate *CIs* were studied and compared for this parameter (thirteen were frequentist and one was Bayesian). The results of the simulation experiments carried out demonstrated that, in very general terms, for small samples $\left(n_i \leq 75\right)$ it is possible to use the Wald *CI* with $h = 0.5$, and that for samples with $n_i \geq 100$ the Wald, logit, arcsine (with $h = 0$) and Bayesian (with $Beta\left(pn^*, qn^*\right)$) intervals can be used. The Bayesian *CI* performs well when for prevalence an a priori distribution $Beta\left(pn^*, qn^*\right)$ is considered, where n^* is the sample size from which the

prevalence has been estimated. If sample size n^* is not known, any value (e.g. 500 or 5000) can be used, since the simulation experiments have demonstrated that this value has practically no effect upon the coverage probability and length of the interval.

Another important question that has been studied is the calculation of the sample sizes n_1 and n_0 needed to estimate the weighted kappa coefficient with a determined precision. Therefore, and once we have set the values of index *c*, of the desired precision δ and of the relation between n_1 and n_0 (ϕ), a method based on the Wald *CI* has been proposed that requires knowledge of the sensitivity, the specificity and the weighted kappa coefficient of the *BDT* (e.g. from pilot samples or other studies). The method proposed depends on the auxiliary information (pilot samples or other studies) and, therefore, it does not guarantee the estimation of the weighted kappa coefficient with the desired precision, and thus it is necessary to check this condition once the new samples have been taken.

The estimation of the weighted kappa coefficient requires knowledge of an estimation of the prevalence obtained from another study e.g. a health survey. To study the effect of a misspecification of the prevalence in the estimation of the weighted kappa coefficient, we carried out simulation experiments similar to those made in Section 2.3.4.2. For this we took as the prevalence for the inference an overestimation (and underestimation) equal to 10% and to 20% of the value of the prevalence set, and we studied the asymptotic behaviour of the intervals recommended (Wald with $h = 0.5$ for small samples and Wald with $h = 0$ for large samples). In the Tables 2.6 and 2.7 we show some of the results obtained. The results demonstrated that slight (10%) and

moderate (20%) misspecifications of the disease prevalence do not have an important effect on the probabilities of coverage of the Wald *CIs*.

Table 2.6. Coverage probabilities and lengths of the Wald *CIs* for $\kappa(c) = 0.1$.

$Se = 0.55$ $Sp = 0.55$ $c = 0.1$ True prevalence = 50%					
		$p =$ True prevalence + $0.10 \times$ True prevalence = 55%		$p =$ True prevalence – $0.10 \times$ True prevalence = 45%	
n_1	n_0	Cov.	Leng.	Cov.	Leng.
25	25	0.962	0.443	0.977	0.399
25	50	0.957	0.374	0.974	0.333
50	25	0.974	0.402	0.982	0.361
50	50	0.955	0.331	0.973	0.295
75	75	0.965	0.278	0.979	0.246
100	100	0.961	0.249	0.979	0.220
200	200	0.965	0.190	0.977	0.167
		$p =$ True prevalence + $0.20 \times$ True prevalence = 60%		$p =$ True prevalence – $0.20 \times$ True prevalence = 40%	
n_1	n_0	Cov.	Leng.	Cov.	Leng.
25	25	0.953	0.460	0.987	0.371
25	50	0.943	0.391	0.979	0.308
50	25	0.969	0.417	0.987	0.336
50	50	0.947	0.345	0.984	0.273
75	75	0.958	0.291	0.987	0.227
100	100	0.957	0.260	0.983	0.202
200	200	0.956	0.199	0.963	0.154

Cov.: coverage probability. Leng.: length.

Table 2.7. Coverage probabilities and lengths of the Wald *CIs* for $\kappa(c) = 0.9$.

$Se = 0.925$ $Sp = 0.975$ $c = 0.9$ True prevalence = 25%					
		$p =$ True prevalence + $0.10 \times$ True prevalence = 27.5%		$p =$ True prevalence – $0.10 \times$ True prevalence = 22.5%	
n_1	n_0	Cov.	Leng.	Cov.	Leng.
25	25	0.982	0.287	0.981	0.280
25	50	0.991	0.269	0.991	0.259
50	25	0.979	0.205	0.979	0.203
50	50	0.986	0.188	0.986	0.182
75	75	0.941	0.150	0.947	0.144
100	100	0.944	0.125	0.933	0.120
200	200	0.941	0.090	0.939	0.086
		$p =$ True prevalence + $0.20 \times$ True prevalence = 30%		$p =$ True prevalence – $0.20 \times$ True prevalence = 20%	
n_1	n_0	Cov.	Leng.	Cov.	Leng.
25	25	0.982	0.292	0.981	0.278
25	50	0.986	0.275	0.991	0.255
50	25	0.962	0.208	0.979	0.204
50	50	0.974	0.192	0.986	0.179
75	75	0.960	0.154	0.947	0.142
100	100	0.944	0.129	0.934	0.118
200	200	0.948	0.092	0.933	0.084

Cov.: coverage probability. Leng.: length.

In this Section we have studied the estimation of the weighted kappa coefficient through *CIs*. Further research should be carried out when, subject to case-control study, a hypothesis test on this parameter is made i.e. $H_0 : \kappa_c = \kappa_{0c}$ vs $H_1 : \kappa_c \neq \kappa_{0c}$, proposing different methods to solve this hypothesis test and also studying the sample sizes necessary to solve the hypothesis test to an error α and to a power θ.

Chapter 3

Comparison and Combination of two Binary Diagnostic Tests

3.1 Introduction

This Chapter studies the comparison of parameters of two *BDTs* subject to a paired design and the combination of two *BDTs*. The comparison of parameters of two *BDTs* is an important topic in the field of Statistical Methods for Diagnosis in Medicine. In this Chapter, the hypothesis tests and the *CIs* to compare the parameters of two *BDTs* subject to a paired design are presented. In this Chapter we also study the combination of the parameters of two *BDTs*. In practice, it is common to combine two *BDTs* in order to increase the accuracy of the diagnosis of the disease. The combination of the sensitivities and of specificities, likelihoods ratios, predictive values and weighted kappa coefficient of two *BDTs* are presented. The combination of the weighted kappa

coefficients of two *BDTs* is the second contribution of this Thesis. Each one of these objectives will now be studied.

3.2. Comparison of two *BDTs*

Let us consider two *BDTs* whose performance is compared in relation to the same *GS*. A paired design consists of applying the two *BDTs* and the *GS* to all the individuals in a random sample sized *n*, and the most common type of sampling when we want to compare the performance of two *BDTs*. Let T_1 and T_2 be the random binary variables that model the results of *Test 1* and *Test 2* respectively, in such a way that $T_k = 1$ when the result of the *BDT* is positive and $T_k = 0$ when the result is negative. Let D be the random binary variable that models the result of the *GS*, in such a way that $D = 1$ when the individual is diseased and $D = 0$ when the individual is non-diseased. Let $Se_k = P\left(T_k = 1 \middle| D = 1\right)$ be the sensitivity of the *k*th *BDT* and $Sp_k = P\left(T_k = 0 \middle| D = 0\right)$ be the specificity; $LR_k^+ = Se_k / \left(1 - Sp_k\right)$ the positive likelihood ratio and $LR_k^- = \left(1 - Se_k\right) / Sp_k$ the negative likelihood ratio; $PPV_k = P\left(D = 1 \middle| T_k = 1\right)$ the positive predictive value and $NPV_k = P\left(D = 0 \middle| T_k = 0\right)$ the negative predictive value; and $\kappa_k\left(c\right)$ the weighted kappa coefficient of the *k*th *BDT*. Let us consider that the two *BDTs* and the *GS* are applied independently to all of the individuals in a sample sized *n*, leading to Table 3.1, where s_{ij} is the number of diseased individuals in which $T_1 = i$ and $T_2 = j$, and r_{ij} is the number of non-diseased individuals in which $T_1 = i$ and $T_2 = j$, with $i, j = 0,1$.

The data in Table 3.1 are the product of a multinomial distribution whose probabilities are shown in Table 3.2, where

$$p_{hk} = p\left[Se_1^h \left(1-Se_1\right)^{1-h} Se_2^k \left(1-Se_2\right)^{1-k} + \delta_{hk}\varepsilon_1 \right] \tag{3.1}$$

and

$$q_{hk} = q\left\lfloor Sp_1^{1-h} \left(1-Sp_1\right)^{h} Sp_2^{1-k} \left(1-Sp_2\right)^{k} + \delta_{hk}\varepsilon_0 \right\rfloor, \tag{3.2}$$

with $p = P\left(D=1\right) = \sum_{h,k=0}^{1} p_{hk}$, $q = 1-p = \sum_{h,k=0}^{1} q_{hk}$ and $h,k = 0,1$. The parameter ε_1 is the covariance (Vacek, 1985) between the two *BDTs* when $D=1$, and the parameter ε_0 is the covariance between the two *BDTs* when $D=0$, with $\delta_{hk}=1$ if $h=k$ and $\delta_{hk}=-1$ if $h\neq k$, and verifying that $0 \le \varepsilon_1 \le \min\left\{Se_1\left(1-Se_2\right), Se_2\left(1-Se_1\right)\right\}$ and $0 \le \varepsilon_0 \le \min\left\{Sp_1\left(1-Sp_2\right), Sp_2\left(1-Sp_1\right)\right\}$. If $\varepsilon_1 = \varepsilon_0 = 0$ then the two *BDTs* are conditionally independent on the disease. In practice the supposition of conditional independence is not realistic, so that usually $\varepsilon_1 > 0$ and/or $\varepsilon_0 > 0$. The Vacek model (1985) treats conditional dependence as an additive factor. Using the transformations $\varepsilon_1 = Se_1 Se_2\left(\alpha_1 - 1\right)$ and $\varepsilon_0 = \left(1-Sp_1\right)\left(1-Sp_2\right)\left(\alpha_0 - 1\right)$, the model treats conditional dependence as a multiplicative factor. Parameter α_1 is the covariance (Berry et al, 2002) between the *BDTs* when $D=1$, and α_0 is the covariance (Berry et al, 2002) between the *BDTs* when $D=0$, and it is verified that $1 \le \alpha_1 \le 1/\max\left\{Se_1, Se_2\right\}$ and $1 \le \alpha_0 \le 1/\max\left\{\left(1-Sp_1\right), \left(1-Sp_2\right)\right\}$. In the case that $\alpha_1 = \alpha_0 = 1$ the two *BDTs* are conditionally independent on the disease.

Table 3.1. Frequencies when comparing two *BDTs* subject to a paired design.

	$T_1=0$		$T_1=1$		
	$T_2=0$	$T_2=1$	$T_2=0$	$T_2=1$	Total
$D=1$	s_{00}	s_{01}	s_{10}	s_{11}	s
$D=0$	r_{00}	r_{01}	r_{10}	r_{11}	r
Total	n_{00}	n_{01}	n_{10}	n_{11}	n

Table 3.2. Probabilities subject to a paired design.

	$T_1=1$		$T_1=0$		
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$	Total
$D=1$	p_{11}	p_{10}	p_{01}	p_{00}	p
$D=0$	q_{11}	q_{10}	q_{01}	q_{00}	q
Total	$p_{11}+q_{11}$	$p_{10}+q_{10}$	$p_{01}+q_{01}$	$p_{00}+q_{00}$	1

3.2.1. Sensitivities and specificities

The hypothesis test to compare two sensitivities is

$$H_0: Se_1 = Se_2 \text{ vs } H_1: Se_1 \neq Se_2,$$

that it is equivalent to checking

$$H_0: p_{01} = p_{10} \text{ vs } H_1: p_{01} \neq p_{10}.$$

Therefore, the comparison of two sensitivities is equivalent to the comparison of two paired binomial proportions. If $s_{01}+s_{10}>10$ this hypothesis test is solved by applying the McNemar's test (with correction continuity), i.e.

$$z=\frac{|s_{01}-s_{10}|-0.5}{\sqrt{s_{01}+s_{10}}}\rightarrow N(0,1). \tag{3.3}$$

If $s_{01}+s_{10}\leq 10$, the hypothesis test is solved by applying the exact test to compare two paired binomial proportions. In this situation, the exact *p-value* is

$$p-value=2\times\sum_{i=0}^{h}\binom{s_{01}+s_{10}}{i}\left(\frac{1}{2}\right)^{s_{01}+s_{10}},\quad h=\min(s_{01},s_{10}). \tag{3.4}$$

The comparison of the two specificities is solved in a similar way. The hypothesis test to compare two specificities is

$$H_0: Sp_1=Sp_2 \text{ vs } H_1: Sp_1\neq Sp_2,$$

that it is equivalent to checking

$$H_0: q_{01}=q_{10} \text{ vs } H_1: q_{01}\neq q_{10}.$$

If $r_{01}+r_{10}>10$, the McNemar's test (with correction continuity) is

$$z=\frac{|r_{01}-r_{10}|-0,5}{\sqrt{r_{01}+r_{10}}}\rightarrow N(0,1). \tag{3.5}$$

If $r_{01}+r_{10}\leq 10$, the exact *p-value* is

$$p-value=2\times\sum_{i=0}^{h}\binom{r_{01}+r_{10}}{i}\left(\frac{1}{2}\right)^{r_{01}+r_{10}},\quad h=\min(r_{01},r_{10}). \tag{3.6}$$

The estimation by *CIs* of the difference between two paired proportions has been object of many studies. Agresti and Min (2005) have proposed a *CI* called $Wald+2$ for the differences of two paired binomial proportions, which is a *CI* that has good asymptotic coverage. In terms of data of Table 3.1, the $100(1-\alpha)\%$ $Wald+2$ *CI* for the difference between the two sensitivities is

$$Se_1 - Se_2 \in \frac{s_{10}-s_{01}}{s+2} \pm z_{1-\alpha/2} \frac{\sqrt{(s_{10}+s_{01}+1)-\frac{(s_{10}-s_{01})^2}{s+2}}}{s+2}, \tag{3.7}$$

and for the difference between the two specificities the *CI* is

$$Sp_1 - Sp_2 \in \frac{r_{01}-r_{10}}{r+2} \pm z_{1-\alpha/2} \frac{\sqrt{(r_{01}+r_{10}+1)-\frac{(r_{01}-r_{10})^2}{r+2}}}{r+2}. \tag{3.8}$$

3.2.2. Likelihood ratios

Roldán-Nofuentes and Luna del Castillo (2007) have studied the comparison of the likelihood ratios of two *BDTs* subject to a paired design. When the results of the two *BDTs* are positive, the hypothesis test to check the equality of the *LRs* is

$$H_0 : \omega^+ = 0 \text{ vs } H_1 : \omega^+ \neq 0,$$

and when the results of both *BDTs* are negative, the hypothesis test is

$$H_0 : \omega^- = 0 \text{ vs } H_1 : \omega^- \neq 0,$$

where $\omega^+ = \ln\left(\frac{LR_1^+}{LR_2^+}\right)$ and $\omega^- = \ln\left(\frac{LR_1^-}{LR_2^-}\right)$. In terms of the probabilities given in Table 3.2, the positive *LRs* are written as

$$LR_1^+ = \frac{1-p}{p}\frac{p_{11}+p_{10}}{q_{11}+q_{10}} \text{ and } LR_2^+ = \frac{1-p}{p}\frac{p_{11}+p_{01}}{q_{11}+q_{01}}, \tag{3.9}$$

and the negative *LR*s are written as

$$LR_1^- = \frac{1-p}{p}\frac{p_{01}+p_{00}}{q_{01}+q_{00}} \text{ and } LR_2^- = \frac{1-p}{p}\frac{p_{10}+p_{00}}{q_{10}+q_{00}}. \tag{3.10}$$

Therefore, it holds that

$$\omega^+ = \ln\left\{\frac{(p_{11}+p_{10})(q_{11}+q_{01})}{(p_{11}+p_{01})(q_{11}+q_{10})}\right\} \text{ and } \omega^- = \ln\left\{\frac{(p_{01}+p_{00})(q_{10}+q_{00})}{(p_{10}+p_{00})(q_{01}+q_{00})}\right\}. \tag{3.11}$$

Let $\boldsymbol{\pi} = (p_{00}, p_{01}, p_{10}, p_{11}, q_{00}, q_{01}, q_{10}, q_{11})^T$ be a vector whose components are the probabilities in Table 3.2. As the probabilities p_{ij} and q_{ij} are probabilities of a multinomial distribution, their *MLE*s are $\hat{p}_{ij} = \frac{s_{ij}}{n_{ij}}$ and $\hat{q}_{ij} = \frac{n_{ij}}{n}$, and the estimated variance-covariance matrix of $\hat{\boldsymbol{\pi}}$ is $\hat{\Sigma}_{\hat{\boldsymbol{\pi}}} = \left\{\text{diag}(\hat{\boldsymbol{\pi}}) - \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^T\right\}/n$. The *MLE*s of the positive *LR*s are

$$LR_1^+ = \frac{r(s_{11}+s_{10})}{s(r_{11}+r_{10})} \text{ and } LR_2^+ = \frac{r(s_{11}+s_{01})}{s(r_{11}+r_{01})}, \tag{3.12}$$

and the *MLE*s of the negative *LR*s are

$$LR_1^- = \frac{r(s_{01}+s_{00})}{s(r_{01}+r_{00})} \text{ and } LR_2^- = \frac{r(s_{10}+s_{00})}{s(r_{10}+r_{00})}. \tag{3.13}$$

Then the *MLE*s of ω^+ and ω^- are

$$\hat{\omega}^+ = \ln\left\{\frac{(s_{11}+s_{10})(r_{11}+r_{01})}{(s_{11}+s_{01})(r_{11}+r_{10})}\right\} \text{ and } \hat{\omega}^- = \ln\left\{\frac{(s_{01}+s_{00})(r_{10}+r_{00})}{(s_{10}+s_{00})(r_{01}+r_{00})}\right\}, \tag{3.14}$$

and applying the delta method the estimated variance of $\hat{\omega}$ is

$$\hat{Var}\left(\hat{\omega}\right)=\left(\frac{\partial\omega}{\partial\boldsymbol{\pi}}\right)_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}\Sigma_{\hat{\boldsymbol{\pi}}}\left(\frac{\partial\omega}{\partial\boldsymbol{\pi}}\right)^{T}_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}, \tag{3.15}$$

where $\hat{\omega}$ is $\hat{\omega}^{+}$ or $\hat{\omega}^{-}$. Finally, the statistic for the hypothesis test $H_0:\omega=0$ vs $H_1:\omega\neq 0$ is

$$z=\frac{\hat{\omega}}{\sqrt{\hat{Var}\left(\hat{\omega}\right)}}\xrightarrow[n\to\infty]{}N\left(0,1\right). \tag{3.16}$$

An approximate *CI* for ω is obtained by inverting the contrast statistics (3.16), i.e.

$$\omega\in\hat{\omega}\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\hat{\omega}\right)}, \tag{3.17}$$

where $z_{1-\alpha/2}$ is the $100\left(1-\alpha/2\right)$th percentile of the normal standard distribution. The $100\left(1-\alpha\right)\%$ *CI* for the ratio of the two positive (negative) *LRs* is

$$\frac{LR_1}{LR_2}\in\exp\left[\hat{\omega}\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left(\hat{\omega}\right)}\right], \tag{3.18}$$

where LR_i is LR_i^{+} or LR_i^{-}.

Roldán-Nofuentes et al (2007) also studied the simultaneous comparison of the *LRs* of the two *BDTs*. This hypothesis test consists of solving the global hypothesis test

$$H_0:\left(\omega^{-}=0\ \text{and}\ \omega^{+}=0\right)\ \text{vs}\ H_1:\left(\omega^{-}\neq 0\ \text{and/or}\ \omega^{+}\neq 0\right).$$

Applying the central theorem of the multivariate limit it holds that

$$\sqrt{n}\begin{pmatrix}\hat{\omega}^{+}-\omega^{+}\\ \hat{\omega}^{-}-\omega^{-}\end{pmatrix}\xrightarrow[n\to\infty]{}N\left[\mu=\begin{pmatrix}0\\0\end{pmatrix},\Sigma_{\boldsymbol{\pi}}=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right], \tag{3.19}$$

where the elements of the variance-covariance matrix are estimated applying the delta method. In this way, if $\boldsymbol{\omega} = \left(\omega^+, \omega^-\right)^T$ then

$$\hat{\Sigma}_{\hat{\boldsymbol{\omega}}} \approx \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\pi}}\right)_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}} \hat{\Sigma}_{\hat{\boldsymbol{\pi}}} \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\pi}}\right)^T_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}}, \tag{3.20}$$

and the test statistic of the global hypothesis test is

$$Q^2 = \hat{\boldsymbol{\omega}}^T \Sigma^{-1}_{\hat{\boldsymbol{\omega}}} \hat{\boldsymbol{\omega}} \xrightarrow[n\to\infty]{} \chi^2_2 . \tag{3.21}$$

Finally, the comparison of the *LRs* of the two *BDTs* subject to a paired design is realised following the next steps: 1) Solve the global hypothesis test $H_0 : \left(\omega^- = 0 \text{ and } \omega^+ = 0\right)$ to an error of α using the statistic $Q^2 = \hat{\boldsymbol{\omega}}^T \Sigma^{-1}_{\hat{\boldsymbol{\omega}}} \hat{\boldsymbol{\omega}}$; 2) If the global hypothesis test is not significant to an error of α, then the homogeneity of the *LRs* of the two *BDTs* is not rejected, but if the global hypothesis test is significant to an error of α, then the study of the causes of the significance is performed by solving hypothesis tests $H_0 : \omega^+ = 0$ and $H_0 : \omega^- = 0$ along with a multiple comparison method (e.g. Bonferroni (1936), Holm (1979) or Hochberg (1988)) to an error of α.

3.2.3. Predictive values

The comparison of the predictive values of two binary diagnostic tests is a topic of great interest in the study of Statistical Methods for Diagnosis, and has been the subject of many papers in the literature of Statistics. The most recent studies are: Leisenring et al (2000), Wang et al (2006), Roldán-Nofuentes et al (2012) and Kosinski (2013). From Tables 3.1 and 3.2, the *MLEs* of *PVs* are

$$\hat{PPV}_1 = \frac{s_{10}+s_{11}}{s_{10}+s_{11}+r_{10}+r_{11}} \text{ and } \hat{PPV}_2 = \frac{r_{00}+r_{01}}{s_{00}+s_{01}+r_{00}+r_{01}} \tag{3.22}$$

for *Test* 1, and

$$\hat{NPV}_1 = \frac{s_{01}+s_{11}}{s_{01}+s_{11}+r_{01}+r_{11}} \text{ and } \hat{NPV}_2 = \frac{r_{00}+r_{10}}{s_{00}+s_{10}+r_{00}+r_{10}} \tag{3.23}$$

for *Test* 2. Applying the delta method, the estimated variances-covariances of the estimators of the predictive values are:

$$\hat{Var}\left(\hat{PPV}_1\right) = \frac{(s_{10}+s_{11})(r_{10}+r_{11})}{n(s_{10}+s_{11}+r_{10}+r_{11})^2}, \hat{Var}\left(\hat{PPV}_2\right) = \frac{(s_{01}+s_{11})(r_{01}+r_{11})}{n(s_{01}+s_{11}+r_{01}+r_{11})^2},$$

$$\hat{Var}\left(\hat{NPV}_1\right) = \frac{(s_{00}+s_{01})(r_{00}+r_{01})}{n(s_{00}+s_{01}+r_{00}+r_{01})^2}, \hat{Var}\left(\hat{NPV}_2\right) = \frac{(s_{00}+s_{10})(r_{00}+r_{10})}{n(s_{00}+s_{10}+r_{00}+r_{10})^2},$$

$$\hat{Cov}\left(\hat{PPV}_1, \hat{PPV}_2\right) = \frac{s_{01}s_{10}r_{11} + s_{11}\left[r_{01}(r_{10}+r_{11}) + r_{11}(s_{01}+s_{10}+s_{11}+r_{10}+r_{11})\right]}{(s_{01}+s_{11}+r_{01}+r_{11})^2(s_{10}+s_{11}+r_{10}+r_{11})^2},$$

$$\hat{Cov}\left(\hat{PPV}_1, \hat{NPV}_1\right) = 0,$$

$$\hat{Cov}\left(\hat{PPV}_1, \hat{NPV}_2\right) = -\frac{s_{00}(s_{10}+s_{11})r_{10} + s_{10}r_{10}(s_{10}+s_{11}+r_{00}+r_{10}) + s_{10}(r_{00}+r_{10})r_{11}}{(s_{00}+s_{10}+r_{00}+r_{10})^2(s_{10}+s_{11}+r_{10}+r_{11})^2},$$

$$\hat{Cov}\left(\hat{PPV}_2, \hat{NPV}_1\right) = -\frac{s_{00}(s_{01}+s_{11})r_{01} + s_{01}r_{01}(s_{01}+s_{11}+r_{00}+r_{01}) + s_{01}(r_{00}+r_{01})r_{11}}{(s_{00}+s_{01}+r_{00}+r_{01})^2(s_{01}+s_{11}+r_{01}+r_{11})^2},$$

$$\hat{Cov}\left(\hat{PPV}_2, \hat{NPV}_2\right) = 0,$$

$$\hat{Cov}\left(\hat{NPV}_1, \hat{NPV}_2\right) = \frac{s_{00}(r_{00}+r_{01})r_{10} + r_{00}\left[r_{00}^2 + s_{01}s_{10} + s_{00}(s_{01}+s_{10}+r_{00}+r_{01})\right]}{(s_{00}+s_{01}+r_{00}+r_{01})^2(s_{00}+s_{10}+r_{00}+r_{10})^2}. \tag{3.24}$$

We will now describe each of these four methods to compare the *PVs* of two *BDTs* subject to a paired design.

3.2.3.1. Study of Leisenring et al

Leisenring et al (2000) studied the comparison of the *PPVs* and *NPVs* of two *BDTs* through marginal regression models. These authors deduced score statistics to compare the *PPVs* and *NPVs* of two *BDTs* in paired designs. The score statistic for the test

$$H_0 : PPV_1 = PPV_2 \quad \text{vs} \quad H_1 : PPV_1 \neq PPV_2 \tag{3.25}$$

is

$$T_{PPV} = \frac{\left(s_{11}\left(1-2\bar{Z}_1\right)+s_{01}\left(1-\bar{Z}_1\right)-s_{10}\bar{Z}_1\right)^2}{s_{11}\left(1-\bar{D}_1\right)^2\left(1-2\bar{Z}_1\right)^2+s_{01}\left(1-\bar{D}_1\right)^2\left(1-\bar{Z}_1\right)^2+s_{10}\left(1-\bar{D}_1\right)^2\bar{Z}_1^2+r_{11}\bar{D}_1^2\left(1-2\bar{Z}_1\right)^2+r_{01}\bar{D}_1^2\left(1-\bar{Z}_1\right)^2+r_{10}\bar{D}_1^2\bar{Z}_1^2}, \tag{3.26}$$

and the score statistic for the test

$$H_0 : NPV_1 = NPV_2 \quad \text{vs} \quad H_1 : NPV_1 \neq NPV_2 \tag{3.27}$$

is

$$T_{NPV} = \frac{\left(r_{00}\left(1-2\bar{Z}_2\right)+r_{10}\left(1-\bar{Z}_2\right)-r_{01}\bar{Z}_2\right)^2}{r_{00}\left(1-\bar{D}_2\right)^2\left(1-2\bar{Z}_2\right)^2+r_{10}\left(1-\bar{D}_2\right)^2\left(1-\bar{Z}_2\right)^2+r_{01}\left(1-\bar{D}_2\right)^2\bar{Z}_2^2+s_{00}\bar{D}_2^2\left(1-2\bar{Z}_2\right)^2+s_{10}\bar{D}_2^2\left(1-\bar{Z}_2\right)^2+s_{01}\bar{D}_2^2\bar{Z}_2^2}. \tag{3.28}$$

The score statistics have a chi-squared distribution with 1 degree of freedom when the null hypothesis is true, and where

$$\bar{Z}_1 = \frac{s_{11}+s_{01}+r_{11}+r_{01}}{2s_{11}+s_{01}+s_{10}+2r_{11}+r_{10}+r_{01}}, \quad \bar{D}_1 = \frac{2s_{11}+s_{01}+s_{10}}{2s_{11}+s_{01}+s_{10}+2r_{11}+r_{10}+r_{01}},$$

$$\bar{Z}_2 = \frac{s_{00}+s_{10}+r_{00}+r_{10}}{2s_{00}+s_{01}+s_{10}+2r_{00}+r_{01}+r_{10}} \text{ and } \bar{D}_2 = \frac{2r_{00}+r_{01}+r_{10}}{2s_{00}+s_{01}+s_{10}+2r_{00}+r_{01}+r_{10}}.$$

3.2.3.2. Study of Wang et al

Wang et al (2006) studied the comparison of the *PVs* of two *BDTs* through a weighted least square method. The statistics proposed for hypothesis tests (3.25) and (3.27) are

$$\chi^2_{PPV} = \frac{\left(P\hat{P}V_1 - P\hat{P}V_2\right)^2}{\hat{V}ar\left(P\hat{P}V_1 - P\hat{P}V_2\right)} \text{ and } \chi^2_{NPV} = \frac{\left(N\hat{P}V_1 - N\hat{P}V_2\right)^2}{\hat{V}ar\left(N\hat{P}V_1 - N\hat{P}V_2\right)}, \tag{3.29}$$

respectively. Both statistics have a chi-squared distribution with 1 degree of freedom, and the variances are estimated by applying the delta method (equations (3.24)).

CIs for the difference between the two *PPVs* and the two *NPVs* are obtained by inverting each contrast statistics (3.29), i.e.

$$PPV_1 - PPV_2 \in \left(P\hat{P}V_1 - P\hat{P}V_2\right) \pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left(P\hat{P}V_1 - P\hat{P}V_2\right)} \tag{3.30}$$

and

$$NPV_1 - NPV_2 \in \left(N\hat{P}V_1 - N\hat{P}V_2\right) \pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left(N\hat{P}V_1 - N\hat{P}V_2\right)}. \tag{3.31}$$

Wang et al have compared this method with the method of Leisenring et al (2000), and they recommend using the weighted least square method.

3.2.3.3. Study of Kosinski

Kosinski (2013) has proposed a weighted generalized score statistic to solve the hypothesis tests of comparison of the *PVs*. The weighted generalized score statistic for the hypothesis test (3.25) is

$$T_{PPV}^{WGS} = \frac{\left(P\hat{P}V_1 - P\hat{P}V_2\right)^2}{\left\{P\hat{P}V_p\left(1-P\hat{P}V_p\right) - 2C_p^{PPV}\right\}\left(\frac{1}{n_{10}+n_{11}} + \frac{1}{n_{01}+n_{11}}\right)}, \tag{3.32}$$

and the weighted generalized score statistic for the hypothesis test (3.27) is

$$T_{NPV}^{WGS} = \frac{\left(N\hat{P}V_1 - N\hat{P}V_2\right)^2}{\left\{N\hat{P}V_p\left(1-N\hat{P}V_p\right) - 2C_p^{NPV}\right\}\left(\frac{1}{n_{00}+n_{01}} + \frac{1}{n_{00}+n_{10}}\right)}. \tag{3.33}$$

Both statistics have a chi-squared distribution with 1 degree of freedom when the null hypothesis is true, and where

$$P\hat{P}V_p = \frac{2s_{11}+s_{10}+s_{01}}{2n_{11}+n_{10}+n_{01}} \text{ and } N\hat{P}V_p = \frac{2r_{00}+r_{01}+r_{10}}{2n_{00}+n_{01}+n_{10}} \tag{3.34}$$

are the pooled *PPV* and pooled *NPV* respectively, and

$$C_p^{PPV} = \frac{s_{11}\left(1-P\hat{P}V_p\right)^2 + r_{11}P\hat{P}V_p^2}{2n_{11}+n_{10}+n_{01}} \text{ and } C_p^{NPV} = \frac{s_{00}N\hat{P}V_p^2 + r_{00}\left(1-N\hat{P}V_p\right)^2}{2n_{00}+n_{01}+n_{10}}. \tag{3.35}$$

CIs for the difference between the two *PPVs* and the two *NPVs* are obtained by inverting the contrast statistics (3.32) and (3.33), i.e.

$$PPV_1 - PPV_2 \in \left(P\hat{P}V_1 - P\hat{P}V_2\right) \pm z_{1-\alpha/2}\sqrt{\left[P\hat{P}V_p\left(1-P\hat{P}V_p\right) - 2C_p^{PPV}\right]\left(\frac{1}{n_{10}+n_{11}} + \frac{1}{n_{01}+n_{11}}\right)} \tag{3.36}$$

and

$$NPV_1 - NPV_2 \in \left(N\hat{P}V_1 - N\hat{P}V_2\right) \pm z_{1-\alpha/2}\sqrt{\left[N\hat{P}V_p\left(1-N\hat{P}V_p\right) - 2C_p^{NPV}\right]\left(\frac{1}{n_{00}+n_{01}} + \frac{1}{n_{00}+n_{10}}\right)}. \tag{3.37}$$

Kosinski has demonstrated that his method performed better in terms of the type I error than methods of Leisenring et al (2000) and of Wang et al (2006).

3.2.3.4. Study of Roldán-Nofuentes et al

Roldán-Nofuentes et al (2012) have studied the simultaneous comparison of the *PVs* of two *BDTs* in paired design, and these authors have shown that the comparison of the *PVs* should be performed simultaneously and not independently (as are the previous methods). The simultaneous comparison of the *PVs* of two *BDTs* consists of solving the hypothesis test

$$\begin{aligned} &H_0 : \left(PPV_1 = PPV_2 \text{ and } NPV_2 = NPV_2\right) \\ &H_1 : \left(PPV_1 \neq PPV_2 \text{ and/or } NPV_2 \neq NPV_2\right). \end{aligned} \tag{3.38}$$

The contrast statistics for the hypothesis test $H_0 : \left(PPV_1 = PPV_2 \text{ and } NPV_2 = NPV_2\right)$ is

$$Q^2 = \hat{\boldsymbol{\eta}}^T \boldsymbol{\varphi}^T \left(\boldsymbol{\varphi} \hat{\Sigma} \boldsymbol{\varphi}^T\right)^{-1} \boldsymbol{\varphi} \hat{\boldsymbol{\eta}} \xrightarrow[n \to \infty]{} \chi^2_2 \tag{3.39}$$

where $\hat{\boldsymbol{\eta}} = \left(P\hat{P}V_1, P\hat{P}V_2, N\hat{P}V_1, N\hat{P}V_2\right)^T$, $\hat{\Sigma}$ is the estimated variance-covariance matrix of $\hat{\boldsymbol{\eta}}$ (its elements are shown in equations (3.24)) and $\boldsymbol{\varphi} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$ is the design matrix. The statistic Q^2 (3.39) is distributed asymptotically according to a central chi-square distribution with two degrees of freedom if H_0 is true. To apply this method it is necessary that all *PVs* can be estimated and that matrix $\boldsymbol{\varphi} \hat{\Sigma} \boldsymbol{\varphi}^T$ is non-singular. Therefore, the method cannot be applied if there are many observed frequencies that are equal to zero. If this global hypothesis test is significant to an error rate of α, the investigation of the causes of the significance is carried out by comparing the *PPVs* and the *NPVs* independently (for example, applying the method of Kosinski (2013)) and subsequently applying a method of multiple comparisons (method of Holm (1979) or method of Hochberg (1988)) to the same error rate of α. Simulation

experiments performed have shown that samples of between 300 and 500 individuals are necessary in order for the power of the global hypothesis test to be high ($\geq 80\%$).

3.2.4. Weighted kappa coefficients

Bloch (1997) has studied the comparison of the weighted kappa coefficients of two *BDTs* subject to a paired design. The hypothesis test is

$$H_0 : \kappa_1(c) = \kappa_2(c) \quad \text{vs} \quad H_1 : \kappa_1(c) \neq \kappa_2(c). \tag{3.40}$$

In terms of the probabilities of the Table 3.2, the weighted kappa coefficient of *Test* 1 is

$$\kappa_1(c) = \frac{(p_{11}+p_{10})(q_{01}+q_{00}) - (p_{01}+p_{00})(q_{10}+q_{11})}{pc\sum_{k=0}^{1}(p_{0k}+q_{0k}) + q(1-c)\sum_{k=0}^{1}(p_{1k}+q_{1k})} \tag{3.41}$$

and that of *Test* 2 is

$$\kappa_2(c) = \frac{(p_{11}+p_{01})(q_{10}+q_{00}) - (p_{10}+p_{00})(q_{01}+q_{11})}{pc\sum_{h=0}^{1}(p_{h0}+q_{h0}) + q(1-c)\sum_{h=0}^{1}(p_{h1}+q_{h1})}. \tag{3.42}$$

As the probabilities p_{hk} and q_{hk} are the probabilities of the multinomial distribution, their estimators are $\hat{p}_{hk} = s_{hk}/n$ and $\hat{q}_{hk} = r_{hk}/n$. By substituting each parameter in the expressions of the weighted kappa coefficients with its estimator, the estimators of the weighted kappa coefficients are

$$\hat{\kappa}_1(c) = \frac{(s_{11}+s_{10})(r_{01}+r_{00}) - (s_{01}+s_{00})(r_{10}+r_{11})}{sc\sum_{k=0}^{1}(s_{0k}+r_{0k}) + r(1-c)\sum_{k=0}^{1}(s_{1k}+r_{1k})} \tag{3.43}$$

and

$$\hat{\kappa}_2(c)=\frac{(s_{11}+s_{01})(r_{10}+r_{00})-(s_{10}+s_{00})(r_{01}+r_{11})}{sc\sum_{h=0}^{1}(s_{h0}+r_{h0})+r(1-c)\sum_{h=0}^{1}(s_{h1}+r_{h1})}. \tag{3.44}$$

Finally, the statistic for the hypothesis test (3.40) is

$$z=\frac{\hat{\kappa}_1(c)-\hat{\kappa}_2(c)}{\sqrt{\hat{V}ar\left[\hat{\kappa}_1(c)\right]+\hat{V}ar\left[\hat{\kappa}_2(c)\right]-2\hat{C}ov\left[\hat{\kappa}_1(c),\hat{\kappa}_2(c)\right]}}\xrightarrow[n\to\infty]{}N(0,1), \tag{3.45}$$

where the expressions of the variances and the covariance have been obtained by Bloch applying the delta method, i.e.

$$\begin{aligned}\hat{V}ar\left[\hat{\kappa}_1(c)\right]=&\frac{1}{n\left[c\hat{p}\frac{n_{00}+n_{01}}{n}+(1-c)\hat{q}\frac{n_{11}+n_{10}}{n}\right]^2}\times\\ &\left\{\left[\hat{q}\left(1-\hat{\kappa}_1(c)\right)\frac{n_{00}+n_{01}}{n}\right]^2\frac{s_{11}+s_{10}}{n}+\left[\hat{p}\left(1-\hat{\kappa}_1(c)\right)\frac{n_{00}+n_{01}}{n}+(1-c)\hat{\kappa}_1(c)\right]^2\frac{r_{11}+r_{10}}{n}+\right.\\ &\left.\left[\frac{n_{11}+n_{10}}{n}\hat{q}\left(1-\hat{\kappa}_1(c)\right)+c\hat{\kappa}_1(c)\right]^2\frac{s_{01}+s_{00}}{n}+\left[\frac{n_{11}+n_{10}}{n}\hat{p}\left(1-\hat{\kappa}_1(c)\right)\right]^2\frac{r_{01}+r_{00}}{n}\right\},\end{aligned} \tag{3.46}$$

$$\begin{aligned}\hat{V}ar\left[\hat{\kappa}_2(c)\right]=&\frac{1}{n\left[c\hat{p}\frac{n_{00}+n_{10}}{n}+(1-c)\hat{q}\frac{n_{11}+n_{01}}{n}\right]^2}\times\\ &\left\{\left[\hat{q}\left(1-\hat{\kappa}_2(c)\right)\frac{n_{00}+n_{10}}{n}\right]^2\frac{s_{11}+s_{01}}{n}+\left[\hat{p}\left(1-\hat{\kappa}_2(c)\right)\frac{n_{00}+n_{10}}{n}+(1-c)\hat{\kappa}_2(c)\right]^2\frac{r_{11}+r_{01}}{n}+\right.\\ &\left.\left[\frac{n_{11}+n_{01}}{n}\hat{q}\left(1-\hat{\kappa}_2(c)\right)+c\hat{\kappa}_2(c)\right]^2\frac{s_{10}+s_{00}}{n}+\left[\frac{n_{11}+n_{01}}{n}\hat{p}\left(1-\hat{\kappa}_2(c)\right)\right]^2\frac{r_{10}+r_{00}}{n}\right\}\end{aligned} \tag{3.47}$$

and

$$\hat{Cov}\left[\hat{\kappa}_1(c),\hat{\kappa}_2(c)\right]=$$
$$\frac{n}{\left[c\hat{p}\left(n_{00}+n_{01}\right)+\left(1-c\right)\hat{q}\left(n_{11}+n_{10}\right)\right]\left[c\hat{p}\left(n_{00}+n_{10}\right)+\left(1-c\right)\hat{q}\left(n_{11}+n_{01}\right)\right]}\times$$
$$\left\{\left(1-\hat{\kappa}_1(c)\right)\left(1-\hat{\kappa}_2(c)\right)\left[\left(\frac{\left(r_{00}-r_{11}\right)\left(n_{11}+n_{10}\right)}{n^2}+\frac{r_{11}}{n}\right)\hat{p}^2+\right.\right.$$
$$\left(\frac{\left(s_{11}-s_{00}\right)\left(n_{11}+n_{10}\right)}{n^2}+\frac{s_{00}}{n}\right)\hat{q}^2-\frac{\left(s_{11}+s_{10}\right)\left(n_{00}+n_{01}\right)\left(n_{11}+n_{01}\right)}{n^3}\hat{q}^2-$$
$$\frac{\left(r_{11}+r_{10}\right)\left(n_{00}+n_{01}\right)\left(n_{11}+n_{01}\right)}{n^3}\hat{p}^2-\frac{\left(s_{01}+s_{00}\right)\left(n_{00}+n_{10}\right)\left(n_{11}+n_{10}\right)}{n^3}\hat{q}^2-$$
$$\left.\frac{\left(r_{01}+r_{00}\right)\left(n_{00}+n_{10}\right)\left(n_{11}+n_{10}\right)}{n^3}\hat{p}^2\right]+$$
$$\left(1-\hat{\kappa}_1(c)\right)\hat{\kappa}_2(c)\left[\left(1-c\right)\hat{p}\frac{r_{11}}{n}-\left(1-c\right)\hat{p}\frac{\left(r_{11}+r_{01}\right)\left(n_{11}+n_{10}\right)}{n^2}+\right.$$
$$\left.c\hat{q}\frac{s_{00}}{n}-c\hat{q}\frac{\left(s_{10}+s_{00}\right)\left(n_{00}+n_{01}\right)}{n^2}\right]+$$
$$\hat{\kappa}_1(c)\left(1-\hat{\kappa}_2(c)\right)\left[\left(1-c\right)\hat{p}\frac{r_{11}}{n}-\left(1-c\right)\hat{p}\frac{\left(r_{11}+r_{10}\right)\left(n_{11}+n_{01}\right)}{n^2}+\right.$$
$$\left.\left.c\hat{q}\frac{s_{00}}{n}-c\hat{q}\frac{\left(s_{01}+s_{00}\right)\left(n_{00}+n_{10}\right)}{n^2}\right]+\hat{\kappa}_1(c)\hat{\kappa}(c)_2\left[c^2\frac{s_{00}}{n}+\left(1-c\right)^2\frac{r_{11}}{n}\right]\right\},\quad(3.48)$$

where $\hat{q}=1-\hat{p}$. An approximate *CI* for $\kappa_1(c)-\kappa_2(c)$ is obtained by inverting the contrast statistics (3.45), i.e.

$$\hat{\kappa}_1(c)-\hat{\kappa}_2(c)\pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\hat{\kappa}_1(c)\right]+\hat{Var}\left[\hat{\kappa}_2(c)\right]-2\hat{Cov}\left[\hat{\kappa}_1(c),\hat{\kappa}_2(c)\right]}.\quad(3.49)$$

3.3. Combination of two *BDTs*

In disease diagnosis it is common to combine two *BDTs* in order to increase the accuracy of the diagnosis. The combination of two *BDTs* can be carried out in parallel testing or in serial testing, and in each case, with the *AND* rule or the *OR* rule (Zhou et al, 2002). In the parallel testing the two *BDTs* are applied to all the individuals in a sample, and in the serial testing the application of a *BDT* depends on the result of the

other *BDT*. The *AND* rule implies that the diagnosis is positive if both *BDTs* are positive and the diagnosis is negative if one of the two *BDTs* is negative (or both are negative). The *OR* rule implies that the diagnosis is positive if one of the two *BDTs* is positive (or both are positive), and the diagnosis is negative if both *BDTs* are negative. From a clinical perspective (Zhou et al, 2002), the serial testing has the advantage of cost-effectiveness, because only a single test needs to be applied to some of the individuals; however its main disadvantage is the delay in applying the treatment for the disease, because this cannot be started until the result of the second test is obtained (where that is necessary). Marshall (1989) has studied the effect of combining two binary tests on the *PVs* of the combination and has proposed some criteria for testing when combining increases the value of the *PVs*. Lin (1999) has studied the effect of the two diagnostic tests on the *Se* and *Sp* of the combination of both tests. Macaskill et al (2002) have studied the estimation of the *LRs* of the combination of two *BDTs* in parallel testing. In this Chapter we summarize the results of these authors, and we make the second contribution of this thesis: the combination of the weighted kappa coefficients of two *BDTs* when these are applied in parallel testing with the *AND* rule or with the *OR* rule.

3.3.1. Study of Marshall

Marshall (1989) studied the combination of the predictive values of two *BDTs* in parallel testing with the *AND* rule and with the *OR* rule, proposing criteria, and their graphic representations, to determine when there is an increase in the *PVs* of the combination of both *BDTs*. The criteria proposed depend on the covariances between the two *BDTs*. Using the notation followed in Chapters 1 and 2, and supposing that

$PPV_1 \geq PPV_2$ and that the covariances α_1 and α_0 verify $1 \leq \alpha_1 \leq 1/\max\{Se_1, Se_2\}$ and $1 \leq \alpha_0 \leq 1/\max\{(1-Sp_1),(1-Sp_2)\}$, since it is assumed that the association between the two *BDTs* is positive $(\alpha_i > 1)$, the regions where there is and is not an increase in the *PPVs* of the combination of the two *BDTs* are:

a) $R\left\{\hat{PPV}_{AND} > \hat{PPV}_1\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\alpha}_1 > \hat{\phi}_1\right\}$, with $\hat{\phi}_1 = \left(1-\hat{Sp}_2\right)\hat{\alpha}_0 / \hat{Se}_2$.

b) $R\left\{\hat{PPV}_{OR} > \hat{PPV}_1\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\alpha}_1 < \hat{\phi}_2\right\}$, with

$$\hat{\phi}_2 = \left(1-\hat{Sp}_2\right)\hat{\alpha}_0 / \hat{Se}_2 + \left[\left(1-\hat{Sp}_1\right)\hat{Se}_2 - \hat{Se}_1\left(1-\hat{Sp}_2\right)\right] / \left[\hat{Se}_1\left(1-\hat{Sp}_1\right)\hat{Se}_2\right].$$

c) $R\left\{\hat{PPV}_1 \geq \max\left(\hat{PPV}_{AND}, \hat{PPV}_{OR}\right)\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\phi}_1 \leq \hat{\alpha}_1 \leq \hat{\phi}_2\right\}$.

Regarding the *NPVs*, the regions are:

a) $R\left\{\hat{NPV}_{AND} > \hat{NPV}_1\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\alpha}_1 > \hat{\varphi}_1\right\}$, with

$$\hat{\varphi}_1 = \hat{e}\hat{\alpha}_0 + \left[\hat{Se}_1 - \left(1-\hat{Sp}_1\right)\right] / \left[\hat{Se}_1\hat{Se}_2\hat{Sp}_1\right]$$

and

$$\hat{e} = \left(1-\hat{Sp}_1\right)\left(1-\hat{Se}_1\right)\left(1-\hat{Sp}_2\right) / \left(\hat{Se}_1\hat{Se}_2\hat{Sp}_1\right).$$

b) $R\left\{\hat{NPV}_{OR} > \hat{NPV}_1\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\alpha}_1 < \hat{\varphi}_2\right\}$, with

$$\hat{\varphi}_2 = \hat{e}\hat{\alpha}_0 + \left[\hat{Se}_1\left(1-\hat{Sp}_2\right) - \hat{Se}_2\left(1-\hat{Sp}_1\right) + \hat{Se}_2 - \left(1-\hat{Sp}_2\right)\right] / \left(\hat{Se}_1\hat{Se}_2\hat{Sp}_1\right).$$

c) $R\left\{\hat{NPV}_1 \geq \max\left(\hat{NPV}_{AND}, \hat{NPV}_{OR}\right)\right\} = \left\{\hat{\alpha}_1, \hat{\alpha}_0;\ \hat{\varphi}_1 \leq \hat{\alpha}_1 \leq \hat{\varphi}_2\right\}$.

In the previous regions, the estimators of α_1 and α_0 are $\hat{\alpha}_1 = \frac{\hat{p}_{11}}{\hat{p}\hat{S}e_1\hat{S}e_2}$ and $\hat{\alpha}_0 = \frac{\hat{q}_{11}}{\hat{q}\left(1-\hat{S}p_1\right)\left(1-\hat{S}p_2\right)}$.

3.3.2. Study of Lin

Lin (1999) studied the sensitivity and specificity of the combination of two *BDTs* in parallel testing with the *AND* rule, discussing the dependence effect between the two *BDTs*. The results obtained by Lin are obvious and they are obtained from the probabilities given by equations (3.1) and (3.2). With the *AND* rule, the sensitivity and the specificity of the combination of the two *BDTs* are

$$Se_{AND} = \frac{p_{11}}{p} \text{ and } Sp_{AND} = \frac{q_{10}+q_{01}+q_{00}}{q}. \tag{3.50}$$

Lin also discussed the dependence effect between the two *BDTs* on the *PVs*, as well as the estimation of all the parameters (sensitivity, specificity and *PVs*).

3.3.3. Study of Macaskill et al

Macaskill et al (2002) studied the combination of the likelihood ratios of two *BDTs* (a main test and another adjunct test) in parallel testing with the *AND* rule and with the *OR* rule, and compared the combined *LRs* in relation to those of the main test. In this situation, the estimators of the *LRs* are

$$\hat{LR}_1^+ = \frac{n_{11}^+ n_{\bar{D}}}{n_{11}^- n_D} \text{ and } \hat{LR}_1^- = \frac{\left(n_{01}^+ + n_{00}^+\right) n_{\bar{D}}}{\left(n_{01}^- + n_{00}^-\right) n_D} \tag{3.51}$$

for the main test, and

$$\hat{LR}^{+}_{AND} = \frac{n^{+}_{11} n_{\bar{D}}}{n^{-}_{11} n_{D}} \text{ and } \hat{LR}^{-}_{AND} = \frac{\left(n^{+}_{10} + n^{+}_{00}\right) n_{\bar{D}}}{\left(n^{-}_{00} + n^{-}_{10}\right) n_{D}} \tag{3.52}$$

for the combination of the positive *LRs* with the *AND* rule, and

$$\hat{LR}^{+}_{OR} = \frac{\left(n^{+}_{11} + n^{+}_{10}\right) n_{\bar{D}}}{\left(n^{-}_{11} + n^{-}_{01}\right) n_{D}} \text{ and } \hat{LR}^{-}_{OR} = \frac{n^{+}_{00} n_{\bar{D}}}{n^{-}_{00} n_{D}} \tag{3.53}$$

for the combination of the positive *LRs* with the *OR* rule, when n^{+}_{ij} $\left(n^{-}_{ij}\right)$ is the number of diseased (non-diseased) individuals in which the combination of the two *BDTs* gives a result *i* and the main test gives a result *j*, with $i, j = 0,1$, and n_{D} $\left(n_{\bar{D}}\right)$ is the total number of diseased (non-diseased) individuals.

In the situation studied by these authors, there structural zeros, since the combination of the two *BDTs* is compared in relation to a main test. With the *AND* rule, the number of diseased (non-diseased) individuals for whom the test is negative and the combination of the two *BDTs* is positive is equal to 0 $\left(n^{+}_{01} = n^{-}_{01} = 0\right)$. With the *AND* rule, in terms of the frequencies in Table 3.1, it is verified that $n^{+}_{11} = s_{11}$, $n^{+}_{10} = s_{10}$, $n^{+}_{00} = s_{01} + s_{00}$, $n^{-}_{11} = r_{11}$, $n^{-}_{10} = r_{10}$ and $n^{-}_{00} = r_{01} + r_{00}$. With the *OR* rule, the number of diseased (non-diseased) individuals for whom the test is positive and the combination of the two *BDTs* is negative is equal to 0 $\left(n^{+}_{10} = n^{-}_{10} = 0\right)$. With the *OR* rule, in terms of the frequencies in Table 3.1, it is verified that $n^{+}_{11} = s_{11} + s_{10}$, $n^{+}_{01} = s_{01}$, $n^{+}_{00} = s_{00}$, $n^{-}_{11} = r_{11} + r_{10}$, $n^{-}_{01} = r_{01}$ and $n^{-}_{00} = r_{00}$. Macaskill et al proposed *CIs* for LR_{AND}/LR_{1} and LR_{OR}/LR_{1}, i.e.

$$\frac{LR_{AND}}{LR_1} \in \exp\left\{\left[\log\left(\hat{LR}_{AND}\right)-\log\left(\hat{LR}_1\right)\right] \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\log\left(\hat{LR}_{AND}\right)-\log\left(\hat{LR}_1\right)\right]}\right\} \qquad (3.54)$$

and

$$\frac{LR_{OR}}{LR_1} \in \exp\left\{\left[\log\left(\hat{LR}_{OR}\right)-\log\left(\hat{LR}_1\right)\right] \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\log\left(\hat{LR}_{OR}\right)-\log\left(\hat{LR}_1\right)\right]}\right\}, \qquad (3.55)$$

where LR is LR^+ or LR^-, and the variances are estimated applying the delta method. If the interval is greater than 1 (the lower limit if greater than 1), then the combination of the two *BDTs* with the corresponding rule is better than the main test.

3.3.4. Combination of two weighted kappa coefficients

Let us consider two *BDTs*, *Test* 1 and *Test* 2, which are evaluated in relation of the same *GS*. Let T_h be the random variable which models the result of the *h*th *BDT*, in such a way that $T_h = 1$ when the result is positive and $T_h = 0$ when it is negative; and let D be the random variable which models the result of the *GS*, so that $D = 1$ when the individual does have the disease and $D = 0$ when the individual does not have the disease. Let Se_h and Sp_h be the sensitivity and specificity of the *h*th *BDT*. When both *BDTs* are applied in parallel testing the probabilities given in Table 3.2 are obtained. In Table 3.3 the associated losses when the *AND* rule or the *OR* rule are used are shown. In this situation the observed frequencies (Table 3.1) are the product of a multinomial distribution with probabilities $\boldsymbol{\pi} = \left(p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00}\right)^T$, verifying that $p = \sum_{h,k} p_{hk}$, $q = \sum_{h,k} q_{hk}$ and $p + q = 1$, where p is the prevalence of the disease. In the following the weighted kappa coefficient for the combination of the two *BDTs* is

defined and the conditions under which the combination of both *BDTs* increases the combined weighted kappa coefficient are studied.

Table 3.3. Losses in parallel testing.

Losses with the *AND* rule				
	$T_1=1$		$T_1=0$	
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$
$D=1$	0	L	L	L
$D=0$	L'	0	0	0
Losses with the *OR* rule				
	$T_1=1$		$T_1=0$	
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$
$D=1$	0	0	0	L
$D=0$	L'	L'	L'	0

3.3.4.1. Combined weighted kappa coefficient

In parallel testing with the *AND* rule the combined weighted kappa coefficient is

$$\kappa_{AND}(c)=\frac{p_{11}q-q_{11}p}{p(1-Q_{AND})c+qQ_{AND}(1-c)}, \tag{3.56}$$

where $Q_{AND}=p_{11}+q_{11}$ is the probability that the combination of the two *BDTs* will be positive and $1-Q_{AND}=1-p_{11}-q_{11}$ is the probability that it will be negative. With the *OR* rule the combined weighted kappa coefficient is

$$\kappa_{OR}(c) = \frac{q_{00}p - p_{00}q}{p(1-Q_{OR})c + qQ_{OR}(1-c)}, \tag{3.57}$$

where $Q_{OR} = 1 - p_{00} - q_{00}$ is the probability that the combination of the two *BDTs* will be positive and $1 - Q_{OR} = p_{00} + q_{00}$ that it will be negative. The proofs are shown in Appendix 3.1. Expressions (3.56) and (3.57) also can be written in terms of sensitivities and specificities by replacing in these equations p_{ij} and q_{ij} by the expressions (3.1) and (3.2) respectively. Then

$$\kappa_{AND}(c) = \frac{pqY_{AND}}{pc + (q-c)[\alpha_1 Se_1 Se_2 - qY_{AND}]} \tag{3.58}$$

where $Y_{AND} = Se_{AND} + Sp_{AND} - 1$ is the Youden index of the combination of the two *BDTs* in parallel testing with the *AND* rule, and $Se_{AND} = p_{11}/p$ and $Sp_{AND} = (q_{10} + q_{01} + q_{00})/q$ are the sensitivity and the specificity of the combination of the two *BDTs*. For the *OR* rule,

$$\kappa_{OR}(c) = \frac{pqY_{OR}}{pc + (q-c)(Se_1 + Se_2 - \alpha_1 Se_1 Se_2 - qY_{OR})} \tag{3.59}$$

where $Y_{OR} = Se_{OR} + Sp_{OR} - 1 = Y_1 + Y_2 - Y_{AND}$ is the Youden index in parallel testing with the *OR* rule, and where $Se_{OR} = (p_{11} + p_{10} + p_{01})/p$ and $Sp_{OR} = q_{00}/q$ are the sensitivity and specificity of the combination of the two *BDTs*, and $Y_h = Se_h + Sp_h - 1$ is the Youden index of the *h*th *BDT*.

If $\kappa_{\cdot}(c)$ is $\kappa_{AND}(c)$ or $\kappa_{OR}(c)$ depending on the rule used, then the combined weighted kappa coefficient $\kappa_{\cdot}(c)$ verifies the following properties:

a) $-1 \le \kappa_{\cdot}(c) \le 1$, however as every *BDT* must have a Youden index between 0 and 1 $(0 < Y_h < 1)$ then $0 \le \kappa_{\cdot}(c) \le 1$.

b) $\kappa_{\cdot}(c)$ is an function increasing in *c* if $Q_{\cdot} > p$, decreasing if $Q_{\cdot} < p$ and it is a constant function equal to the combined Youden index $Y_{\cdot}$ if $Q_{\cdot} = p$, where $Q_{\cdot}$ is equal to Q_{AND} or Q_{OR}, and $Y_{\cdot}$ is equal to Y_{AND} or Y_{OR}, respectively.

These properties are similar to those of the weighted kappa coefficient of a *BDT*. The object pursued when two *BDTs* are combined is to increase the accuracy of the combination of both tests. In the following, the conditions under which the combination in parallel testing increases the value of the weighted kappa coefficient of the combination of both *BDTs* are studied, both with the *AND* rule and with the *OR* rule. Let us consider that the combination in parallel testing of the two *BDTs* increases the value of the combined weighted kappa coefficient if it is verified that $\kappa_{\cdot}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$ for a fixed value of the weighting index *c*. If this is not the case, $\kappa_{\cdot}(c) \le \max\{\kappa_1(c), \kappa_2(c)\}$ for a fixed value of *c*, the combination of the *BDTs* does not increase the value of the weighted kappa coefficient of the combination. In this situation, the combination of the two *BDTs* should not be carried out for this value of *c*, because it does not increase the value of $\kappa_{\cdot}(c)$.

In all that follows we shall consider that if the subindex *i* is equal to 1 (or 2) then the subindex *j* is equal to 2 (or 1). Moreover it is assumed that the Youden indices of the two *BDTs* are between 0 and 1 $(0 < Y_h < 1,\ h = 1,2)$. The situation in which a Youden index (or both) is equal to 1 is not contemplated, since the weighted kappa coefficient of the corresponding *BDT* is always equal to 1, and then it is always verified that

$\kappa_{AND}(c) \leq \kappa_i(c)$ and $\kappa_{OR}(c) \leq \kappa_i(c)$ (an increase in the combined weighted kappa coefficient is never produced).

3.3.4.1.1. Increase with the AND rule

Let c be the weighting index fixed by the clinician and let $c_h = \frac{q(Q_h Y_{AND} - Q_{AND} Y_h)}{Y_h(p - Q_{AND}) - Y_{AND}(p - Q_h)}$ with $h = 1,2$. Then the combination of the two *BDTs* with the *AND* rule increases the value of the combined weighted kappa coefficient for the weighting index c, that is $\kappa_{AND}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$, if one of the two following conditions is verified:

a) $0 < c < \min\{c_1, c_2\}$, with $0 < c_1 < 1$ and $0 < c_2 < 1$. Moreover, it is always verified that $\kappa_{AND}(1) < \min\{\kappa_1(1), \kappa_2(1)\}$ and $\kappa_{AND}(0) > \max\{\kappa_1(0), \kappa_2(0)\}$.

b) $0 \leq c < c_i < 1$, $c_j > 1$ (or $c_j < 0$) and $Y_j < Y_{AND}$. In this situation it is verified that $\kappa_{AND}(c) > \kappa_i(c)$ for $0 \leq c < c_i < 1$, and $\kappa_{AND}(c) > \kappa_j(c)$ for $0 \leq c \leq 1$.

In any other situation that is different to the previous ones, the combination of the two *BDTs* in parallel testing with the *AND* rule never increases the value of $\kappa_{AND}(c)$ for any value of c. The proofs of these results may be seen in Appendix 3.2.

In Figure 3.1 two graphics are shown for $\kappa_{AND}(c)$ and for different values of sensitivities, specificities, prevalence and covariances. Graphic 1 corresponds to the situation in which c_1 and c_2 are between 0 and 1, and hence $\kappa_{AND}(c)$ is greater than $\kappa_1(c)$ and $\kappa_2(c)$ for $0 \leq c < \min\{c_1, c_2\} = 0.583$; and in Graphic 2 it is verified that c_1

is between 0 and 1, c_2 is greater than 1 and $Y_2 < Y_{AND}$, and therefore $\kappa_{AND}(c) > \kappa_1(c)$ for $0 \le c < 0.095$ and $\kappa_{AND}(c) > \kappa_2(c)$ for any value of $c \in [0,1]$.

Figure 3.1. Examples of combined weighted kappa coefficients with the *AND* rule.

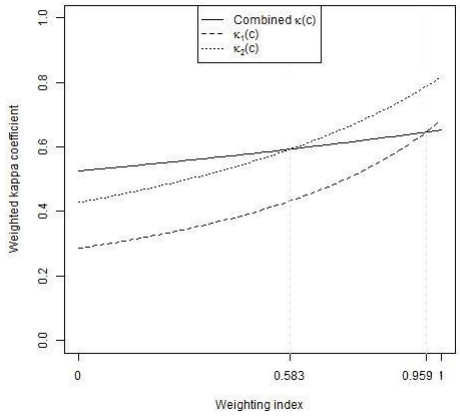
Graphic 1

$Se_1 = 0.75$ $Sp_1 = 0.85$ $Se_2 = 0.85$ $Sp_2 = 0.90$ $p = 10\%$

$\alpha_1 = 1.09$ $\alpha_0 = 3.83$ $Se_{AND} = 0.695$ $Sp_{AND} = 0.943$

$Y_1 = 0.60$ $Y_2 = 0.75$ $Y_{AND} = 0.637$

$c_1 = 0.959$ $c_2 = 0.583$



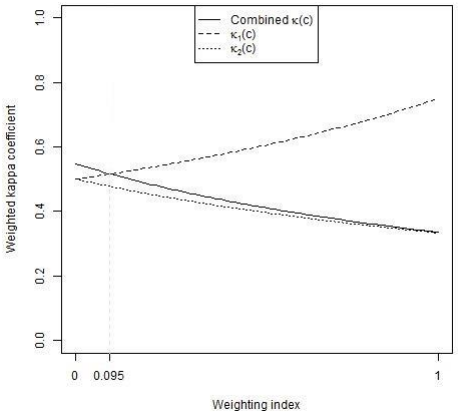
Graphic 2

$Se_1 = 0.90$ $Sp_1 = 0.70$ $Se_2 = 0.60$ $Sp_2 = 0.80$ $p = 50\%$

$\alpha_1 = 1.09$ $\alpha_0 = 2.87$ $Se_{AND} = 0.589$ $Sp_{AND} = 0.828$

$Y_1 = 0.60$ $Y_2 = 0.40$ $Y_{AND} = 0.417$

$c_1 = 0.095$ $c_2 = 1.161$



3.3.4.1.2. Increase with the OR rule

Let c be the weighting index fixed by the clinician and let $c'_h = \dfrac{q(Q_h Y_{OR} - Q_{OR} Y_h)}{Y_h(p - Q_{OR}) - Y_{OR}(p - Q_h)}$ with $h = 1,2$. Then the combination in parallel testing with the *OR* rule increases the value of the combined weighted kappa coefficient for the weighting index c, that is $\kappa_{OR}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$, if one of the two following conditions is verified:

a) $\max\{c_1', c_2'\} < c < 1$, with $0 < c_1' < 1$ and $0 < c_2' < 1$. Moreover, it is always verified that $\kappa_{OR}(0) < \min\{\kappa_1(0), \kappa_2(0)\}$ and $\kappa_{OR}(1) > \max\{\kappa_1(1), \kappa_2(1)\}$.

b) $0 \le c_i' < c < 1$, $c_j' > 1$ (or $c_j' < 0$) and $Y_j < Y_{OR}$. In this situation it is verified that $\kappa_{OR}(c) > \kappa_i(c)$ for $0 \le c < c_i' < 1$, and $\kappa_{OR}(c) > \kappa_j(c)$ for $0 \le c \le 1$.

In any other situation that is different to the previous ones, the combination of the two *BDTs* in parallel with the *OR* rule never increases the value of $\kappa_{OR}(c)$ for any value of *c*. The proofs of these results may be seen in Appendix 3.3.

In Figure 3.2 two graphics for $\kappa_{OR}(c)$ are shown, and for different values of sensitivities, specificities, prevalence and covariances. Graphic 1 corresponds to the situation in which c_1' and c_2' are between 0 and 1, and therefore $\kappa_{OR}(c)$ is greater than $\kappa_1(c)$ and $\kappa_2(c)$ for $\max\{c_1', c_2'\} = 0.423 < c \le 1$; and in Graphic 2 it is verified that c_1' is between 0 and 1, c_2' is greater than 1 and $Y_2 < Y_{OR}$, so that $\kappa_{OR}(c) > \kappa_1(c)$ for $0.972 < c \le 1$ and $\kappa_{OR}(c) > \kappa_2(c)$ for any value of $c \in [0,1]$.

Figure 3.2. Examples of combined weighted kappa coefficients with the *OR* rule.

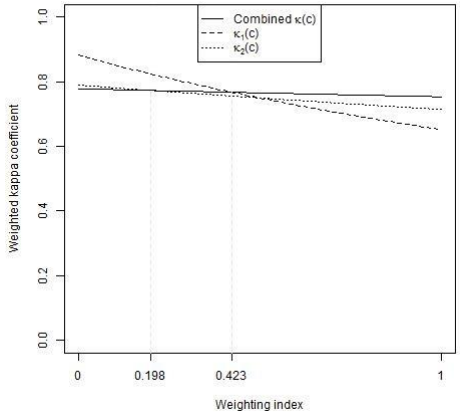
Graphic 1

$Se_1 = 0.80$ $Sp_1 = 0.95$ $Se_2 = 0.85$ $Sp_2 = 0.90$ $p = 50\%$

$\alpha_1 = 1.14$ $\alpha_0 = 8.2$ $Se_{OR} = 0.875$ $Sp_{OR} = 0.891$

$Y_1 = 0.75$ $Y_2 = 0.75$ $Y_{OR} = 0.766$

$c_1' = 0.423$ $c_2' = 0.198$



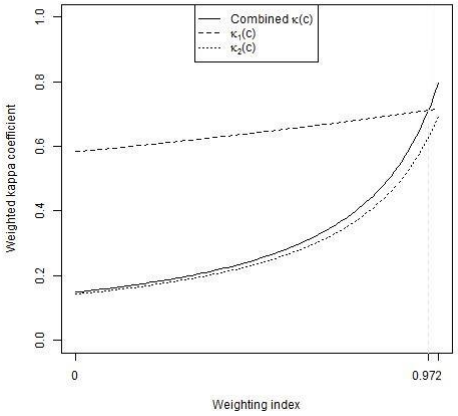
Graphic 2

$Se_1 = 0.75$ $Sp_1 = 0.95$ $Se_2 = 0.80$ $Sp_2 = 0.70$ $p = 10\%$

$\alpha_1 = 1.13$ $\alpha_0 = 2.17$ $Se_{OR} = 0.872$ $Sp_{OR} = 0.683$

$Y_1 = 0.70$ $Y_2 = 0.50$ $Y_{OR} = 0.555$

$c_1' = 0.972$ $c_2' = 3.147$



3.3.4.2. Estimation and CIs

By substituting each parameter in the expressions of the combined weighted kappa coefficients with its estimator, the estimators of the combined weighted kappa coefficients are obtained, i.e.

$$\hat{\kappa}_{AND}(c) = \frac{s_{11}r - r_{11}s}{s(n - s_{11} - r_{11})c + r(s_{11} + r_{11})(1-c)} \tag{3.60}$$

and

$$\hat{\kappa}_{OR}(c) = \frac{r_{00}s - s_{00}r}{s(s_{00} + r_{00})c + r(n - s_{00} - r_{00})(1-c)}. \tag{3.61}$$

Their asymptotic variances are obtained by applying the delta method. As $\boldsymbol{\pi}$ is the vector of probabilities of a multinomial distribution, then $\Sigma_{\hat{\pi}} = \left\{\text{diag}\left(\boldsymbol{\pi}\right) - \boldsymbol{\pi}\boldsymbol{\pi}^T\right\}/n$ and applying the delta method we obtain that

$$Var\left[\hat{\kappa}_{\cdot}\left(c\right)\right] = \left(\frac{\partial \kappa_{\cdot}\left(c\right)}{\partial \boldsymbol{\pi}}\right)\Sigma_{\hat{\pi}}\left(\frac{\partial \kappa_{\cdot}\left(c\right)}{\partial \boldsymbol{\pi}}\right)^T, \tag{3.62}$$

where $\hat{\kappa}_{\cdot}\left(c\right)$ is $\hat{\kappa}_{AND}\left(c\right)$ or $\hat{\kappa}_{OR}\left(c\right)$. By realizing the algebraic operations and substituting each parameter by its estimator, the expressions of the estimated variance of $\hat{\kappa}_{\cdot}\left(c\right)$ is obtained. These are long and complex expressions that require the use of software for their calculation.

In Section 3.3.4.1 the situations in which the combination of the two *BDTs* produce an increase of the combined weighted kappa coefficient have been analysed. These conditions studied are theoretical (for the parameters), so that its practical application does not guarantee that the combination of the two *BDTs* produces and increase in the combined weighted kappa coefficient. Thus, in practice it is necessary to study the increase in $\kappa_{\cdot}\left(c\right)$ using *CIs*. Therefore, once the two *BDTs* have been applied to all the *n* individuals of a random sample and the value of the weighting index *c* has been fixed by the clinician, it is necessary to study if the combination of the two *BDTs* produce such an increase. For this we propose a *CI* for $\theta_{\cdot h} = \kappa_{\cdot}\left(c\right)/\kappa_h\left(c\right)$ with $h = 1,2$. If the *CI* is greater than 1 (that is, the lower limit is greater than 1) then $\kappa_{\cdot}\left(c\right)$ is (to the fixed confidence) larger than $\kappa_h\left(c\right)$. If this happens for $\theta_{\cdot 1}$ and $\theta_{\cdot 2}$, then $\kappa_{\cdot}\left(c\right)$ is greater that $\kappa_1\left(c\right)$ and that $\kappa_2\left(c\right)$, and the combination of the two *BDTs* increases the value of the combined weighted kappa coefficient. In the case that a *CI* (or both) contains the value

1 (or is lower than the said value), then the combination of the two *BDTs* does not increase the value of the combined weighted kappa coefficient. In the following, a *CI* for $\theta_{\cdot h}$ is proposed applying the Fieller method (1940). The Fieller method is a method that has been used for calculating a *CI* for the ratio of two parameters. For this, we shall base it on the approximation of the estimators to the normal bivariate distribution, i.e. $\left(\hat{\kappa}_{\cdot}(c),\hat{\kappa}_h(c)\right)^T \xrightarrow[n\to\infty]{} N\left(0,\Sigma_h\right)$, where the matrix

$$\Sigma_h = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

is estimated by applying the delta method, i.e.

$$\Sigma_h = \left[\frac{\partial\left(\kappa_{\cdot}(c),\kappa_h(c)\right)}{\partial\boldsymbol{\pi}}\right]\Sigma_{\hat{\boldsymbol{\pi}}}\left[\frac{\partial\left(\kappa_{\cdot}(c),\kappa_h(c)\right)}{\partial\boldsymbol{\pi}}\right]^T, \tag{3.63}$$

where $\sigma_{11} = Var\left\lfloor \kappa_{\cdot}(c)\right\rfloor$, $\sigma_{12} = \sigma_{21} = Cov\left\lfloor \kappa_{\cdot}(c),\kappa_h(c)\right\rfloor$ and $\sigma_{22} = Var\left\lfloor \kappa_h(c)\right\rfloor$. By applying the Fieller method, it is verified that $\hat{\kappa}_{\cdot}(c) - \theta_{\cdot h}\hat{\kappa}_h(c) \xrightarrow[n\to\infty]{} N\left(0,\sigma_{11} - 2\theta_{\cdot h}\sigma_{12} + \theta_{\cdot h}^2\sigma_{12}\right)$. The Fieller *CI* is obtained by searching for the set of values for $\theta_{\cdot h}$ that satisfy the inequality

$$\frac{\left[\hat{\kappa}_{\cdot}(c) - \theta_{\cdot h}\hat{\kappa}_h(c)\right]^2}{\hat{\sigma}_{11} - 2\theta_{\cdot h}\hat{\sigma}_{12} + \theta_{\cdot h}^2\hat{\sigma}_{22}} < z_{1-\alpha/2}^2, \tag{3.64}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution. By resolving the equation (3.64) one obtains that the *CI* for $\theta_{\cdot h} = \kappa_{\cdot}(c)/\kappa_h(c)$ is

$$\theta_{\cdot h} \in \frac{\hat{\Delta}_1 \pm \sqrt{\hat{\Delta}_1^2 - \hat{\Delta}_2\hat{\Delta}_3}}{\hat{\Delta}_3}, \tag{3.65}$$

where $\hat{\Delta}_1 = \hat{\kappa}_{\cdot}(c)\hat{\kappa}_h(c) - \hat{\sigma}_{12}z^2_{1-\alpha/2}$, $\hat{\Delta}_2 = \hat{\kappa}_{\cdot}(c) - \hat{\sigma}_{11}z^2_{1-\alpha/2}$ and $\hat{\Delta}_3 = \hat{\kappa}_h(c) - \hat{\sigma}_{22}z^2_{1-\alpha/2}$. This interval is valid when $\hat{\Delta}_1^2 > \hat{\Delta}_2\hat{\Delta}_3$ and $\hat{\Delta}_3 \neq 0$.

On the other hand, if the combination of the two *BDTs* increases the value of the combined weighted kappa coefficient then it is of interest to calculate a *CI* for this parameter. As the values of $\kappa_{AND}(c)$ and of $\kappa_{OR}(c)$ are between 0 and 1, the logit transformations can be applied. The logit transformation of $\hat{\kappa}_{\cdot}(c)$ is closer to a normal distribution with mean $\text{logit}\left[\kappa_{\cdot}(c)\right]$. Thus, the $100(1-\alpha)\%$ *CI* for $\text{logit}\left[\kappa_{\cdot}(c)\right]$ is $\text{logit}\left[\hat{\kappa}_{\cdot}(c)\right] \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right)\right]}$, where $\hat{Var}\left[\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right)\right]$ is obtained by applying the delta method. Finally the $100(1-\alpha)\%$ logit *CI* for $\kappa_{\cdot}(c)$ is

$$\kappa_{\cdot}(c) \in \frac{\exp\left\{\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right)\right]}\right\}}{1+\exp\left\{\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right) \pm z_{1-\alpha/2}\sqrt{\hat{Var}\left[\text{logit}\left(\hat{\kappa}_{\cdot}(c)\right)\right]}\right\}}. \qquad (3.66)$$

3.3.4.3. Simulation experiments

Monte Carlo simulation experiments have been carried out for studying the coverage probabilities of the proposed *CIs* in the previous Section. For this, 10000 random samples of multinomial distributions have been generated, whose probabilities have been calculated from expressions (3.1) and (3.2), and with the sizes $n = \{100, 200, 300, 400, 500, 1000\}$. The random samples have been generated from values of weighted kappa coefficients, and not setting the values of *Se* and of *Sp*, from equations (1.11) and (1.12), i.e.

$$\kappa_i(0) = \frac{Sp_i - (1 - Q_i)}{Q_i} \text{ and } \kappa_i(1) = \frac{Se_i - Q_i}{1 - Q_i},$$

in the following way. As disease prevalence we have taken the values 10%, 25% and 50%; as weighting index *c* we have taken values 0.1, 0.5 and 0.9; and for each one of the two *BDTs* we have taken as $\kappa_h(0)$ and $\kappa_h(1)$ the values $\{0.10, 0.20,..., 0.80, 0.90\}$, with $h = 1,2$. Once the values of *p*, $\kappa_h(0)$ and $\kappa_h(1)$ were set, the values of *Se* and of *Sp* of each *BDT* were calculated solving (through the Newton-Raphson method) the system formed by equations (1.11) and (1.12). Finally, the value of the weighted kappa coefficient was calculated with equation (1.10) (or (1.13)), i.e.

$$\kappa_h(c) = \frac{pq(Se_h + Sp_h - 1)}{p(1 - Q_h)c + qQ_h(1 - c)},$$

where $Q_h = pSe_h + q(1 - Sp_h)$. As values of each weighted kappa coefficient only the values $\kappa_h(c) = \{0.2, 0.4, 0.6, 0.8\}$ have been considered. Therefore, considering the classification of the values of the weighted kappa coefficient given by Cicchetti (2001), values of $\kappa_h(c)$ were considered with different levels of clinical significance: poor $(\kappa_h(c) < 0.40)$, fair $(0.40 \le \kappa_h(c) \le 0.59)$, good $(0.60 \le \kappa_h(c) \le 0.74)$ and excellent $(0.75 \le \kappa_h(c) \le 1)$. As covariances α_1 and α_0 we have taken intermediate and high values. All the samples have been generated in such a way that in all of them it has been possible to estimate all the parameters and their variances-covariances. For the whole study the confidence level has been taken as 95%.

In Table 3.4 (Fieller *CI*) the coverage probabilities and the average lengths of the *CI* of $\kappa_{AND}(c)/\kappa_1(c)$ are shown, for different values of $\kappa_1(c)$ and $\kappa_2(c)$, indicating in

each case the values of p, Se_i, Sp_i, α_1 and α_0 with those with which we have calculated the value of $\kappa_1(c)$, $\kappa_2(c)$ and $\kappa_{AND}(c)$. We show situations in which there is an increase of $\kappa_{AND}(c)$ and situations in which there is no an increase of $\kappa_{AND}(c)$ for the values $c=\{0.1, 0.5, 0.9\}$. The results for the *CI* of $\kappa_{AND}(c)/\kappa_2(c)$ are not shown because they are very similar to those obtained for $\kappa_{AND}(c)/\kappa_1(c)$. In general terms the Fieller *CI* presents a coverage probability that fluctuates around 95%, although the coverage probability may overestimate the desired coverage of 95% when the sample size is relatively small $(n=100\text{-}200)$. With regard to the *CI* for $\kappa_{AND}(c)$ the results are shown in Table 3.4 (Logit *CI*). The scenarios are the same as for the Fieller *CI*. In general terms, the coverage probabilities of the logit *CI* fluctuate around 95%, although they may overestimate the desired coverage of 95% when the sample size is 100-200.

In Table 3.5 the coverage probabilities and the average lengths of the *CI* of $\kappa_{OR}(c)/\kappa_1(c)$ (Fieller *CI*) and of the *CI* of $\kappa_{OR}(c)$ (Logit *CI*) are shown, for different values of all parameters. We show situations in which there is an increase of $\kappa_{OR}(c)$ and situations in which there is no an increase of $\kappa_{OR}(c)$ for $c=\{0.1, 0.5, 0.9\}$. The results for the *CI* of $\kappa_{OR}(c)/\kappa_2(c)$ are not shown because they are very similar to those obtained for $\kappa_{OR}(c)/\kappa_1(c)$. In general terms the Fieller *CI* of $\kappa_{OR}(c)/\kappa_1(c)$ and the Logit *CI* of $\kappa_{OR}(c)$ have a behaviour very similar to the *CIs* of $\kappa_{AND}(c)/\kappa_1(c)$ and of $\kappa_{AND}(c)$.

Table 3.4. Coverage probabilities and average lengths of the *CI* for $\kappa_{AND}(c)/\kappa_1(c)$ and of the *CI* for $\kappa_{AND}(c)$.

	$\kappa_1(0.1)=0.6\ \ \kappa_2(0.1)=0.2$ $\kappa_{AND}(0.1)=0.426$ $p=10\%$ $Se_1=0.85\ \ Sp_1=0.90$ $Se_2=0.75\ \ Sp_2=0.85$ $\alpha_1=1.281\ \ \alpha_0=6.75$				$\kappa_1(0.5)=0.4\ \ \kappa_2(0.5)=0.4$ $\kappa_{AND}(0.5)=0.449$ $p=10\%$ $Se_1=0.6727\ \ Sp_1=0.8727$ $Se_2=0.6727\ \ Sp_2=0.8727$ $\alpha_1=1.243\ \ \alpha_0=4.429$				$\kappa_1(0.9)=0.6\ \ \kappa_2(0.9)=0.8$ $\kappa_{AND}(0.9)=0.572$ $p=10\%$ $Se_1=0.64\ \ Sp_1=0.96$ $Se_2=0.82\ \ Sp_2=0.98$ $\alpha_1=1.110\ \ \alpha_0=13$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.941	1.357	0.987	0.639	0.978	1.702	0.980	0.485	0.985	0.795	0.989	0.522
200	0.951	0.877	0.969	0.528	0.975	0.843	0.974	0.370	0.981	0.440	0.978	0.403
300	0.959	0.719	0.959	0.458	0.953	0.594	0.956	0.306	0.963	0.318	0.961	0.338
400	0.957	0.619	0.958	0.407	0.950	0.491	0.952	0.268	0.961	0.260	0.959	0.296
500	0.947	0.552	0.958	0.370	0.947	0.429	0.956	0.241	0.956	0.226	0.957	0.267
1000	0.946	0.385	0.953	0.269	0.950	0.292	0.944	0.172	0.948	0.155	0.954	0.191
	$\kappa_1(0.1)=0.6\ \ \kappa_2(0.1)=0.4$ $\kappa_{AND}(0.1)=0.530$ $p=25\%$ $Se_1=0.70\ \ Sp_1=0.90$ $Se_2=0.55\ \ Sp_2=0.85$ $\alpha_1=1.386\ \ \alpha_0=6.1$				$\kappa_1(0.5)=0.4\ \ \kappa_2(0.5)=0.4$ $\kappa_{AND}(0.5)=0.458$ $p=25\%$ $Se_1=0.76\ \ Sp_1=0.72$ $Se_2=0.76\ \ Sp_2=0.72$ $\alpha_1=1.158\ \ \alpha_0=2.286$				$\kappa_1(0.9)=0.8\ \ \kappa_2(0.9)=0.8$ $\kappa_{AND}(0.9)=0.820$ $p=25\%$ $Se_1=0.9429\ \ Sp_1=0.7429$ $Se_2=0.9429\ \ Sp_2=0.7429$ $\alpha_1=1.030\ \ \alpha_0=2.444$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.947	0.494	0.969	0.435	0.971	1.065	0.976	0.367	0.978	0.328	0.939	0.348
200	0.951	0.293	0.953	0.319	0.957	0.525	0.962	0.265	0.971	0.185	0.959	0.224
300	0.955	0.226	0.957	0.264	0.944	0.408	0.952	0.218	0.960	0.141	0.961	0.174
400	0.957	0.193	0.953	0.230	0.954	0.347	0.945	0.190	0.958	0.119	0.956	0.148
500	0.946	0.169	0.948	0.207	0.951	0.305	0.957	0.170	0.953	0.104	0.957	0.131
1000	0.949	0.118	0.955	0.147	0.958	0.211	0.947	0.121	0.935	0.073	0.941	0.092
	$\kappa_1(0.1)=0.4\ \ \kappa_2(0.1)=0.6$ $\kappa_{AND}(0.1)=0.619$ $p=50\%$ $Se_1=0.70\ \ Sp_1=0.70$ $Se_2=0.80\ \ Sp_2=0.80$ $\alpha_1=1.125\ \ \alpha_0=2.167$				$\kappa_1(0.5)=0.8\ \ \kappa_2(0.5)=0.6$ $\kappa_{AND}(0.5)=0.7$ $p=50\%$ $Se_1=0.90\ \ Sp_1=0.90$ $Se_2=0.80\ \ Sp_2=0.80$ $\alpha_1=1.1\ \ \alpha_0=4.6$				$\kappa_1(0.9)=0.8\ \ \kappa_2(0.9)=0.8$ $\kappa_{AND}(0.9)=0.824$ $p=50\%$ $Se_1=0.9692\ \ Sp_1=0.5846$ $Se_2=0.9692\ \ Sp_2=0.5846$ $\alpha_1=1.016\ \ \alpha_0=1.704$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.943	2.519	0.966	0.362	0.939	0.249	0.957	0.275	0.980	0.307	0.935	0.317
200	0.941	0.997	0.958	0.265	0.942	0.166	0.956	0.195	0.972	0.172	0.951	0.201
300	0.944	0.746	0.953	0.218	0.947	0.133	0.953	0.160	0.961	0.131	0.960	0.156
400	0.947	0.620	0.946	0.190	0.951	0.114	0.958	0.139	0.959	0.111	0.956	0.133
500	0.945	0.541	0.951	0.171	0.948	0.101	0.941	0.124	0.956	0.098	0.952	0.118
1000	0.956	0.372	0.941	0.121	0.952	0.071	0.958	0.088	0.947	0.069	0.948	0.083

Cov.: coverage probability. Leng.: average length.

Table 3.5. Coverage probabilities and average lengths of the *CI* for $\kappa_{OR}(c)/\kappa_1(c)$ and of the *CI* for $\kappa_{OR}(c)$.

	$\kappa_1(0.1)=0.6\ \ \kappa_2(0.1)=0.2$ $\kappa_{OR}(0.1)=0.392$ $p=10\%$ $Se_1=0.64\ \ Sp_1=0.96$ $Se_2=0.28\ \ Sp_2=0.92$ $\alpha_1=1.281\ \ \alpha_0=6.75$				$\kappa_1(0.5)=0.4\ \ \kappa_2(0.5)=0.4$ $\kappa_{OR}(0.5)=0.474$ $p=10\%$ $Se_1=0.3368\ \ Sp_1=0.9789$ $Se_2=0.3368\ \ Sp_2=0.9789$ $\alpha_1=1.438\ \ \alpha_0=7.171$				$\kappa_1(0.9)=0.6\ \ \kappa_2(0.9)=0.8$ $\kappa_{OR}(0.9)=0.828$ $p=10\%$ $Se_1=0.64\ \ Sp_1=0.96$ $Se_2=0.82\ \ Sp_2=0.98$ $\alpha_1=1.110\ \ \alpha_0=13$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.943	0.975	0.977	0.442	0.980	5.117	0.985	0.490	0.966	3.121	0.979	0.425
200	0.946	0.456	0.967	0.334	0.971	3.323	0.983	0.393	0.938	1.070	0.974	0.297
300	0.947	0.335	0.957	0.278	0.953	1.331	0.960	0.333	0.945	0.794	0.961	0.239
400	0.953	0.283	0.951	0.242	0.943	0.929	0.957	0.293	0.938	0.670	0.942	0.207
500	0.950	0.251	0.952	0.218	0.946	0.662	0.955	0.264	0.947	0.577	0.945	0.185
1000	0.948	0.173	0.947	0.155	0.948	0.405	0.956	0.189	0.949	0.393	0.947	0.132
	$\kappa_1(0.1)=0.6\ \ \kappa_2(0.1)=0.4$ $\kappa_{OR}(0.1)=0.478$ $p=25\%$ $Se_1=0.70\ \ Sp_1=0.90$ $Se_2=0.55\ \ Sp_2=0.85$ $\alpha_1=1.386\ \ \alpha_0=6.1$				$\kappa_1(0.5)=0.4\ \ \kappa_2(0.5)=0.4$ $\kappa_{OR}(0.5)=0.410$ $p=25\%$ $Se_1=0.40\ \ Sp_1=0.9429$ $Se_2=0.76\ \ Sp_2=0.72$ $\alpha_1=1.158\ \ \alpha_0=2.286$				$\kappa_1(0.9)=0.8\ \ \kappa_2(0.9)=0.4$ $\kappa_{OR}(0.9)=0.514$ $p=25\%$ $Se_1=0.9429\ \ Sp_1=0.7429$ $Se_2=0.8667\ \ Sp_2=0.40$ $\alpha_1=1.030\ \ \alpha_0=1.333$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.941	0.420	0.969	0.368	0.942	3.218	0.976	0.328	0.976	0.483	0.980	0.393
200	0.940	0.270	0.956	0.270	0.944	1.162	0.959	0.238	0.966	0.289	0.974	0.266
300	0.948	0.216	0.958	0.223	0.945	0.782	0.960	0.195	0.955	0.225	0.959	0.213
400	0.953	0.186	0.954	0.195	0.946	0.640	0.952	0.170	0.953	0.191	0.957	0.185
500	0.947	0.166	0.955	0.175	0.948	0.558	0.947	0.152	0.948	0.169	0.948	0.165
1000	0.949	0.116	0.954	0.124	0.947	0.377	0.951	0.108	0.948	0.118	0.946	0.117
	$\kappa_1(0.1)=0.4\ \ \kappa_2(0.1)=0.4$ $\kappa_{OR}(0.1)=0.447$ $p=50\%$ $Se_1=0.2286\ \ Sp_1=0.9429$ $Se_2=0.2286\ \ Sp_2=0.9429$ $\alpha_1=2.688\ \ \alpha_0=9.25$				$\kappa_1(0.5)=0.8\ \ \kappa_2(0.5)=0.6$ $\kappa_{OR}(0.5)=0.7$ $p=50\%$ $Se_1=0.90\ \ Sp_1=0.90$ $Se_2=0.80\ \ Sp_2=0.80$ $\alpha_1=1.1\ \ \alpha_0=4.6$				$\kappa_1(0.9)=0.8\ \ \kappa_2(0.9)=0.8$ $\kappa_{OR}(0.9)=0.824$ $p=50\%$ $Se_1=0.9692\ \ Sp_1=0.5846$ $Se_2=0.9429\ \ Sp_2=0.2286$ $\alpha_1=1.016\ \ \alpha_0=1.148$			
	Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*		Fieller *CI*		Logit *CI*	
n	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.	Cov.	Leng.
100	0.961	3.521	0.978	0.469	0.940	0.251	0.954	0.275	0.979	0.601	0.981	0.462
200	0.959	1.491	0.974	0.347	0.942	0.166	0.958	0.195	0.961	0.389	0.976	0.328
300	0.957	0.887	0.958	0.285	0.946	0.132	0.952	0.160	0.958	0.305	0.960	0.266
400	0.956	0.569	0.954	0.249	0.950	0.114	0.956	0.139	0.949	0.260	0.958	0.229
500	0.949	0.477	0.957	0.224	0.943	0.101	0.949	0.124	0.950	0.230	0.952	0.205
1000	0.951	0.305	0.950	0.160	0.958	0.071	0.957	0.088	0.947	0.162	0.947	0.146

Cov.: coverage probability. Leng.: average length.

3.3.4.4. Program in R

A program has been written in R to solve this problem. This program, called "*cwkc*" (Combination of Weighted Kappa Coefficients), is executed with the command

$$\text{cwkc}\left(s_{11},s_{10},s_{01},s_{00},r_{11},r_{10},r_{01},r_{00},indexc,rule\right)$$

when the *CIs* are calculated to a confidence of 95%, where s_{ij} and r_{ij} are the observed frequencies, *indexc* the value of the weighting index *c*, and *rule* is equal to "and" (including commas) when the *AND* rule is used and is equal to "or" when the *OR* rule is used. The command is

$$\text{cwkc}\left(s_{11},s_{10},s_{01},s_{00},r_{11},r_{10},r_{01},r_{00},indexc,rule,conflevel\right)$$

when the confidence level is $\left(100\times conflevel\right)\%$. First, the program always checks that all the initial values are correct; where this is not so, an error message is given. The program produces the estimations of the weighted kappa coefficients of the *BDTs* and their standard errors, the estimation of the combined weighted kappa coefficient and its standard error, and the *CIs* proposed in Section 3.3.4.2. The program also carries out a similar graph to those shown in Figures 3.1 and 3.2, and which is kept in a file called "Graph_cwkc.jpg" in the same folder from where the program is run. In this graph the values of c_h (or c_h') are indicated, provided that their values are between 0 and 1. Similarly, the results obtained when running the program are kept in a file called "Results_cwkc.txt" in the same folder from where the program is run. The program is written in the Appendix 3.4.

3.3.4.5. Example

The results obtained have been applied to the study of Weiner et al (1979) on the diagnosis of coronary disease, using the clinical history (*Test* 1) and the exercise test (*Test* 2) as diagnostic tests, and the coronary angiography as *GS*. Weiner et al applied the two *BDTs* and the *GS* to a sample of 871 individuals. The results of this study are shown in Table 3.6, where the variable T_1 models the result of the clinical history, T_2 models the result of the exercise test, and *D* the result of the coronary angiography. In the following, the results of the two *BDTs* being combined with the *AND* rule and with the *OR* are analysed.

Table 3.6. Data of the study of Weiner et al (1979).

	$T_1=1$		$T_1=0$		
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$	Total
$D=1$	473	81	29	25	608
$D=0$	22	44	46	151	263
Total	495	125	75	176	871

3.3.4.5.1. AND rule

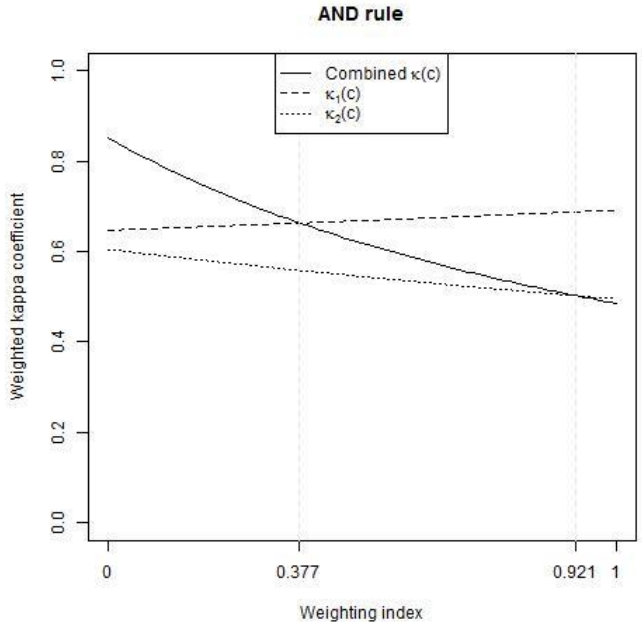
By executing the program "*cwkc*" with the command

$$\text{cwkc}\left(473,81,29,25,22,44,46,151,\textit{indexc},\textit{"and"}\right),$$

one obtains the Figure 3.3, independently of the index *c* chosen by the researcher, as well as the values of c_h, where $c_1=0.377$ and $c_2=0.921$. By analysing this graph one

finds that $\hat{\kappa}_{AND}(c) > \{\hat{\kappa}_1(c), \hat{\kappa}_2(c)\}$ if $c < 0.377$. This result indicates that, in terms of the estimators, the combination of the two *BDTs* increases the value of $\hat{\kappa}_{AND}(c)$ if the clinician is more concerned with the false positives of the combination of the two *BDTs* (for example, when the combination of the two *BDTs* is used as a previous step to a risky treatment), but always for $0 \leq c < 0.377$. If the clinician is more concerned with the false negatives of the combination of the two *BDTs* (for example, when the combination of the two *BDTs* is used as a screening test), then $\hat{\kappa}_{AND}(c) < \hat{\kappa}_1(c)$ for any value $0.5 < c \leq 1$ and the combination of the two *BDTs* never increases the value of $\hat{\kappa}_{AND}(c)$.

Figure 3.3. Graphic of the study of Weiner et al with the *AND* rule.



In order to illustrate the method, let us consider the values 0.1 and 0.3, that are lower values than 0.377, as weighting indices. In Table 3.7 the results are shown. For $c = 0.1$ the *CIs* for $\kappa_{AND}(0.1)/\kappa_1(0.1)$ and $\kappa_{AND}(0.1)/\kappa_2(0.1)$ are larger than 1 (the lower limit is greater than 1) and this indicates that $\kappa_{AND}(0.1) > \{\kappa_1(0.1), \kappa_2(0.1)\}$, and so, the combination of the two *BDTs* increases (to a 95% confidence level) the value of the weighted kappa coefficient. By interpreting the logit *CI* one obtains that $\kappa_{AND}(0.1)$ is, with the confidence of 95%, a value between 0.738 and 0.839, which indicates that the level of clinical significance between the combination of the two *BDTs* and the *GS* is excellent (Cicchetti, 2001). For $c = 0.3$ the *CI* for $\kappa_{AND}(0.3)/\kappa_1(0.3)$ contains the value 1, which indicates that $\kappa_{AND}(0.3)$ is not (to a 95% confidence level) greater than $\kappa_1(0.3)$. Therefore, the combination of the two *BDTs* does not increase (to a 95% confidence level) the value of the combined weighted kappa coefficient. The *CI* for $\kappa_{AND}(0.3)/\kappa_2(0.3)$ is greater than 1, and this indicates that $\kappa_{AND}(0.3)$ is (to a 95% confidence level) greater than $\kappa_2(0.3)$. Therefore the combination of the two *BDTs* has a weighted kappa coefficient greater than that of the *Test* 2, but is not higher than that of the *Test* 1.

For any value of the index *c* greater than 0.377, the value of $\hat{\kappa}_{AND}(c)/\hat{\kappa}_1(c)$ is less than 1. In this situation, the *CI* contains the value $\hat{\kappa}_{AND}(c)/\hat{\kappa}_1(c)$, so that the *CI* is less than 1 (the upper limit of the *CI* is less than 1) or the *CI* contains the value 1, and in both cases the combination of the two *BDTs* with the *AND* rule does not produce (to a 95% confidence level) an increase of $\kappa_{AND}(c)$.

Table 3.7. Results obtained when combining the two weighted kappa coefficients with the *AND* rule.

	$c=0.1$		$c=0.3$	
	Test 1	Test 2	Test 1	Test 2
$\hat{\kappa}_h(c)\pm SE$	0.652 ± 0.0322	0.592 ± 0.0345	0.660 ± 0.0291	0.567 ± 0.0311
$\hat{Y}_h$	0.660	0.567	0.660	0.567
$\hat{\kappa}_{AND}(c)\pm SE$	0.793 ± 0.0256		0.695 ± 0.0249	
$\hat{Y}_{AND}$	0.694		0.694	
Fieller *CI* for $\kappa_{AND}(c)/\kappa_1(c)$	$(1.121\ ,\ 1.328)$		$(0.972\ ,\ 1.144)$	
Fieller *CI* for $\kappa_{AND}(c)/\kappa_2(c)$	$(1.222\ ,\ 1.483)$		$(1.128\ ,\ 1.340)$	
Logit *CI* for $\kappa_{AND}(c)$	$(0.738\ ,\ 0.839)$		$(0.644\ ,\ 0.742)$	

3.3.4.5.2. OR rule

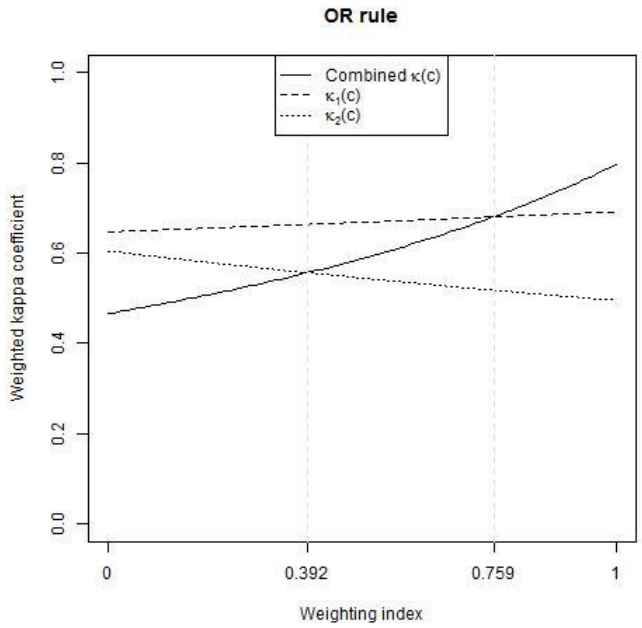
By executing the program "*cwkc*" with the command

$$\text{cwkc}(473,81,29,25,22,44,46,151,indexc,"or"),$$

one obtains the Figure 3.4, independently of the index *c* chosen, as well as the values of c'_h, with $c'_1=0.759$ and $c'_2=0.392$. By analysing this graph one finds that $\hat{\kappa}_{OR}(c)>\{\hat{\kappa}_1(c),\hat{\kappa}_2(c)\}$ if $c>0.759$. This result indicates that, in terms of the estimators, the combination of the two *BDTs* increases the value of $\hat{\kappa}_{OR}(c)$ if the clinician has more concern for the false negative of the combination of the two *BDTs*, but always for $0.759<c\leq1$. If the clinician has more concern for the false positives

$(0 \le c < 0.5)$ of the combination of the two *BDTs*, then $\hat{\kappa}_{OR}(c) < \{\hat{\kappa}_1(c), \hat{\kappa}_2(c)\}$ and the combination of the two *BDTs* never increases the value of $\hat{\kappa}_{OR}(c)$.

Figure 3.4. Graphic of the study of Weiner et al with the *OR* rule.



In order to illustrate the method, we shall take as weighting indices the values 0.8 and 0.9, which are lower values than 0.759. In Table 3.8 the results obtained are shown. For $c = 0.8$ the *CI* for $\kappa_{OR}(0.8)/\kappa_1(0.8)$ contains the value 1, which indicates that $\kappa_{OR}(0.8)$ is not (95% confidence level) greater than $\kappa_1(0.8)$. Therefore, the combination of the two *BDTs* does not increase (to a 95% confidence level) the value of the combined weighted kappa coefficient. The *CI* for $\kappa_{OR}(0.8)/\kappa_2(0.8)$ is greater than 1, which indicates that $\kappa_{OR}(0.8)$ is (to a 95% confidence level) greater than $\kappa_2(0.8)$. Hence, the combination of the two *BDTs* has a weighted kappa coefficient greater than

that of *Test* 2, but is not greater than that of *Test* 1. For $c = 0.9$ the *CIs* for $\kappa_{OR}(0.9)/\kappa_1(0.9)$ and $\kappa_{OR}(0.9)/\kappa_2(0.9)$ are greater than 1, which indicates than $\kappa_{OR}(0.9) > \{\kappa_1(0.9), \kappa_2(0.9)\}$ and therefore the combination of the two *BDTs* increases (to a 95% confidence level) the value of the combined weighted kappa coefficient. By interpreting the logit *CI* one obtains that $\kappa_{OR}(0.9)$ is, with a 95% confidence level, a value between 0.676 and 0.802, which indicates that the level of clinical significance between the combination of the two *BDTs* and the *GS* is a value between good and excellent (Cicchetti, 2001).

For any value of the index *c* lower than 0.759, the value of $\hat{\kappa}_{OR}(c)/\hat{\kappa}_1(c)$ is less than 1. The *CI* will contains the value $\hat{\kappa}_{OR}(c)/\hat{\kappa}_1(c)$, so that the *CI* is less than 1 or it contains the value 1, and in both cases the combination of the two *BDTs* with the *OR* rule does not produce (to a 95% confidence level) an increase of $\kappa_{OR}(c)$.

Table 3.8. Results obtained when combining the two weighted kappa coefficients with the *OR* rule.

	$c=0.8$		$c=0.9$	
	Test 1	Test 2	Test 1	Test 2
$\hat{\kappa}_h(c)\pm SE$	0.682 ± 0.0300	0.514 ± 0.0318	0.687 ± 0.0319	0.505 ± 0.0327
$\hat{Y}_h$	0.660	0.567	0.660	0.567
$\hat{\kappa}_{OR}(c)\pm SE$	0.698 ± 0.0304		0.744 ± 0.0322	
$\hat{Y}_{OR}$	0.533		0.533	
Fieller *CI* for $\kappa_{OR}(c)/\kappa_1(c)$	$(0.951\ ,\ 1.099)$		$(1.008\ ,\ 1.164)$	
Fieller *CI* for $\kappa_{OR}(c)/\kappa_2(c)$	$(1.225\ ,\ 1.515)$		$(1.325\ ,\ 1.654)$	
Logit *CI* for $\kappa_{OR}(c)$	$(0.635\ ,\ 0.754)$		$(0.676\ ,\ 0.802)$	

3.3.4.6. Discussion

The combination of *BDTs* for incrementing the accuracy of the diagnosis of a disease is common in clinical practice. It consists in combining the results of the two tests using a suitable method and then estimating the parameters of the combination of both *BDTs*. In Section 3.3.4 we study the combination of the weighted kappa coefficients of two *BDTs* when the combination is in parallel testing under the rules *AND* and *OR*. The combined weighted kappa coefficient and its properties have been defined and the conditions under which the combination of the two *BDTs* increases the value of the combined weighted kappa coefficient have been studied. The conditions studied are theoretical

conditions (for the parameters), so that its practical application does not guarantee that the increase in the combined weighted kappa coefficient will be produced, although they can help to give a view of the problem. Thus, in practice, and as we carried out in the example of Weiner et al (1979), it is necessary to study the increase in $\kappa_{\cdot}(c)$ using *CIs*.

The *CIs* proposed are approximate intervals, based on the asymptotic normality of the estimators. The simulation experiments carried out have shown that the coverage probabilities of these *CIs* can exceed the confidence level when the samples are not very large (sizes 100-200), and this may be due to the fact that the convergence of the multinomial distribution (which is the probability distribution inherent of the observed data) with the normal distribution is slow and requiring larger sample sizes. For larger sample sizes, the *CIs* proposed have a coverage probability fluctuating around the 95% confidence level.

An alternative *CI* to Fieller method is obtained by applying the transformation of the naperian logarithm. Assuming that $\ln\left(\hat{\theta}_{\cdot h}\right) \xrightarrow[n\to\infty]{} N\left(0, Var\left[\ln\left(\theta_{\cdot h}\right)\right]\right)$, then an approximate *CI* for $\theta_{\cdot h}$ is $\hat{\theta}_{\cdot h} \times \exp\left\{\pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left[\ln\left(\hat{\theta}_{\cdot h}\right)\right]}\right\}$, where $\hat{V}ar\left[\ln\left(\hat{\theta}_{\cdot h}\right)\right]$ is obtained by applying the delta method in a similar way as is indicated in the equation (3.62) but with $\ln\left(\hat{\theta}_{\cdot h}\right)$. Simulation experiments (similar to those made in Section 3.3.4.3) have been carried out, obtaining that this *CI* has a very similar behaviour to the Fieller method when the sample size is ≥ 300; for samples of sizes 100 or 200, the coverage probability of this *CI* exceeds 95% or does not attain that coverage (and in some cases is below 90%).

We have only studied the combination of the two *BDTs* in parallel testing. A different scenario is when the serial testing is used, in which case the application of *Test* 2 depends on the result of *Test* 1. From a statistical point of view, with the serial testing we cannot estimate [8]: all the probabilities p_{ij} and q_{ij}, the sensitivity and specificity of *Test* 2, and the dependency factors between the two *BDTs*.

Conclusions

The main parameters to estimate and compare the accuracy of binary diagnostic tests are sensitivity and specificity, which only depend on the intrinsic ability of the diagnostic test to distinguish between diseased individuals and non-diseased individuals. Other parameters are likelihood ratios which only depend on the sensitivity and specificity of the binary diagnostic test. Both sensitivity and specificity, like likelihood ratios, are superior measures of the accuracy of a binary diagnostic test, as they depend on the characteristics of the test. The predictive values of a binary diagnostic test are also parameters that are widely used to assess and compare binary diagnostic tests, and they represent the clinical accuracy of the test. The predictive values depend on the sensitivity and specificity of the test and the disease prevalence. When considering the losses or costs of an erroneous classification with the binary diagnostic test, the performance of a diagnostic test is measured through the weighted kappa coefficient. The weighted kappa coefficient depends on the sensitivity and specificity of the diagnostic test, on the disease prevalence and the weighting index c. The weighting index c represents the relative importance between the false positives and

the false negatives. In Chapter 1, all these parameters are defined and their properties are studied.

In Chapter 2, we study the estimations and confidence intervals of the parameters defined in Chapter 1, subject to two types of study or sample: a cross-sectional study and a case-control study. The contribution made by this Chapter is the estimation of the weighted kappa coefficient subject to a case-control study. Different approximate confidence intervals are proposed for the weighted kappa coefficient: a Wald type interval, an interval with logit transformation, an interval with arcsine transformation, an interval using bootstrap and a Bayesian interval. For the first three intervals, we used four corrections which are frequent in this type of studies $\left(h=0,\ 0.5,\ 2,\ z_{1-\alpha/2}^2/2\right)$. As a bootstrap interval, we used a bias-corrected one, and for the Bayesian interval we used an interval based on the Monte Carlo method based on quartiles with distribution which a priori are non-informative and also with distribution which a priori are informative. Simulation experiments to study the asymptotic coverage of these intervals, from which some general rules of application were given: when the sample sizes are small $\left(n_i \le 75\right)$ use the Wald *CI* with $h=0.5$, and for other sample sizes $\left(n_i \ge 100\right)$ use the Wald *CI* with $h=0$. A method has also been proposed to calculate the sample sizes (of cases and of controls) necessary to estimate the weighted kappa coefficient with a determined precision δ to the confidence $100\left(1-\alpha\right)\%$. The results obtained have been applied to a real example of the diagnosis of coronary disease.

Finally, in Chapter 3 we studied the comparison of the parameters defined in Chapter 1 of two binary diagnostic tests subject to a paired design, presenting hypothesis tests to check the equality of the corresponding parameters and *CIs* for the respective differences. We also studied the combination of parameters of two binary diagnostic tests in parallel testing. The contribution of this Chapter is the combination of the weighted kappa coefficients of two binary diagnostic tests in parallel testing, both with the *AND* rule and with the *OR* rule. The combined weighted kappa coefficient was defined when using the *AND* rule and when using the *OR* rule, and we studied its properties, finding them to be similar to those of the weighted kappa coefficient of a binary diagnostic test. We studied the conditions under which the combination of two binary diagnostic tests produces an increase in the combined weighted kappa coefficient. Whether or not an increase occurs in the combined weighted kappa coefficient will depend on the value of the weighting index c set by the researcher, although in some situations such an increase never occurs. These conditions studied are theoretical, and therefore in a practical problem, whether or not there is an increase in the combined weighted kappa coefficient must be studied through a confidence interval. For this purpose, Fieller's theorem was applied, obtaining a confidence interval for the ratio between the combined weighted kappa coefficient and each weighted kappa coefficient. Furthermore, a logit confidence interval was studied for the combined weighted kappa coefficient. Monte Carlo simulation experiments were carried out to study the asymptotic behaviour of the confidence intervals proposed, finding that, in general terms, these intervals have an asymptotic coverage that fluctuates around 95% even with relatively small samples, although in some situations the average coverage

can exceed 95%. An *R* programme was written to solve the problem posed. The results have been applied to a real example on the diagnosis of coronary disease.

Appendices

Appendix 2.1: Estimation of the variance

The variance of the weighted kappa coefficient can be estimated applying the delta method (Agresti, 2002), i.e.

$$Var\left[\hat{\kappa}(c)\right] \approx \left(\frac{\partial \kappa(c)}{\partial Se}\right)^2 Var\left(\hat{Se}\right) + \left(\frac{\partial \kappa(c)}{\partial Sp}\right)^2 Var\left(\hat{Sp}\right),$$

since $Cov\left(\hat{Se}, \hat{Sp}\right) = 0$ (as the sensitivity is estimated from the case sample and the specificity from the control sample, and both samples are independent), and where $Var\left(\hat{Se}\right) = \frac{Se(1-Se)}{n_1}$ and $Var\left(\hat{Sp}\right) = \frac{Sp(1-Sp)}{n_0}$. Carrying out the algebraic operations, it holds that

$$\frac{\partial \kappa(c)}{\partial Se} = \frac{\kappa(c)}{pqY}\left[pq - \kappa(c)\left\{p(q-c)\right\}\right] \text{ and } \frac{\partial \kappa(c)}{\partial Sp} = \frac{\kappa(c)}{pqY}\left[pq + \kappa(c)\left\{q(c-q)\right\}\right],$$

where $Y = Se + Sp - 1$ is the Youden index. Thus, it holds that

$$Var\left[\hat{\kappa}(c)\right] \approx \left(\frac{\kappa(c)}{pqY}\right)^2 \times \left\{\left[pq - \kappa(c)\left\{p(q-c)\right\}\right]^2 \frac{Se(1-Se)}{n_1} + \left[pq + \kappa(c)\left\{q(c-q)\right\}\right]^2 \frac{Sp(1-Sp)}{n_0}\right\}. \quad (3.67)$$

Substituting in this equation each parameter with its estimator and carrying out the algebraic operations we obtain equation (2.71). The variances of the logit of $\kappa(c)$ and of the arcsine are estimated in a similar way to the previous case, i.e.

$$Var\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\} \approx \left(\frac{\partial \text{logit}\left[\kappa(c)\right]}{\partial Se}\right)^2 Var\left(\hat{Se}\right) + \left(\frac{\partial \text{ogit}\left[\kappa(c)\right]}{\partial Sp}\right)^2 Var\left(\hat{Sp}\right)$$

and

$$Var\left\{\arcsin\left[\sqrt{\hat{\kappa}(c)}\right]\right\} \approx \left(\frac{\partial \arcsin\left[\sqrt{\kappa(c)}\right]}{\partial Se}\right)^2 Var\left(\hat{Se}\right) + \left(\frac{\partial \arcsin\left[\sqrt{\kappa(c)}\right]}{\partial Sp}\right)^2 Var\left(\hat{Sp}\right).$$

The expression of the variance obtained here (equation (2.71)) is different to that obtained by Roldán-Nofuentes et al (2009). Equation (2.71) has a term of variability less than the variance deduced by Roldán-Nofuentes et al (2009), because in a case-control study it is necessary to know a value for the prevalence in order to estimate $\kappa(c)$. This means that the *CIs* for $\kappa(c)$ have a length which is lower in the case-control sample than in the cross-sectional sample.

Appendix 2.2: Comparison of the performance of the *CIs*

"Step 1" of this method to choose the optimum *CI* establishes that the probability of coverage should be higher than 93%, or in other words, that the *CI* does not fail. $\Delta\alpha = \alpha - \alpha^* = \gamma^* - \gamma$ is defined, where $\gamma = 1-\alpha = 0.95$ is the nominal confidence of the *CI* and γ^* the probability of coverage calculated. The hypothesis test (two-tailed) for the weighted kappa coefficient is $H_0 : \kappa(c) = \kappa'(c)$ vs $H_1 : \kappa(c) \neq \kappa'(c)$, where $\kappa'(c)$ is the value of $\kappa(c)$ subject to H_0. This test can be solved through different methods. If the test is solved applying the Wald method, the contrast statistic is:

$$z_W = \frac{\left|\hat{\kappa}(c) - \kappa(c)\right|}{\sqrt{\hat{V}ar\left[\hat{\kappa}(c)\right]}}.$$

Other alternatives are to use the logit and arcsine transformations. Applying the logit transformation, the test is $H_0 : \text{logit}\left[\kappa(c)\right] = \text{logit}\left[\kappa'(c)\right]$ vs $H_1 : \text{logit}\left[\kappa(c)\right] \neq \text{logit}\left[\kappa'(c)\right]$, and the contrast statistic is

$$z_L = \frac{\left|\text{logit}\left[\hat{\kappa}(c)\right] - \text{logit}\left[\hat{\kappa}'(c)\right]\right|}{\sqrt{\hat{V}ar\left\{\text{logit}\left[\hat{\kappa}(c)\right]\right\}}},$$

and applying the arcsine transformation, the test $H_0 : \arcsin\left[\kappa(c)\right] = \arcsin\left[\kappa'(c)\right]$ vs $H_1 : \arcsin\left[\kappa(c)\right] \neq \arcsin\left[\kappa'(c)\right]$ is solved with the statistic

$$z_A = \frac{\left|\sin^{-1}\sqrt{\hat{\kappa}(c)} - \sin^{-1}\sqrt{\kappa'(c)}\right|}{\sqrt{\hat{V}ar\left\{\sin^{-1}\sqrt{\hat{\kappa}(c)}\right\}}}.$$

In Step 1 of the method, a *CI* has a failure if its probability of coverage is $\leq 93\%$, i.e. if $\Delta\alpha \leq -2$. In this situation, the type I error of the corresponding hypothesis test is $\geq 7\%$, and therefore it is a very liberal hypothesis test and can give false significances. The 93% criteria has been used by other authors (Agreti and Coull, 1998; Price and Bonnett, 2004; Martín-Andrés and Álvarez-Hernández, 2014a and 2014b). If $\Delta\alpha > 2\%$, i.e. the probability of coverage is greater than 97%, then the corresponding hypothesis test is very conservative (its type I error is very small, $< 3\%$), but at least it does not give false significances. Consequently, the choice of the optimum *CI* is linked to the decisions of the corresponding hypothesis test, and it is preferable to choose a conservative test rather than a very liberal one (then there will not be any false significances as its type I error is lower than the nominal one).

Appendix 3.1: Combined weighted kappa coefficient

The weighted kappa coefficient of *BDT* is defined [1] as

$$\kappa = \frac{\text{Random loss} - \text{Expected loss}}{\text{Random loss}}. \tag{3.68}$$

The expected loss is the average loss which occurs when erroneously classifying a diseased or non-diseased patient with the *BDT*. The random expected loss is the expected loss which occurs when the *BDT* and the *GS* are independent, i.e. when $P(T=i|D=j) = P(T=i)$. In parallel testing with the *AND* rule, the random loss is

$$\text{Random loss} = Lp(p_{10} + p_{01} + p_{00} + q_{10} + q_{01} + q_{00}) + L'q(p_{11} + q_{11}) = Lp(1 - p_{11} - q_{11}) + L'q(p_{11} + q_{11}),$$

and the expected loss is

$$\text{Expected loss} = L\left(p_{10} + p_{01} + p_{00}\right) + L'q_{11} = L\left(p - p_{11}\right) + L'q_{11}.$$

By substituting the two previous expressions in equation (3.68) and realizing the algebraic operations, equation (3.56) is obtained. With the *OR* rule, the random loss is

$$\text{Random loss} = Lp\left(p_{00} + q_{00}\right) + L'q\left(p_{11} + p_{10} + p_{01} + q_{11} + q_{10} + q_{01}\right) = Lp\left(p_{00} + q_{00}\right) + L'q\left(1 - p_{00} - q_{00}\right),$$

and the expected loss is

$$\text{Expected loss} = Lp_{00} + L'\left(q_{11} + q_{01} + q_{00}\right) = Lp_{00} + L'\left(q - q_{00}\right).$$

By substituting these two previous expressions in equation (3.68) and realizing the algebraic operations, equation (3.57) is obtained.

Appendix 3.2: The *AND* rule

In all of the following, it will be taken that all the Youden indices are between 0 and 1 $\left(0 < Y < 1\right)$, and that the prevalence of the disease is also a value between 0 and 1 $\left(0 < p < 1\right)$. Let us consider the following equalities:

$$\omega_1 = Q_1 Y_{AND} - Q_{AND} Y_1 = \frac{p_{11}q_{10} - p_{10}q_{11}}{pq},\ \omega_2 = Q_2 Y_{AND} - Q_{AND} Y_2 = \frac{p_{11}q_{01} - p_{01}q_{11}}{pq},$$

$$\xi_1 = Y_{AND} - Y_1 = \frac{q_{10}p - p_{10}q}{pq},\ \xi_2 = Y_{AND} - Y_2 = \frac{q_{01}p - p_{01}q}{pq},$$

$$Q_1 = p_{11} + p_{10} + q_{11} + q_{10},\ Q_2 = p_{11} + p_{01} + q_{11} + q_{01},$$

$$\tau_1 = Q_1 - Q_{AND} = p_{10} + q_{10},\ \tau_2 = Q_2 - Q_{AND} = p_{01} + q_{01},$$

$$\nu_{AND} = pc\left(1-Q_{AND}\right) + q\left(1-c\right)Q_{AND} \text{ and } \nu_h = pc\left(1-Q_h\right) + q\left(1-c\right)Q_h,\ h = 1,2,$$

where ν_{AND} and ν_h are the denominators of $\kappa_{AND}(c)$ and $\kappa_h(c)$ respectively. It is verified that $-1 < \omega_h < 1$, $-1 < \xi_h < 1$, $0 < Q_h < 1$ and $0 < \tau_h < 1$. By realizing algebraic operations it is verified that

$$\kappa_{AND}(c) - \kappa_h(c) = \frac{pq\xi_h - (c-q)\tau_h\kappa_h(c)}{\nu_{AND}} = pq\frac{\xi_h\nu_h - (c-q)\tau_h Y_h}{\nu_{AND}\nu_h}. \qquad (3.69)$$

Then $\kappa_{AND}(c) = \kappa_h(c)$ if

$$c = c_h = \frac{q\left(Q_h Y_{AND} - Q_{AND}Y_h\right)}{Y_h\left(p - Q_{AND}\right) - Y_{AND}\left(p - Q_h\right)} = \frac{q\omega_h}{\omega_h - p\xi_h} = \frac{q\omega_h}{\psi_h},\ h = 1,2, \qquad (3.70)$$

where $\psi_h = Y_h\left(p - Q_{AND}\right) - Y_{AND}\left(p - Q_h\right) = \omega_h - p\xi_h$. The values c_1 and c_2 can take any real value. In terms of c_h the equation (3.69) is written as

$$\kappa_{AND}(c) - \kappa_h(c) = \frac{pq^2\omega_h\left(c_h - c\right)}{c_h\nu_{AND}\nu_h}. \qquad (3.71)$$

In what follows the conditions under which the combination of the two *BDTs* with the *AND* rule produces an increase of the combined weighted kappa coefficient are shown.

a) If $0 < c_h < 1$ with $h = 1,2$, then $\kappa_h(c_h)$ and $\kappa_{AND}(c_h)$, can be calculated, so obtaining $\kappa_h(c_h) = \dfrac{Y_h}{\left(\omega_h - \xi_h Q_h\right)/\psi_h}$ and $\kappa_{AND}(c_h) = \dfrac{Y_{AND}}{\left(\omega_h - \xi_h Q_{AND}\right)/\psi_h}$. It is verified that

$$\frac{\omega_h - \xi_h Q_h}{\psi_h} > 0 \text{ and } \frac{\omega_h - \xi_h Q_{AND}}{\psi_h} > 0,$$

since both expressions are the denominators of weighted kappa coefficients. As it is always verified that $Q_h > Q_{AND}$, if $\xi_h = Y_{AND} - Y_h > 0$ then

$$\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} > 0.$$

By realizing algebraic operations one obtains

$$\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} = \frac{\xi_h \tau_h}{\psi_h} > 0,$$

and as ξ_h and τ_h are greater than 0, then $\psi_h > 0$. If $\xi_h = Y_{AND} - Y_h < 0$ then

$$\frac{\omega_h - \xi_h Q_h}{\psi_h} - \frac{\omega_h - \xi_h Q_{AND}}{\psi_h} > 0,$$

and by realizing algebraic operations one obtains

$$\frac{\omega_h - \xi_h Q_h}{\psi_h} - \frac{\omega_h - \xi_h Q_{AND}}{\psi_h} = \frac{-\xi_h \tau_h}{\psi_h} > 0.$$

As $-\xi_h$ and τ_h are greater than 0, then $\psi_h > 0$. Thus, whatever the value of ξ_h, it is always verified that $\psi_h > 0$. As $c_h = \frac{q\omega_h}{\psi_h} > 0$ then $\omega_h > 0$. If $0 \le c < c_h$ then, from equation (3.71), $\kappa_{AND}(c) > \kappa_h(c)$ with $h = 1,2$. Finally, if $0 \le c < \min\{c_1, c_2\}$ then $\kappa_{AND}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$.

If $c = 1$ then $c_h - 1 < 0$, and from equation (3.71) it is verified that $\kappa_{AND}(1) < \kappa_h(1)$ with $h = 1,2$. Therefore $\kappa_{AND}(1) < \min\{\kappa_1(1), \kappa_2(c)\}$.

b) From *a)* it can be verified that $\kappa_{AND}(c) > \kappa_i(c)$ if $0 \le c < c_i < 1$. Let us suppose that $Y_j < Y_{AND}$, then $\xi_j = Y_{AND} - Y_j > 0$. By multiplying by Q_j one obtains

$$\xi_j Q_j > 0 \Rightarrow Q_j Y_{AND} - Q_j Y_j = Q_j Y_{AND} - \left(\tau_j + Q_{AND}\right) Y_j = \omega_j - \tau_j Y_j > 0 \Rightarrow \omega_j > \tau_j Y_j > 0,$$

and by multiplying by Q_{AND} one obtains

$$\xi_j Q_{AND} > 0 \Rightarrow Q_{AND} Y_{AND} - Q_{AND} Y_j = \left(Q_j - \tau_j\right) Y_{AND} - Q_{AND} Y_j = \omega_j - \tau_j Y_{AND} > 0$$
$$\Rightarrow \omega_j > \tau_j Y_{AND} > 0.$$

Hence, it can be verified that $\omega_j > \tau_j Y_{AND} > \tau_j Y_j > 0$. As $c_j > 1$ and $\omega_j > 0$, then $\psi_j = \omega_j - p\xi_j$ must be greater than 0, and it is verified that $\xi_j > \omega_j$. Finally, as $\omega_j > 0$ then, from equation (3.71), it is verified that $\kappa_{AND}(c) > \kappa_j(c)$ for $0 \le c \le 1$.

In the case that $Y_j > Y_{AND}$, then $\xi_j < 0$. Let us suppose that $\omega_j > 0$, then as $c_j > 1$ it must be verified that $\omega_j - p\xi_j > 0$, and so one obtains

$$q\omega_j > \omega_j - p\xi_j \Rightarrow \xi_j > \omega_j > 0,$$

which is contradictory to $\xi_j < 0$. Thus, if $Y_j > Y_{AND}$ then $\omega_j < 0$, and from equation (3.71) it can be verified that $\kappa_j(c) > \kappa_{AND}(c)$ for $0 \le c \le 1$.

If $c_j < 0$ the reasoning is the same as for $c_j > 1$. In this case $\omega_j - p\xi_j < 0$, so that it is also verified that $\xi_j > p\xi_j > \omega_j$.

For any situation different to those that have been dealt with in points *a)* and *b)*, the combination of the two *BDTs* does not increase the value of $\kappa_{AND}(c)$ for any value of *c*. These situations are:

1) $0 \le c < c_i < 1$, $c_j > 1$ (or $c_j < 0$) and $Y_{AND} \le Y_j$,

2) $\min\{c_1, c_2\} > 1$,

3) $\max\{c_1, c_2\} < 0$,

4) $c_i > 1$ and $c_j < 0$.

In the following each of these is analysed.

1) The situation $0 \le c < c_i < 1$, $c_j > 1$ (or $c_j < 0$) and $Y_{AND} < Y_j$ has been dealt with in point *b*) above. Let us study the same situation but supposing that $Y_{AND} = Y_j$. If $Y_{AND} = Y_j$ then $\xi_j = 0$. As $c_j = \dfrac{q\omega_j}{\omega_j - p\xi_j}$, if $\xi_j = 0$ then $c_j = q$, but this is incompatible with $c_j > 1$ or $c_j < 0$, because $0 < q < 1$.

2) Let us consider that $c_1 > 1$ and $c_2 > 1$, i.e. $\min\{c_1, c_2\} > 1$. Each equation $c_h > 1$, $h = 1, 2$, has two solutions:

$$0 < p\xi_h < \omega_h < \xi_h < 1 \tag{3.72}$$

and

$$-1 < \xi_h < \omega_h < p\xi_h < 0. \tag{3.73}$$

Let us consider

$$c_h(1 - c_h) = pq\frac{\omega_h(\omega_h - \xi_h)}{\psi_h^2} < 0,$$

then $c_1(1 - c_1)c_2(1 - c_2) > 0$, i.e.

$$c_1(1-c_1)c_2(1-c_2) = p^2q^2\frac{\omega_1\omega_2(\omega_1-\xi_1)(\omega_2-\xi_2)}{\psi_1^2\psi_2^2} > 0. \tag{3.74}$$

From equation (3.74) it is verified that $c_1(1-c_1)c_2(1-c_2) > 0$ if

$$\omega_1\omega_2(\omega_1-\xi_1)(\omega_2-\xi_2) > 0. \tag{3.75}$$

Moreover,

$$c_1c_2 = q^2\frac{\omega_1\omega_2}{\psi_1\psi_2} > 1. \tag{3.76}$$

Solving the system formed by equations (3.75) and (3.76), the solutions for this system which are compatible with solutions (3.72) and (3.73) are

$$0 < p\xi_1 < \omega_1 < \xi_1 < 1 \text{ and } -1 < \xi_2 < \omega_2 < p\xi_2 < 0$$

and

$$\xi_1 < \omega_1 < p\xi_1 < 0 \text{ and } 0 < p\xi_2 < \omega_2 < \xi_2.$$

In general terms, the solution is

$$-1 < \xi_j < \omega_j < p\xi_j < 0 < p\xi_i < \omega_i < \xi_i < 1.$$

As $\omega_j < 0$ y $c_j > 1$, then applying equation (3.34) it holds that $\kappa_{AND}(c) < \kappa_j(c)$ for $0 \le c \le 1$. As $\omega_i > 0$ and $c_i > 1$, then $\kappa_{AND}(c) > \kappa_i(c)$ for $0 \le c \le 1$. Finally,

$$\kappa_i(c) < \kappa_{AND}(c) < \kappa_j(c), \quad 0 \le c \le 1.$$

Consequently, if $\min\{c_1, c_2\} > 1$ then the combination of the two weighted kappa coefficients never increases the value of the combined weighted kappa coefficient. Furthermore, as $\xi_j < 0$ then $Y_{AND} < Y_j$, and as $\xi_i > 0$ then $Y_{AND} > Y_i$, i.e.

$$Y_i < Y_{AND} < Y_j .$$

3) Another possible situation is that $c_1 < 0$ and $c_2 < 0$, i.e. $\max\{c_1, c_2\} < 0$, then the solutions for each equation $c_h < 0$, $h = 1,2$, are $0 < \omega_h < p\xi_h < \xi_h < 1$ or $-1 < \xi_h < p\xi_h < \omega_h < 0$. Solving the system formed by the equations

$$c_1(1-c_1)c_2(1-c_2) = p^2q^2\frac{\omega_1\omega_2(\omega_1-\xi_1)(\omega_2-\xi_2)}{\psi_1^2\psi_2^2} > 0$$

and

$$c_1c_2 = q^2\frac{\omega_1\omega_2}{\psi_1\psi_2} > 0,$$

the appropriate solution is $-1 < \xi_j < p\xi_j < \omega_j < 0 < \omega_i < p\xi_i < \xi_i < 1$. Finally, it holds that

$$\kappa_i(c) < \kappa_{AND}(c) < \kappa_j(c), \quad 0 \le c \le 1,$$

and

$$Y_i < Y_{AND} < Y_j .$$

4) Finally, another possible situation is that $c_i > 1$ and $c_j < 0$. Solving the system formed by the equations

$$c_1(1-c_1)c_2(1-c_2) = p^2q^2\frac{\omega_1\omega_2(\omega_1-\xi_1)(\omega_2-\xi_2)}{\psi_1^2\psi_2^2} > 0$$

and

$$c_i c_j = q^2 \frac{\omega_i \omega_j}{\psi_i \psi_j} < 0,$$

the appropriate solution is $-1 < \xi_j < p\xi_j < \omega_j < 0 < p\xi_i < \omega_i < \xi_i < 1$. In this situation, it is verified that if $\xi_j < 0$ then $-1 < \xi_j < p\xi_j < \omega_j < q\omega_j < 0$ and $\psi_j = \omega_j - p\xi_j > 0$, and therefore $c_j < 0$. Consequently, if $c_i > 1$ and $c_j < 0$, the solution is $-1 < \xi_j < p\xi_j < \omega_j < 0 < p\xi_i < \omega_i < \xi_i < 1$, and it is again verified that

$$\kappa_i(c) < \kappa_{AND}(c) < \kappa_j(c), \quad 0 \le c \le 1,$$

and

$$Y_i < Y_{AND} < Y_j.$$

Appendix 3.3: The *OR* rule

In all of the following, it will be taken that all the Youden indices are between 0 and 1 $(0 < Y < 1)$, and that the prevalence of the disease is also a value between 0 and 1 $(0 < p < 1)$. Let us take the following equalities

$$\omega_1' = Q_1 Y_{OR} - Q_{OR} Y_1 = \frac{p_{01}(q_{11} + q_{10}) - q_{01}(p_{11} + p_{10})}{pq},$$

$$\omega_2' = Q_2 Y_{OR} - Q_{OR} Y_2 = \frac{p_{10}(q_{11} + q_{01}) - q_{10}(p_{11} + p_{01})}{pq},$$

$$\xi_1' = Y_{OR} - Y_1 = \frac{p_{01}q - q_{01}p}{pq},\ \xi_2' = Y_{OR} - Y_2 = \frac{p_{10}q - q_{10}p}{pq},\ \tau_1' = Q_1 - Q_{OR} = -\left(p_{01} + q_{01}\right),$$

$$\tau_2' = Q_2 - Q_{OR} = -\left(p_{10} + q_{10}\right) \text{ and } \nu_{OR} = pc\left(1 - Q_{OR}\right) + q\left(1 - c\right)Q_{OR}.$$

It is verified that $-1 < \omega_h' < 1$, $-1 < \xi_h' < 1$ and $-1 < \tau_h' < 0$. By realizing algebraic operations it can be verified that

$$\kappa_{OR}\left(c\right) - \kappa_h\left(c\right) = \frac{pq\xi_h' - \left(c - q\right)\tau_h'\kappa_h\left(c\right)}{\nu_{OR}} = pq\frac{\xi_h'\nu_h - \left(c - q\right)\tau_h'Y_h}{\nu_{OR}\nu_h}, \tag{3.77}$$

so verifying that $\kappa_{OR}\left(c\right) = \kappa_h\left(c\right)$ if

$$c = c_h' = \frac{q\left(Q_hY_{OR} - Q_{OR}Y_h\right)}{Y_h\left(p - Q_{OR}\right) - Y_{OR}\left(p - Q_h\right)} = \frac{q\omega_h'}{\omega_h' - p\xi_h'} = \frac{q\omega_h'}{\psi_h'},\ h = 1,2, \tag{3.78}$$

with $\psi_h' = Y_h\left(p - Q_{OR}\right) - Y_{OR}\left(p - Q_h\right) = \omega_h' - p\xi_h'$. The values c_1' and c_2' may take any real value. In terms of c_h', equation (3.77) is written as

$$\kappa_{OR}\left(c\right) - \kappa_h\left(c\right) = \frac{pq^2\omega_h'\left(c_h' - c\right)}{c_h'\nu_{OR}\nu_h}. \tag{3.79}$$

In what follows the conditions under which the combination of the two *BDTs* with the *AND* rule produces an increase of the combined weighted kappa coefficient are shown.

a) If $0 < c_h' < 1$ with $h = 1,2$, then $\kappa_h\left(c_h\right)$ and $\kappa_{OR}\left(c_h\right)$, can be calculated, so obtaining $\kappa_h\left(c_h'\right) = \dfrac{Y_h}{\left(\omega_h' - \xi_h'Q_h\right)/\psi_h'}$ and $\kappa_{OR}\left(c_h\right) = \dfrac{Y_{OR}}{\left(\omega_h' - \xi_h'Q_{OR}\right)/\psi_h'}$. It is verified that

$$\frac{\omega_h' - \xi_h'Q_h}{\psi_h'} > 0 \text{ and } \frac{\omega_h' - \xi_h'Q_{OR}}{\psi_h'} > 0,$$

as both expressions are the denominators of weighted kappa coefficients. As it is always verified that $Q_{OR} > Q_h$, if $\xi'_h = Y_{OR} - Y_h > 0$ then

$$\frac{\omega'_h - \xi'_h Q_{OR}}{\psi'_h} - \frac{\omega'_h - \xi'_h Q_h}{\psi'_h} > 0 .$$

By realizing algebraic operations one obtains

$$\frac{\omega'_h - \xi'_h Q_{OR}}{\psi'_h} - \frac{\omega'_h - \xi'_h Q_h}{\psi'_h} = \frac{\xi'_h \tau_h}{\psi'_h} > 0 ,$$

and as $\xi'_h > 0$ and $\tau_h < 0$, then $\psi_h < 0$. If $\xi_h = Y_{OR} - Y_h < 0$ then

$$\frac{\omega'_h - \xi'_h Q_{OR}}{\psi'_h} - \frac{\omega'_h - \xi'_h Q_h}{\psi'_h} < 0 ,$$

and by realizing algebraic operations one obtains

$$\frac{\omega'_h - \xi'_h Q_{OR}}{\psi'_h} - \frac{\omega'_h - \xi'_h Q_h}{\psi'_h} = \frac{\xi'_h \tau'_h}{\psi'_h} < 0 .$$

As $\xi'_h < 0$ and $\tau'_h < 0$, then $\psi'_h < 0$. Thus, whatever the value of ξ'_h, it is always verified that $\psi'_h < 0$. As $c'_h = \frac{q\omega'_h}{\psi'_h} > 0$ then $\omega'_h < 0$. If $0 < c'_h < c \le 1$ then, from equation (3.79), $\kappa_{OR}(c) > \kappa_h(c)$ with $h = 1,2$. Finally, if $0 < \max\{c'_1, c'_2\} < c \le 1$ then $\kappa_{OR}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$.

If $c = 0$ then $c'_h - 1 < 0$, and from equation (3.79) it is verified that $\kappa_{OR}(0) < \kappa_h(0)$ with $h = 1,2$. Therefore $\kappa_{OR}(0) < \min\{\kappa_1(0), \kappa_2(0)\}$.

b) From *a)* it can be verified that $\kappa_{OR}(c) > \kappa_i(c)$ if $0 < c_i' < c \leq 1$. Let us suppose that $c_j' > 1$ and that $Y_j < Y_{OR}$, then $\xi_j' = Y_{OR} - Y_j > 0$. Let us suppose that $\omega_j' < 0$, as $c_j' > 1$ then $\psi_j' = \omega_j' - p\xi_j'$ must be smaller than 0 and $q\omega_j' < \omega_j' - p\xi_j' < 0$, and therefore it is verified that $p\left(\xi_j' - \omega_j'\right) < 0$ and therefore $\xi_j' < \omega_j'$. As $\xi_j' > 0$ it is not possible that $\omega_j' < 0$. Finally $\omega_j' > 0$ and from equation (3.79) it is verified that $\kappa_{OR}(c) > \kappa_j(c)$ for $0 \leq c \leq 1$.

In the case that $Y_j > Y_{OR}$, then $\xi_j < 0$. By multiplying by Q_j one obtains

$$\xi_j' Q_j < 0 \Rightarrow Q_j Y_{OR} - Q_j Y_j = Q_j Y_{OR} - \left(\tau_j' + Q_{OR}\right) Y_j = \omega_j' - \tau_j' Y_j < 0 \Rightarrow \omega_j' < \tau_j' Y_j < 0,$$

since $\tau_j' < 0$. Multiplying by Q_{OR} one obtains

$$\xi_j' Q_{OR} < 0 \Rightarrow Q_{OR} Y_{OR} - Q_{OR} Y_j = \left(Q_j - \tau_j'\right) Y_{OR} - Q_{OR} Y_j = \omega_j' - \tau_j' Y_{OR} < 0 \Rightarrow \omega_j' < \tau_j' Y_{OR} < 0.$$

Hence, it is verified that $\omega_j' < \tau_j' Y_j < \tau_j' Y_{OR} < 0$. Therefore, if $Y_j < Y_{OR}$ then $\omega_j' < 0$, and from equation (3.79) it can be verified that $\kappa_j(c) > \kappa_{OR}(c)$ for $0 \leq c \leq 1$.

If $c_j' < 0$ the reasoning is the same as for $c_j' > 1$. If $\xi_j > 0$ then $\omega_j' > 0$, and from equation (3.79) it is verified that $\kappa_{OR}(c) > \kappa_j(c)$ for $0 \leq c \leq 1$. Therefore, if $\xi_j < 0$ then $\omega_j' < 0$, and from equation (3.79) it can be verified that $\kappa_j(c) > \kappa_{OR}(c)$ for $0 \leq c \leq 1$.

For any situation different to those that have been dealt with in points *a)* and *b)*, the combination of the two *BDTs* does not increase the value of $\kappa_{OR}(c)$ for any value of *c*. These situations are:

1) $0 < c_i' < c \le 1$, $c_j' > 1$ (or $c_j' < 0$) and $Y_{OR} \le Y_j$,

2) $\min\{c_1', c_2'\} > 1$,

3) $\max\{c_1', c_2'\} < 0$,

4) $c_i' > 1$ and $c_j' < 0$.

In the following each of these is analysed.

1) The situation $0 < c_i' < c \le 1$, $c_j' > 1$ (or $c_j' < 0$) and $Y_{OR} < Y_j$ has been dealt with in point *b*) above. Let us study the same situation but supposing that $Y_{OR} = Y_j$. If $Y_{OR} = Y_j$ then $\xi_j' = 0$. As $c_j' = \dfrac{q\omega_j'}{\omega_j' - p\xi_j'}$, if $\xi_j' = 0$ then $c_j' = q$, but this is incompatible with $c_j' > 1$ or $c_j' < 0$, because $0 < q < 1$.

2) Let us consider that $c_1' > 1$ and $c_2' > 1$, i.e. $\min\{c_1', c_2'\} > 1$. Each equation $c_h' > 1$, $h = 1,2$, has two solutions:

$$0 < p\xi_h' < \omega_h' < \xi_h' < 1 \tag{3.80}$$

and

$$-1 < \xi_h' < \omega_h' < p\xi_h' < 0. \tag{3.81}$$

Let us consider

$$c_h'(1 - c_h') = pq\frac{\omega_h'(\omega_h' - \xi_h')}{\psi_h'^2} < 0,$$

then $c_1'(1 - c_1')c_2'(1 - c_2') > 0$, i.e.

$$c_1'\left(1-c_1'\right)c_2'\left(1-c_2'\right)=p^2q^2\,\frac{\omega_1'\omega_2'\left(\omega_1'-\xi_1'\right)\left(\omega_2'-\xi_2'\right)}{\psi_1'^2\psi_2'^2}>0\,. \tag{3.82}$$

From equation (3.74) it is verified that $c_1'\left(1-c_1'\right)c_2'\left(1-c_2'\right)>0$ if

$$\omega_1'\omega_2'\left(\omega_1'-\xi_1'\right)\left(\omega_2'-\xi_2'\right)>0\,. \tag{3.83}$$

Moreover,

$$c_1'c_2'=q^2\,\frac{\omega_1'\omega_2'}{\psi_1'\psi_2'}>1. \tag{3.84}$$

Solving the system formed by equations (3.83) and (3.84) with $0<p<1$, the solutions for this system which are compatible with solutions (3.80) and (3.81) are

$$0<p\xi_1'<\omega_1'<\xi_1'<1 \text{ and } -1<\xi_2'<\omega_2'<p\xi_2'<0$$

and

$$-1<\xi_1'<\omega_1'<p\xi_1'<0 \text{ and } 0<p\xi_2'<\omega_2'<\xi_2'<1\,.$$

In general terms, the solution is

$$-1<\xi_j'<\omega_j'<p\xi_j'<0<p\xi_i'<\omega_i'<\xi_i'<1\,.$$

As $\omega_j'<0$ y $c_j'>1$, then applying equation (3.79) it holds that $\kappa_{OR}\left(c\right)<\kappa_j\left(c\right)$ for $0\le c\le 1$. As $\omega_i'>0$ and $c_i'>1$, then $\kappa_{OR}\left(c\right)>\kappa_i\left(c\right)$ for $0\le c\le 1$. Finally,

$$\kappa_i\left(c\right)<\kappa_{OR}\left(c\right)<\kappa_j\left(c\right),\;\; 0\le c\le 1\,.$$

Consequently, if $\min\left\{c_1',c_2'\right\}>1$ then the combination of the two weighted kappa coefficients never increases the value of the combined weighted kappa coefficient. Furthermore, as $\xi_j'<0$ then $Y_{OR}<Y_j$, and as $\xi_i'>0$ then $Y_{OR}>Y_i$, i.e.

$$Y_i < Y_{OR} < Y_j .$$

3) Another possible situation is that $c_1' < 0$ and $c_2' < 0$, i.e. $\max\{c_1', c_2'\} < 0$, then the solutions for each equation $c_h' < 0$, $h = 1, 2$, are $0 < \omega_h' < p\xi_h' < \xi_h' < 1$ or $-1 < \xi_h' < p\xi_h' < \omega_h' < 0$. Solving the system formed by the equations

$$c_1'(1-c_1')c_2'(1-c_2') = p^2q^2 \frac{\omega_1'\omega_2'(\omega_1' - \xi_1')(\omega_2' - \xi_2')}{\psi_1'^2\psi_2'^2} > 0$$

and

$$c_1'c_2' = q^2 \frac{\omega_1'\omega_2'}{\psi_1'\psi_2'} > 0,$$

the appropriate solution is $-1 < \xi_j' < p\xi_j' < \omega_j' < 0 < \omega_i' < p\xi_i' < \xi_i' < 1$. Finally, it holds that

$$\kappa_i(c) < \kappa_{OR}(c) < \kappa_j(c), \quad 0 \le c \le 1,$$

and

$$Y_i < Y_{OR} < Y_j .$$

4) Finally, another possible situation is that $c_i > 1$ and $c_j < 0$. Solving the system formed by the equations

$$c_1'(1-c_1')c_2'(1-c_2') = p^2q^2 \frac{\omega_1'\omega_2'(\omega_1' - \xi_1')(\omega_2' - \xi_2')}{\psi_1'^2\psi_2'^2} > 0$$

and

$$c_1'c_2' = q^2 \frac{\omega_1'\omega_2'}{\psi_1'\psi_2'} < 0,$$

the solution is $-1 < \xi_j' < p\xi_j' < \omega_j' < 0 < p\xi_i' < \omega_i' < \xi_i' < 1$. In this situation, it is verified that if $\xi_j' < 0$ then $p\xi_j' < \omega_j' < q\omega_j' < 0$ and $\psi'_j = \omega_j' - p\xi_j' > 0$, and therefore $c_j' < 0$. Consequently, if $c_i' > 1$ and $c_j' < 0$, the solution is $-1 < \xi_j' < p\xi_j' < \omega_j' < 0 < p\xi_i' < \omega_i' < \xi_i' < 1$, and it is again verified that

$$\kappa_i(c) < \kappa_{OR}(c) < \kappa_j(c), \quad 0 \le c \le 1,$$

and

$$Y_i < Y_{OR} < Y_j.$$

Appendix 3.4: The program "cwkc"

```
cwkc <- function(s11, s10, s01, s00, r11, r10, r01, r00, indexc, rule, conflevel = 0.95)
{
 library(graphics)

 if (s11 < 0 | s10 < 0 | s01 < 0 | s00 < 0 | r11 < 0 | r10 < 0 | r01 < 0 | r00 < 0)
   {
    cat("\n")
    stop("Any observed frequency can be negative. Introduces new values \n")
    cat("\n")
   }
```

```
 if (abs(s00 - trunc (s00)) > 0  | abs(s01 - trunc (s01)) > 0  | abs(s10 - trunc (s10)) > 0  | abs(s11 - trunc (s11)) > 0  | abs(r00 - trunc (r00)) > 0  | abs(r01 - trunc (r01)) > 0  | abs(r10 - trunc (r10)) > 0  | abs(r11 - trunc (r11)) > 0)
   {
    cat("\n")
    stop("Observed frequencies can not have decimals. Introduces new values \n")
    cat("\n")
   }

 if ((s11 + s10 + s01 + s00) == 0 | (r11 + r10 + r01 + r00) == 0)
   {
    cat("\n")
    stop("Accuracy of a Binary Test cannot be estimated. There are many observed frequencies equal to zero. Introduces new values \n")
    cat("\n")
   }

 if (indexc > 1 | indexc < 0)
   {
    cat("\n")
    stop("Weighting index c should take a value between 0 and 1. Introduces a new value \n")
    cat("\n")
   }

 n <- s00 + s01 + s10 + s11 + r00 + r01 + r10 + r11

 p <- (s00 + s01 + s10 + s11) / n

 Se1 <- (s11 + s10) / (s11 + s10 + s01 + s00)

 Se2 <- (s11 + s01) / (s11 + s10 + s01 + s00)
```

```
Sp1 <- (r01 + r00) / (r11 + r10 + r01 + r00)

Sp2 <- (r10 + r00) / (r11 + r10 + r01 + r00)

Y1 <- Se1 + Sp1 - 1

Y2 <- Se2 + Sp2 - 1

if (Y1 <= 0 | Y2 <= 0)
  {
   cat("\n")
   cat("Estimated Youden index of Binary Test 1 is ",Y1, "\n")
   cat("Estimated Youden index of Binary Test 2 is ",Y2, "\n")
   stop("Estimated Youden index of a Binary Test must be greater than zero. Introduces new values \n")
   cat("\n")
  }

 if (Y1 == 1 | Y2 == 1)
  {
   cat("\n")
   cat("Estimated Youden index of Binary Test 1 is ",Y1, "\n")
   cat("Estimated Youden index of Binary Test 2 is ",Y2, "\n")
   stop("A Binary Test is a gold standard. Introduces new values \n")
   cat("\n")
  }

y <- as.character(rule)

y1 <- "and"
y2 <- "or"
```

```
if(y !=  y1 && y != y2)
  {
   cat("\n")
   stop("Write and for AND rule. Write or for OR rule  \n")
   cat("\n")
  }

r1 <- identical (y,y1)

r2 <- identical (y, y2)

if(isTRUE(r1))
  {
   kappacomb <- expression((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01
+ p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00)
* (1- indexc) * (p11 + q11)))

   quot1 <- expression(log((((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01
+ p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00)
* (1- indexc) * (p11 + q11))) / (((p10 + p11) * (q00 + q01) - (p00 + p01) * (q10 + q11))
/ ((p11 + p10 + p01 + p00) * indexc * (1- p11 - p10 - q11 - q10) + (q11 + q10 + q01 +
q00) * (1 - indexc)* (p11 + p10 + q11 + q10))))))

   quot2 <- expression(log(((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01
+ p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00)
* (1- indexc) * (p11 + q11))) / (((p01 + p11) * (q00 + q10) - (p00 + p10) * (q01 + q11))
/ ((p11 + p10 + p01 + p00) * indexc * (1- p11 - p01 - q11 - q01) + (q11 + q10 + q01 +
q00) * (1 - indexc)* (p11 + p01 + q11 + q01)))))

   arcsink <- expression(asin(sqrt((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 +
p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 +
q00) * (1- indexc) * (p11 + q11)))))
```

```
    logitk <- expression(log(((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01
+ p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00)
* (1- indexc) * (p11 + q11))) / (1 - ((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10
+ p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01
+ q00) * (1- indexc) * (p11 + q11))))))

  }
 else
  {
    kappacomb <- expression((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01
+ q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) *
(1- indexc) * (1 - p00 - q00)))

    quot1 <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01
+ q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) *
(1- indexc) * (1 - p00 - q00))) / (((p10 + p11) * (q00 + q01) - (p00 + p01) * (q10 +
q11)) / ((p11 + p10 + p01 + p00) * indexc * (1- p11 - p10 - q11 - q10) + (q11 + q10 +
q01 + q00) * (1 - indexc)* (p11 + p10 + q11 + q10)))))

    quot2 <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01
+ q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) *
(1- indexc) * (1 - p00 - q00))) / (((p01 + p11) * (q00 + q10) - (p00 + p10) * (q01 +
q11)) / ((p11 + p10 + p01 + p00) * indexc * (1- p11 - p01 - q11 - q01) + (q11 + q10 +
q01 + q00) * (1 - indexc)* (p11 + p01 + q11 + q01)))))

    arcsink <- expression(asin(sqrt(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10
+ q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 +
q00) * (1- indexc) * (1 - p00 - q00))))))
    logitk <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01
+ q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) *
(1- indexc) * (1 - p00 - q00))) / (1 - ((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10
```

```
+ q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 +
q00) * (1- indexc) * (1 - p00 - q00))))))
   }

  Q1 <- p * Se1 + (1 - p) * (1 - Sp1)

  k1 <- (p * (1 - p) * Y1) / (p * indexc * (1 - Q1) + (1 - p) * (1 - indexc) * Q1)

  Q2 <- p * Se2 + (1 - p) * (1 - Sp2)

  k2 <- (p * (1 - p) * Y2) / (p * indexc * (1 - Q2) + (1 - p) * (1 - indexc) * Q2)

  z  = qnorm((1 + conflevel)/2,0,1)

  alpha <- 1 - conflevel

  kappa1 <- expression(((p10 + p11) * (q00 + q01) - (p00 + p01) * (q10 + q11)) / ((p11
+ p10 + p01 + p00) * indexc * (1- p11 - p10 - q11 - q10) + (q11 + q10 + q01 + q00) *
(1 - indexc)* (p11 + p10 + q11 + q10)))

  kappa2 <- expression(((p01 + p11) * (q00 + q10) - (p00 + p10) * (q01 + q11)) / ((p11
+ p10 + p01 + p00) * indexc * (1- p11 - p01 - q11 - q01) + (q11 + q10 + q01 + q00) *
(1 - indexc)* (p11 + p01 + q11 + q01)))

# Derivatives

  derivk1p00 <- deriv(kappa1, "p00")
  derivk1p01 <- deriv(kappa1, "p01")
  derivk1p10 <- deriv(kappa1, "p10")
  derivk1p11 <- deriv(kappa1, "p11")
  derivk1q00 <- deriv(kappa1, "q00")
  derivk1q01 <- deriv(kappa1, "q01")
```

```
derivk1q10 <- deriv(kappa1, "q10")
derivk1q11 <- deriv(kappa1, "q11")

derivk2p00 <- deriv(kappa2, "p00")
derivk2p01 <- deriv(kappa2, "p01")
derivk2p10 <- deriv(kappa2, "p10")
derivk2p11 <- deriv(kappa2, "p11")
derivk2q00 <- deriv(kappa2, "q00")
derivk2q01 <- deriv(kappa2, "q01")
derivk2q10 <- deriv(kappa2, "q10")
derivk2q11 <- deriv(kappa2, "q11")

derivkappap00 <- deriv(kappacomb, "p00")
derivkappap01 <- deriv(kappacomb, "p01")
derivkappap10 <- deriv(kappacomb, "p10")
derivkappap11 <- deriv(kappacomb, "p11")
derivkappaq00 <- deriv(kappacomb, "q00")
derivkappaq01 <- deriv(kappacomb, "q01")
derivkappaq10 <- deriv(kappacomb, "q10")
derivkappaq11 <- deriv(kappacomb, "q11")

dlogitkappap00 <- deriv(logitk, "p00")
dlogitkappap01 <- deriv(logitk, "p01")
dlogitkappap10 <- deriv(logitk, "p10")
dlogitkappap11 <- deriv(logitk, "p11")
dlogitkappaq00 <- deriv(logitk, "q00")
dlogitkappaq01 <- deriv(logitk, "q01")
dlogitkappaq10 <- deriv(logitk, "q10")
dlogitkappaq11 <- deriv(logitk, "q11")
```

```
# Variances and covariances

  p00 <- s00 / n
  p01 <- s01 / n
  p10 <- s10 / n
  p11 <- s11 / n

  q00 <- r00 / n
  q01 <- r01 / n
  q10 <- r10 / n
  q11 <- r11 / n

  vec1 <- vector("numeric", 8)

  vec1[1] <- p00
  vec1[2] <- p01
  vec1[3] <- p10
  vec1[4] <- p11
  vec1[5] <- q00
  vec1[6] <- q01
  vec1[7] <- q10
  vec1[8] <- q11

  matp <- matrix(0, 8, 8)

  matp[1,1] <- p00
  matp[2,2] <- p01
  matp[3,3] <- p10
  matp[4,4] <- p11
  matp[5,5] <- q00
  matp[6,6] <- q01
  matp[7,7] <- q10
```

```
matp[8,8] <- q11

sigmap <- matrix(0, 8, 8)

sigmap <- (1 / n) * (matp - vec1 %*% t(vec1))

vec2 <- vector("numeric", 8)

vec2[1] <- attr(eval(derivk1p00), "gradient")[1]
vec2[2] <- attr(eval(derivk1p01), "gradient")[1]
vec2[3] <- attr(eval(derivk1p10), "gradient")[1]
vec2[4] <- attr(eval(derivk1p11), "gradient")[1]

vec2[5] <- attr(eval(derivk1q00), "gradient")[1]
vec2[6] <- attr(eval(derivk1q01), "gradient")[1]
vec2[7] <- attr(eval(derivk1q10), "gradient")[1]
vec2[8] <- attr(eval(derivk1q11), "gradient")[1]

vec3 <- vector("numeric", 8)

vec3[1] <- attr(eval(derivk2p00), "gradient")[1]
vec3[2] <- attr(eval(derivk2p01), "gradient")[1]
vec3[3] <- attr(eval(derivk2p10), "gradient")[1]
vec3[4] <- attr(eval(derivk2p11), "gradient")[1]

vec3[5] <- attr(eval(derivk2q00), "gradient")[1]
vec3[6] <- attr(eval(derivk2q01), "gradient")[1]
vec3[7] <- attr(eval(derivk2q10), "gradient")[1]
vec3[8] <- attr(eval(derivk2q11), "gradient")[1]

mat1 <- rbind(vec2, vec3)
```

```
sigma1 <- matrix(0, 2, 2)

sigma1 <- mat1 %*% sigmap %*% t(mat1)

vec4 <- vector("numeric", 8)

vec4[1] <- attr(eval(derivkappap00), "gradient")[1]
vec4[2] <- attr(eval(derivkappap01), "gradient")[1]
vec4[3] <- attr(eval(derivkappap10), "gradient")[1]
vec4[4] <- attr(eval(derivkappap11), "gradient")[1]

vec4[5] <- attr(eval(derivkappaq00), "gradient")[1]
vec4[6] <- attr(eval(derivkappaq01), "gradient")[1]
vec4[7] <- attr(eval(derivkappaq10), "gradient")[1]
vec4[8] <- attr(eval(derivkappaq11), "gradient")[1]

mat2 <- rbind(vec4, vec2)

sigma2 <- matrix(0, 2, 2)

sigma2 <- mat2 %*% sigmap %*% t(mat2)

mat3 <- rbind(vec4, vec3)

sigma3 <- matrix(0, 2, 2)

sigma3 <- mat3 %*% sigmap %*% t(mat3)

vec5 <- vector("numeric", 8)

vec5[1] <- attr(eval(dlogitkappap00), "gradient")[1]
vec5[2] <- attr(eval(dlogitkappap01), "gradient")[1]
```

```
  vec5[3] <- attr(eval(dlogitkappap10), "gradient")[1]
  vec5[4] <- attr(eval(dlogitkappap11), "gradient")[1]

  vec5[5] <- attr(eval(dlogitkappaq00), "gradient")[1]
  vec5[6] <- attr(eval(dlogitkappaq01), "gradient")[1]
  vec5[7] <- attr(eval(dlogitkappaq10), "gradient")[1]
  vec5[8] <- attr(eval(dlogitkappaq11), "gradient")[1]
  var1 <- t(vec5) %*% sigmap %*% vec5

  if(isTRUE(r1))
    {
     combSe <- p11 / p

     combSp <- (q10 + q01 + q00) / (1 - p)

     combQ <- p11 + q11
    }

  else
    {
     combSe <- (p11 + p10 + p01) / p

     combSp <- q00 / (1 - p)

     combQ <- 1 - p00 - q00
    }

  combY <- combSe + combSp - 1

  combwkc <- (p * (1 - p) * combY) / (p * indexc * (1 - combQ) + (1 - p) * (1 - indexc)
* combQ)
```

```
logitcwkc <- log(combwkc / (1 - combwkc))

# Fieller CI

d11 <- combwkc * k1 - sigma2[1,2] * z^2

d21 <- combwkc^2 - sigma2[1,1] * z^2

d31 <- k1^2 - sigma2[2,2] * z^2

LFiellerkappak1 <- (d11 - sqrt(d11^2 - d21 * d31)) / d31

UFiellerkappak1 <- (d11 + sqrt(d11^2 - d21 * d31)) / d31

d12 <- combwkc * k2 - sigma3[1,2] * z^2

d22 <- combwkc^2 - sigma3[1,1] * z^2

d32 <- k2^2 - sigma3[2,2] * z^2
LFiellerkappak2 <- (d12 - sqrt(d12^2 - d22 * d32)) / d32

UFiellerkappak2 <- (d12 + sqrt(d12^2 - d22 * d32)) / d32

# Logit CI

Llogitkappa <- exp(logitcwkc - z * sqrt(var1[1])) /(1 + exp(logitcwkc - z * sqrt(var1[1])))

Ulogitkappa <- exp(logitcwkc + z * sqrt(var1[1])) /(1 + exp(logitcwkc + z * sqrt(var1[1])))
```

```
  # Solutions for Kand = Ki and Kor = Ki
  c1 <- (1 - p) *(Q1 * combY - combQ * Y1)/ (Y1 * (p - combQ) - combY * (p - Q1))

  c2 <- (1 - p) *(Q2 * combY - combQ * Y2)/ (Y2 * (p - combQ) - combY * (p - Q2))

  # Graphics
  funct1 <- function (x) (p * (1 - p) * Y1) / (p * x * (1 - Q1) + (1 - p) * (1 - x) * Q1)
  funct2 <- function (x) (p * (1 - p) * Y2) / (p * x * (1 - Q2) + (1 - p) * (1 - x) * Q2)
  functkappa <- function (x) (p * (1 - p) * combY) / (p * x * (1 - combQ) + (1 - p) * (1 -
x) * combQ)

  if(isTRUE(r1)) tex <- "AND rule" else tex <- "OR rule"

  jpeg("Graph_cwkc.jpg")
  curve(functkappa, 0, 1,  lty = 1, ylim = c(0, 1), xlim = c(0, 1), xaxp = c(0, 1, 1), main =
tex, xlab = "Weighting index", ylab = "Weighted kappa coefficient")
  curve(funct1, 0, 1, lty = 2, add = TRUE)
  curve(funct2, 0, 1, lty = 3, add = TRUE)
  axis(1, at = c(round(c1, digits = 3), round(c2, digits = 3)))
  abline(v = c(round(c1, digits = 3), round(c2, digits = 3)), lty = 2, lwd = .1, col =
"gray90")
  lab <- expression("Combined " * kappa * "(c)", kappa[1] * "(c)", kappa[2] * "(c)")
  legend("top", lab, lty = c(1, 2, 3), ncol = 1, merge = TRUE)
  dev.off()

  #Result
  sink("Results_cwkc.txt", split=TRUE)
  cat("\n")
  cat(" R E S U L T S \n")
  cat("---------------\n")
  cat("\n")
  cat("     WEIGHTED KAPPA COEFFICIENTS OF THE BDTs \n")
```

```
  cat("\n")
  cat("Weighting index c is:",indexc, "\n")
  cat("\n")
  cat("Estimated weighted kappa coefficient of Test 1 is ",k1," and its standard error is",
sqrt(sigma1[1,1]), "\n")
  cat("\n")
  cat("Estimated weighted kappa coefficient of Test 2 is ",k2," and its standard error is",
sqrt(sigma1[2,2]), "\n")
  cat("\n")
  cat("Estimated Youden index of Test 1 is: ",Y1, "\n")
  cat("\n")
  cat("Estimated Youden index of Test 2 is: ",Y2, "\n")
  cat("\n")

  cat("     COMBINATION OF THE TWO BDTs WITH THE ", rule," RULE \n")
  cat("\n")
  cat("Estimated combined weighted kappa coefficient is ",combwkc," and its standard
error is", sqrt(sigma2[1,1]), "\n")
  cat("\n")
  cat("Estimated Youden index is: ",combY, "\n")
  cat("\n")
  cat("Estimated combined weighted kappa coefficient is equal to weighted kappa
coefficient of Test 1 if c index is: ",c1, "\n")
  cat("\n")
  cat("Estimated combined weighted kappa coefficient is equal to weighted kappa
coefficient of Test 2 if c index is: ",c2, "\n")
  cat("\n")
  cat(100 * conflevel,"% Fieller CI for combined weighted kappa coefficient / K1(c) is:
(",LFiellerkappak1, " ; ",UFiellerkappak1,") \n")
  cat("\n")
  cat(100 * conflevel,"% Fieller CI for combined weighted kappa coefficient / K2(c) is:
(",LFiellerkappak2, " ; ",UFiellerkappak2,") \n")
```

```
  cat("\n")
  cat(100 * conflevel,"% Logit CI for combined weighted kappa coefficient is:
(",Llogitkappa, " ; ",Ulogitkappa,") \n")
  cat("\n")
  sink()
}
```

References

Agresti, A., (2002). Categorical data analysis. John Wiley and Sons, New York.

Agresti, A., Coull, B.A., (1998). Approximate is better than "exact" for interval estimation of binomial proportions. American Statistician 52: 119-126.

Agresti, A., Min, Y., (2005). Sample improved confidence intervals for comparing matched proportions. Statistics in Medicine, 24: 729-740.

Anscombe, F.J., (1956). On estimating binomial response relations. Biometrics 43: 461-464.

Berry, G., Smith, C., Macaskill, P., Irwig, L., (2002). Analytic methods for comparing two dichotomous screening or diagnostic tests applied to two populations of differing disease prevalence when individuals negative on both tests are unverified. Statistics in Medicine 21: 853-862.

Bloch D.A., (1997). Comparing two diagnostic tests against the same "gold standard" in the same sample. Biometrics, 53: 73-85.

Bonferroni, C.E., (1936). Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8: 3-62.

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

Cicchetti, D.V., Volkmar, F., Klin, A., Showalter, D., (1995). Diagnosing autism using ICD-10 criteria: a comparison of neural networks and standard multivariate procedures. Child Neuropsychology, 1: 26-37.

Cicchetti D.V., (2001). Methodological Commentary The Precision of Reliability and Validity Estimates Re-Visited: Distinguishing Between Clinical and Statistical Significance of Sample Size Requirements. Journal of Clinical and Experimental Neuropsychology, 23: 695-700.

Clopper C.P., Pearson E.S., (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26: 404-413.

Efron B., Tibshirani R., (1993). An introduction to the Bootstrap. Chapman & Hall, London.

Fieller, E.C., (1940). The biological standardization of insulin. Journal of the Royal Statistical Society, 7, Supplement: 1-64.

Gart, J.J., Nam, J., (1988). Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness. Biometrics, 44: 323-338.

Gupta, R., Joshi, P., Mohan, V., Reddy, K.S., Yusuf, S., (2008). Epidemiology and causation of coronary heart disease and stroke in India. Heart, 94: 16-26.

Hochberg, Y., (1988). A sharper Bonferroni procedure for multiple test of significance. Biometrika, 75: 800-802.

Holm, S., (1979). A simple sequential rejective multiple testing procedure. Scandinavian Journal of Statistics, 6: 65-70.

Jannarone, R.J., Macera, C.A., Garrison, C.Z., (1987). Evaluating interrater agreement through case-control sampling. Biometrics, 43: 433-437.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Lehnen, R.G., (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. Biometrics, 33: 133-158.

Kosinski A.S., (2013). A weighted generalized score statistic for comparison of predictive values of diagnostic test. Statistics in Medicine, 32: 964-77.

Kraemer, H.C., Bloch, D.A., (1990). A Note on case-control sampling to estimate kappa coefficients. Biometrics, 46: 49-59.

Kraemer H.C., (1992). Evaluating medical test. Objective and quantitative guidelines. Sage publications, Newbury Park.

Kraemer H.C., Periyakoil V.S., Noda A., (2002). Kappa coefficients in medical research. Statistics in Medicine, 21: 2109-2129.

Leisenring W., Alonzo T., Pepe M.S., (2000). Comparisons of predictive values of binary medical diagnostic test for paired designs. Biometrics, 56: 344–351.

Lin, S.C.C., (1999). Some results on combinations of two binary screening tests. Journal of Biopharmaceutical Statistics, 9: 81-88.

Macaskill, P., Walter, S.D., Irwig, L., Franco, E.L., (2002). Assessing the gain in diagnostic performance when combining two diagnostic tests. Statistics in Medicine, 21: 2527-2546.

Marshall, R.J., (1989). The predictive values of simple rules for combining two diagnostic tests. Biometrics, 45: 1213-1222.

Martín Andrés, A., Álvarez Hernández M., (2014a). Two-tailed asymptotic inferences for a proportion. Journal of Applied Statistics, 41: 1516-1529.

Martín Andrés, A., Álvarez Hernández M., (2014b). Two-tailed approximate confidence intervals for the ratio of proportions. Statistics and Computing, 24: 65-75.

Martín Andrés, A., Álvarez Hernández M., (2016). Comment on 'An improved score interval with a modified midpoint for a binomial proportion'. Journal of Statistical Computation and Simulation, 86: 388-393.

Mercaldo, N.D., Kit, F.L., Zhou, X.H., (2007). Confidence intervals for predictive values with an emphasis to case-control studies. Statistics in Medicine, 26: 2170-2183.

Patil, S.M., Banker, M.P., Padalkar, R.K., Pathak, A.P., Bhagat, S.S., Ghone, R.A., Phatake, A.S., (2013). The clinical assessment of ischaemia modified albumin and troponin I in the early diagnosis of the acute coronary syndrome. Journal of Clinical and Diagnostic Research, 7: 804-808.

Pepe, M.S. (2003). The statistical evaluation of medical tests for classification and prediction. Oxford University Press, New York.

Price, R.M., Bonett, D.G., (2004). An improved confidence interval for a linear function of binomial proportions. Computational Statistics and Data Analysis, 45: 449-456.

Roldán-Nofuentes J.A., Luna del Castillo J.D., (2007) Comparing of the likelihood ratios of two binary diagnostic test in paired designs. Statistics in Medicine, 26:4179–4201.

Roldán-Nofuentes J.A., Luna del Castillo J.D., Montero-Alonso M.A., (2009). Confidence intervals of weighted kappa coefficient of a binary diagnostic test. Communications in Statistics - Simulation and Computation, 38:1562-1578.

Roldán-Nofuentes J.A., Luna del Castillo J.D., Montero-Alonso M.A., (2012). Global hypothesis test to simultaneously compare the predictive values of two binary diagnostic tests. Computational statistics & data analysis, Special issue "Computational Statistics for Clinical Research", 56(5):1161-1173.

Simel, D.L., Samsa, G.P., Matchar, D.B., (1991). Likelihood ratios with confidence: sample size estimation for diagnostic test studies. Journal of Clinical Epidemiology, 44: 763-770.

Stamey, J.D., Holt, M.M., (2010). Bayesian interval estimation for predictive values from case-control studies. Communications in Statistics - Simulation and Computation, 39: 101-110.

Vacek, P.M., (1985). The effect of conditional dependence on the evaluation of diagnostic test. Biometrics, 41:959–968.

Wang, W., Davis, C.S., Soong, S., (2006). Comparison of predictive values of two diagnostic test from the same sample of subjects using weighted least squares. Statistics in Medicine, 25: 2215-2229.

Weiner D.A., Ryan T.J., Mccabe C.H., Kennedy J.W., Schloss M., Tristani F., Chaitman B.R., Fisher L.D., (1979). Exercise stress testing. Correlations among history

of angina, ST-segmant response and prevalence of coronary-artery disease in the coronary artery surgery study (CASS). The New England Journal of Medicine, 301: 230-235.

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

Youden W.J., (1950). Index for rating diagnostic test. Cancer, 3: 32-35.

Yu W., Gou X., Xu W., (2014). An improved score interval with a modified midpoint for a binomial proportion. Journal of Statistical Computation and Simulation. 84: 1022-1038.

Zhou X.H., Obuchowski N.A., McClish D.K., (2002). Statistical methods in diagnostic medicine. John Wiley and Sons, New York.