

Article

# Confidence Intervals and Sample Size to Compare the Predictive Values of Two Diagnostic Tests

José Antonio Roldán-Nofuentes <sup>1,\*</sup>  and Saad Bouh Regad <sup>2</sup><sup>1</sup> Department of Statistics, School of Medicine, University of Granada, 18016 Granada, Spain<sup>2</sup> Epidemiology and Public Health Research Unit and URMCD, University of Nouakchott Alaasriya, BP 880 Nouakchott, Mauritania; regad@una.mr

\* Correspondence: jaroldan@ugr.es

**Abstract:** A binary diagnostic test is a medical test that is applied to an individual in order to determine the presence or the absence of a certain disease and whose result can be positive or negative. A positive result indicates the presence of the disease, and a negative result indicates the absence. Positive and negative predictive values represent the accuracy of a binary diagnostic test when it is applied to a cohort of individuals, and they are measures of the clinical accuracy of the binary diagnostic test. In this manuscript, we study the comparison of the positive (negative) predictive values of two binary diagnostic tests subject to a paired design through confidence intervals. We have studied confidence intervals for the difference and for the ratio of the two positive (negative) predictive values. Simulation experiments have been carried out to study the asymptotic behavior of the confidence intervals, giving some general rules for application. We also study a method to calculate the sample size to compare the parameters using confidence intervals. We have written a program in R to solve the problems studied in this manuscript. The results have been applied to the diagnosis of colorectal cancer.



**Citation:** Roldán-Nofuentes, J.A.; Regad, S.B. Confidence Intervals and Sample Size to Compare the Predictive Values of Two Diagnostic Tests. *Mathematics* **2021**, *9*, 1462. <https://doi.org/10.3390/math9131462>

Academic Editor: Lev Klebanov

Received: 11 May 2021  
Accepted: 18 June 2021  
Published: 22 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** binary diagnostic test; confidence interval; positive predictive value; negative predictive value; sample size

## 1. Introduction

A diagnostic test is medical test that is applied to an individual in order to determine the presence of a certain disease. Binary diagnostic tests are a very common type of diagnostic test in clinical practice. A binary diagnostic test (BDT) is a diagnostic test whose possible result is positive or negative. A positive result indicates the presence of the disease, and a negative result indicates the absence. Mammography for the diagnosis of breast cancer is an example of BDT. Accuracy of a BDT is measured in terms of two fundamental parameters: sensitivity and specificity. Sensitivity ( $Se$ ) is the probability of the result of the BDT being positive when the individual has the disease, and specificity ( $Sp$ ) is the probability of the result of the BDT being negative when the individual does not have the disease. Therefore,  $Se$  and  $Sp$  are probabilities of getting the disease diagnosis right, and they represent the intrinsic accuracy of the BDT, since these parameters depend on the physical, chemical, or biological properties upon which the BDT is developed. Other parameters that are used to assess and compare two BDTs are the positive and negative predictive values. Positive predictive value ( $\tau$ ) is the probability of an individual having the disease when the result of the BDT is positive, and the negative predictive value ( $\nu$ ) is the probability of an individual not having the disease when the result of the BDT is negative. Predictive values represent the accuracy of the diagnostic test when it is applied to a cohort of individuals, and they are measures of the clinical accuracy of the BDT. Predictive values

depend on  $Se$ ,  $Sp$ , and on the disease prevalence ( $p$ ), and are easily calculated applying Bayes theorem, i.e.,

$$\tau = \frac{p \times Se}{p \times Se + (1 - p) \times (1 - Sp)} \text{ and } v = \frac{(1 - p) \times Sp}{p \times (1 - Se) + (1 - p) \times Sp}. \quad (1)$$

The accuracy of a BDT is assessed in relation to a gold standard. A gold standard (GS) is a medical test that determines without error whether or not an individual has the disease. Biopsy for the diagnosis of breast cancer is an example of GS.

On the other hand, the comparison of parameters of two BDTs is an important topic in the study of statistical methods for diagnosis in Medicine. The most frequent sample design to compare the parameters of two BDTs is the paired design. The paired design consists of applying the two BDTs and the GS to all individuals of a random sample sized  $n$ . The comparison of the predictive values of two BDTs subject to a paired design has been the subject of different studies. Bennett [1,2], Leisenring et al. [3], Wang et al. [4], Kosinski [5], Tsou [6], and Takahashi and Yamamoto [7] have studied hypothesis tests to compare the two positive predictive values and the negative predictive values independently. Roldán-Nofuentes et al. [8] studied a global hypothesis test to simultaneously compare the positive and negative predictive values of two BDTs. However, the comparison of predictive values using confidence intervals has been little studied. If the hypothesis test is significant to an  $\alpha$  error, the confidence interval (CI) allows determining how much one predictive value is greater than the other. Moskowitz and Pepe [9] have proposed a Wald-type CI for the ratio of the two positive (negative) predictive values. A Wald-type CI for the difference of the two positive (negative) predictive values is easily obtained by inverting the contrast statistic of the hypothesis test studied by Wang et al. [4].

The objective of this manuscript is study CIs to compare the positive (negative) predictive values of two BDTs subject to a paired design. For this, we have studied CIs for the difference and for the ratio of the two positive (negative) predictive values. If a CI for the difference (ratio) does not contain the zero (one) value, then we reject the equality between the two positive (negative) predictive values, and we estimate how much bigger one predictive value is than another one. The problem of calculating the sample size to compare the two positive (negative) predictive values through a CI is also studied.

This manuscript is structured in the following way. In Section 2, the existing CIs are presented and other new CIs are proposed, both for the ratio and for the difference between the positive (negative) predictive values. In Section 3, simulation experiments are carried out to study the coverage probabilities and the average lengths of the CIs. In Section 4, a method to calculate the sample size to compare the parameters through CIs is proposed. In Section 5, we present the program called “cicpvbdt”, which is a program written in R that solves the problems studied in this manuscript. In Section 6, the results were applied to an example on the diagnosis of colorectal cancer, and in Section 7, the results are discussed.

## 2. Confidence Intervals

Let us consider two BDTs that are assessed in relation to the same GS. Let  $T_i$  be the variable that models the result of the  $i$ th BDT, with  $i = 1, 2$ .  $T_i = 0$  indicates that the test result is negative, and  $T_i = 1$  indicates that the test result is positive. Let  $D$  be the random variable that models the result of the GS, so that  $D = 1$  when the individual is diseased and  $D = 0$  when the individual is non-diseased. Let  $Se_i$  and  $Sp_i$  be the sensitivity and specificity of the  $i$ th BDT. Table 1 shows the observed frequencies obtained subject to a paired design. The frequencies  $s_{jk}$  and  $r_{jk}$  are the product of a multinomial distribution whose probabilities are  $p_{jk} = P(D = 1, T_1 = j, T_2 = k)$  and  $q_{jk} = P(D = 0, T_1 = j, T_2 = k)$ , with  $j, k = 0, 1$ . Applying the conditional dependence model of Vacek [10], probabilities  $p_{jk}$  and  $q_{jk}$  are written as

$$p_{jk} = p \left[ Se_1^j (1 - Se_1)^{1-j} Se_2^k (1 - Se_2)^{1-k} + \delta_{jk} \varepsilon_1 \right] \quad (2)$$

and

$$q_{jk} = q \left[ Sp_1^{1-j} (1 - Sp_1)^j Sp_2^{1-k} (1 - Sp_2)^k + \delta_{jk} \varepsilon_0 \right], \tag{3}$$

where  $\delta_{jk} = 1$  if  $j = k$  and  $\delta_{jk} = -1$  if  $j \neq k$ , with  $j, k = 0, 1$ ,  $\varepsilon_1$  is the dependence factor between the two BDTs when  $D = 1$  and  $\varepsilon_0$  is the dependence factor between the two BDTs when  $D = 0$ . It is verified that  $0 \leq \varepsilon_1 \leq \text{Min}\{Se_1(1 - Se_2), Se_2(1 - Se_1)\}$  and  $0 \leq \varepsilon_0 \leq \text{Min}\{Sp_1(1 - Sp_2), Sp_2(1 - Sp_1)\}$ . If  $\varepsilon_1 = \varepsilon_0 = 0$ , then the two BDTs are conditionally independent on the disease. This assumption is not realistic, so in practice, it is verified that  $\varepsilon_1 > 0$  and/or  $\varepsilon_0 > 0$ . Let  $\pi = (p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00})^T$  be the vector of probabilities of the multinomial distribution,  $p = \sum_{i,j=0}^1 p_{ij}$  and  $q = 1 - p = \sum_{i,j=0}^1 q_{ij}$ .

The maximum likelihood estimators of  $p_{jk}$  and  $q_{jk}$  are:

$$\hat{p}_{jk} = \frac{s_{jk}}{n} \text{ and } \hat{q}_{jk} = \frac{r_{jk}}{n}.$$

**Table 1.** Observed frequencies subject to a paired design.

	$T_1=1$		$T_1=0$		Total
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$	
Diseased ( $D = 1$ )	$s_{11}$	$s_{10}$	$s_{01}$	$s_{00}$	$s$
Non-diseased ( $D = 0$ )	$r_{11}$	$r_{10}$	$r_{01}$	$r_{00}$	$r$
Total	$s_{11} + r_{11}$	$s_{10} + r_{10}$	$s_{01} + r_{01}$	$s_{00} + r_{00}$	$n$

From Equation (1), the sensitivity and specificity of each BDT are written, in terms of the predictive values and of  $p$ , as

$$Se_i = \frac{\tau_i(v_i - q)}{pY_i} \text{ and } Sp_i = \frac{v_i(\tau_i - p)}{qY_i} \tag{4}$$

where  $q = 1 - p$  and  $Y_i = \tau_i + v_i - 1$ . Then

$$0 \leq \varepsilon_1 \leq \text{Min} \left\{ \frac{\tau_1(v_1 - q)}{pY_1} \left( 1 - \frac{\tau_2(v_2 - q)}{pY_2} \right), \frac{\tau_2(v_2 - q)}{pY_2} \left( 1 - \frac{\tau_1(v_1 - q)}{pY_1} \right) \right\} \tag{5}$$

and

$$0 \leq \varepsilon_0 \leq \text{Min} \left\{ \frac{v_1(\tau_1 - p)}{qY_1} \left( 1 - \frac{v_2(\tau_2 - p)}{qY_2} \right), \frac{v_2(\tau_2 - p)}{qY_2} \left( 1 - \frac{v_1(\tau_1 - p)}{qY_1} \right) \right\}. \tag{6}$$

In terms of predictive values, Equations (2) and (3) are written as

$$p_{jk} = p \left[ \frac{\tau_1^j (v_1 - q)^j (\tau_1 + p)^{1-j} (1 - v_1)^{1-j}}{p^j p^{1-j} Y_1^j Y_1^{1-j}} \times \frac{\tau_2^k (v_2 - q)^k (\tau_2 + p)^{1-k} (1 - v_2)^{1-k}}{p^k p^{1-k} Y_2^k Y_1^{1-k}} + \delta_{jk} \varepsilon_1 \right] \tag{7}$$

and

$$q_{jk} = q \left[ \frac{(1 - \tau_1)^j (v_1 - q)^j (\tau_1 - p)^{1-j} v_1^{1-j}}{q^j q^{1-j} Y_1^j Y_1^{1-j}} \times \frac{(1 - \tau_2)^k (v_2 - q)^k (\tau_2 - p)^{1-k} v_2^{1-k}}{q^k q^{1-k} Y_2^k Y_2^{1-k}} + \delta_{jk} \varepsilon_0 \right]. \tag{8}$$

The estimator of sensitivities and specificities are

$$\hat{Se}_1 = \frac{s_{11} + s_{10}}{s}, \hat{Se}_2 = \frac{s_{11} + s_{01}}{s}, \hat{Sp}_1 = \frac{r_{01} + r_{00}}{r} \text{ and } \hat{Sp}_2 = \frac{r_{10} + r_{00}}{r}, \tag{9}$$

and applying the delta method, the estimators of the variances-covariances of  $\hat{S}e_i$  and  $\hat{S}p_i$  are

$$\hat{V}ar(\hat{S}e_i) = \frac{\hat{S}e_i(1-\hat{S}e_i)}{s}, \quad Var(\hat{S}p_i) = \frac{\hat{S}p_i(1-\hat{S}p_i)}{r},$$

$$\hat{C}ov(\hat{S}e_1, \hat{S}e_2) = \frac{\hat{\epsilon}_1}{s} \text{ and } \hat{C}ov(\hat{S}p_1, \hat{S}p_2) = \frac{\hat{\epsilon}_0}{r},$$

where  $\hat{\epsilon}_1 = \frac{n\hat{p}_{11}}{s} - \hat{S}e_1\hat{S}e_2 = \frac{s_{11}s_{00} - s_{10}s_{01}}{s^2}$  and  $\hat{\epsilon}_0 = \frac{n\hat{q}_{00}}{r} - \hat{S}p_1\hat{S}p_2 = \frac{r_{11}r_{00} - r_{10}r_{01}}{r^2}$ , and the estimator of the disease prevalence is

$$\hat{p} = \frac{s}{n}. \tag{10}$$

Let  $Q_i = pSe_i + q(1 - Sp_i)$  be the probability that the result of the *i*th BDT is positive and let  $\bar{Q}_i = 1 - Q_i = p(1 - Se_i) + qSp_i$  be the probability that the result is negative. Its estimators are:

$$\hat{Q}_1 = \frac{s_{10} + s_{11} + r_{10} + r_{11}}{n} \text{ and } \hat{Q}_2 = \frac{s_{01} + s_{11} + r_{01} + r_{11}}{n},$$

and

$$\hat{\bar{Q}}_1 = \frac{s_{01} + s_{00} + r_{01} + r_{00}}{n} \text{ and } \hat{\bar{Q}}_2 = \frac{s_{10} + s_{00} + r_{10} + r_{00}}{n},$$

respectively. With respect to the predictive values, their estimators are:

$$\hat{\tau}_1 = \frac{s_{11} + s_{10}}{s_{11} + s_{10} + r_{11} + r_{10}}, \quad \hat{\tau}_2 = \frac{s_{11} + s_{01}}{s_{11} + s_{01} + r_{11} + r_{01}}, \quad \hat{v}_1 = \frac{r_{01} + r_{00}}{s_{01} + s_{00} + r_{01} + r_{00}}$$

and

$$\hat{v}_2 = \frac{r_{10} + r_{00}}{s_{10} + s_{00} + r_{10} + r_{00}}.$$

Applying the delta method, the estimators of the variances-covariances of  $\hat{\tau}_i$  and  $\hat{v}_i$  are [8]:

$$\hat{V}ar(\hat{\tau}_1) = \frac{(s_{10}+s_{11})(r_{10}+r_{11})}{(s_{10}+s_{11}+r_{10}+r_{11})^3}, \quad \hat{V}ar(\hat{v}_1) = \frac{(s_{00}+s_{01})(r_{00}+r_{01})}{(s_{00}+s_{01}+r_{00}+r_{01})^3}$$

$$\hat{V}ar(\hat{\tau}_2) = \frac{(s_{01}+s_{11})(r_{01}+r_{11})}{(s_{01}+s_{11}+r_{01}+r_{11})^3}, \quad \hat{V}ar(\hat{v}_2) = \frac{(s_{00}+s_{10})(r_{00}+r_{10})}{(s_{00}+s_{10}+r_{00}+r_{10})^3},$$

$$\hat{C}ov(\hat{\tau}_1, \hat{\tau}_2) = \frac{s_{11}r_{10}r_{01}+r_{11}[s_{01}(s_{10}+s_{11})+s_{11}(s_{11}+s_{10}+r_{11}+r_{10}+r_{01})]}{(s_{01}+s_{11}+r_{01}+r_{11})^2(s_{10}+s_{11}+r_{10}+r_{11})^2}$$

and

$$\hat{C}ov(\hat{v}_1, \hat{v}_2) = \frac{s_{10}(s_{00} + s_{01})r_{00} + s_{00}[r_{00}^2 + r_{01}r_{10} + r_{00}(s_{00} + s_{01} + r_{10} + r_{01})]}{(s_{00} + s_{01} + r_{00} + r_{01})^2(s_{00} + s_{10} + r_{00} + r_{10})^2}.$$

When two parameters are compared in Statistics, the interest is to study the difference or the ratio between them. Then, we compare the positive (negative) predictive values of two BDTs through CIs for the difference, i.e.,  $\delta_\tau = \tau_1 - \tau_2$  and  $\delta_v = v_1 - v_2$ , and for the ratio, i.e.,  $\rho_\tau = \tau_1/\tau_2$  and  $\rho_v = v_1/v_2$ .

### 2.1. CIs for the Difference

Three CIs for each difference  $\delta_\tau$  and  $\delta_v$  are studied: Wald CI, bias-corrected bootstrap CI, and Monte Carlo Bayesian CI.

#### 2.1.1. Wald CI

Wang et al. [4] have studied the comparison of the PVs of two BDTs through the weighted least square method. The test statistics for  $H_0 : \tau_1 = \tau_2$  and  $H_0 : v_1 = v_2$  are

$$z_\tau = \frac{\hat{\delta}_\tau}{\sqrt{\hat{V}ar(\hat{\delta}_\tau)}} \text{ and } z_v = \frac{\hat{\delta}_v}{\sqrt{\hat{V}ar(\hat{\delta}_v)}},$$

respectively. Both test statistics follow a standard normal distribution where

$$\hat{V}ar(\hat{\delta}_\tau) = \hat{V}ar(\hat{\tau}_1) + \hat{V}ar(\hat{\tau}_2) - 2\hat{C}ov(\hat{\tau}_1, \hat{\tau}_2)$$

and

$$\hat{V}ar(\hat{\delta}_v) = \hat{V}ar(\hat{v}_1) + \hat{V}ar(\hat{v}_2) - 2\hat{C}ov(\hat{v}_1, \hat{v}_2)$$

are the estimators of the variances of  $\hat{\delta}_\tau$  and  $\hat{\delta}_v$ , respectively. Inverting the two test statistics, the Wald CIs for  $\delta_\tau$  and for  $\delta_v$  are

$$\delta_\tau \in \hat{\delta}_\tau \pm z_{1-\alpha/2} \sqrt{\hat{V}ar(\hat{\delta}_\tau)} \text{ and } \delta_v \in \hat{\delta}_v \pm z_{1-\alpha/2} \sqrt{\hat{V}ar(\hat{\delta}_v)},$$

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

### 2.1.2. Bias-Corrected Bootstrap CI

The bias-corrected bootstrap CI is calculated from  $B$  random samples with replacement generated from the sample of  $n$  individuals. In each of the  $B$  samples, we calculate  $\hat{\tau}_{1b}$ ,  $\hat{\tau}_{2b}$ ,  $\hat{v}_{1b}$ ,  $\hat{v}_{2b}$ ,  $\hat{\delta}_{\tau b} = \hat{\tau}_{1b} - \hat{\tau}_{2b}$  and  $\hat{\delta}_{vb} = \hat{v}_{1b} - \hat{v}_{2b}$ , with  $b = 1, \dots, B$ . Then, the average differences are calculated as  $\hat{\delta}_{\tau B} = \frac{1}{B} \sum_{b=1}^B \hat{\delta}_{\tau b}$  and  $\hat{\delta}_{vB} = \frac{1}{B} \sum_{b=1}^B \hat{\delta}_{vb}$ . Assuming that the bootstrap statistics  $\hat{\delta}_{\tau B}$  and  $\hat{\delta}_{vB}$  can be transformed to a normal distribution, the bias-corrected bootstrap CIs [11] are calculated in the following way. Let  $A_\tau = \#(\hat{\delta}_{\tau b} < \hat{\delta}_\tau)$  be the number of bootstrap estimators  $\hat{\delta}_{\tau b}$  that are lower than the maximum likelihood estimator (MLE)  $\hat{\delta}_\tau$ , and let  $A_v = \#(\hat{\delta}_{vb} < \hat{\delta}_v)$  be the number of bootstrap estimators  $\hat{\delta}_{vb}$  that are lower than the MLE  $\hat{\delta}_v$ . Let  $\hat{z}_\tau = \Phi^{-1}(A_\tau/B)$  and  $\hat{z}_v = \Phi^{-1}(A_v/B)$ , where  $\Phi^{-1}(\cdot)$  is the inverse function of the standard normal cumulative distribution function. Let  $\alpha_{1\tau} = \Phi(2\hat{z}_\tau - z_{1-\alpha/2})$ ,  $\alpha_{2\tau} = \Phi(2\hat{z}_\tau + z_{1-\alpha/2})$ ,  $\alpha_{1v} = \Phi(2\hat{z}_v - z_{1-\alpha/2})$ , and  $\alpha_{2v} = \Phi(2\hat{z}_v + z_{1-\alpha/2})$ ; then, the bias-corrected bootstrap CI for  $\delta_\tau$  is

$$\delta_\tau \in \left( \hat{\delta}_{\tau B}^{(\alpha_{1\tau})}, \hat{\delta}_{\tau B}^{(\alpha_{2\tau})} \right)$$

and the bias-corrected bootstrap CI for  $\delta_v$  is

$$\delta_v \in \left( \hat{\delta}_{vB}^{(\alpha_{1v})}, \hat{\delta}_{vB}^{(\alpha_{2v})} \right)$$

where  $\hat{\delta}_{\tau B}^{(\gamma)}$  is the  $\gamma$ th quantile of the distribution of the  $B$  bootstrap estimations of  $\delta_\tau$ , and  $\hat{\delta}_{vB}^{(\gamma)}$  is the  $\gamma$ th quantile of the distribution of the  $B$  bootstrap estimations of  $\delta_v$ .

### 2.1.3. Monte Carlo Bayesian CI

The number of diseased individuals ( $s$ ) is the product of a binomial distribution, i.e.,  $s \rightarrow B(n, p)$ . Conditioning on  $D = 1$ , it is verified that

$$s_{11} + s_{10} \rightarrow B(s, Se_1) \text{ and } s_{11} + s_{01} \rightarrow B(s, Se_2).$$

The number of non-diseased individuals ( $r$ ) is the product of a binomial distribution, i.e.,  $r \rightarrow B(n, q)$ . Conditioning on  $D = 0$ , it is verified that

$$r_{01} + r_{00} \rightarrow B(r, Sp_1) \text{ and } r_{10} + r_{00} \rightarrow B(r, Sp_2).$$

On the other hand, the estimators  $\hat{S}e_i$ ,  $\hat{S}p_i$ , and  $\hat{p}$  (Equations (9) and (10)) are estimators of binomial proportions. Therefore, for these estimators we propose conjugate beta prior distributions, i.e.,

$$\hat{S}e_i \rightarrow \text{Beta}(\alpha_{Se_i}, \beta_{Se_i}), \hat{S}p_i \rightarrow \text{Beta}(\alpha_{Sp_i}, \beta_{Sp_i}) \text{ and } \hat{p} \rightarrow \text{Beta}(\alpha_p, \beta_p). \tag{11}$$

Let  $\mathbf{n} = (s_{11}, s_{10}, s_{01}, s, r_{11}, r_{10}, r_{01}, n - s)$  be the vector of observed frequencies, with  $s_{00} = s - s_{11} - s_{10} - s_{01}$ ,  $r = n - s$ , and  $r_{00} = n - s - r_{11} - r_{10} - r_{01}$ . Then, the posteriori distributions for the estimators of  $Se_i$ ,  $Sp_i$ , and  $p$  are:

$$\begin{aligned} \hat{Se}_1 | \mathbf{n} &\rightarrow \text{Beta}(s_{11} + s_{10} + \alpha_{Se_1}, s - s_{11} - s_{10} + \beta_{Se_1}), \\ \hat{Sp}_1 | \mathbf{n} &\rightarrow \text{Beta}(r_{01} + r_{00} + \alpha_{Sp_1}, n - s - r_{01} - r_{00} + \beta_{Sp_1}), \\ \hat{Se}_2 | \mathbf{n} &\rightarrow \text{Beta}(s_{11} + s_{01} + \alpha_{Se_2}, s - s_{11} - s_{01} + \beta_{Se_2}), \\ \hat{Sp}_2 | \mathbf{n} &\rightarrow \text{Beta}(r_{10} + r_{00} + \alpha_{Sp_2}, n - s - r_{10} - r_{00} + \beta_{Sp_2}), \\ \hat{p} | \mathbf{n} &\rightarrow \text{Beta}(s + \alpha_p, n - s + \beta_p). \end{aligned} \tag{12}$$

The posteriori distribution for the positive (negative) predictive value of each BDT, and for  $\delta_\tau$  and  $\delta_v$ , can be approximated applying the Monte Carlo method. The Monte Carlo method is a computational method that consists of generating M values from the posteriori distributions (12). In the mth iteration, the values generated for  $Se_j^{(m)}$ ,  $Sp_j^{(m)}$ , and  $p^{(m)}$  are plugged in the equations

$$\tau_i^{(m)} = \frac{p^{(m)} \times Se_i^{(m)}}{p^{(m)} \times Se_i^{(m)} + (1 - p^{(m)}) \times (1 - Sp_i^{(m)})}$$

and

$$v_i^{(m)} = \frac{(1 - p^{(m)}) \times Sp_i^{(m)}}{p^{(m)} \times (1 - Se_i^{(m)}) + (1 - p^{(m)}) \times Sp_i^{(m)'}}$$

with  $i = 1, 2$ , and then,  $\delta_\tau^{(m)} = \tau_1^{(m)} - \tau_2^{(m)}$  and  $\delta_v^{(m)} = v_1^{(m)} - v_2^{(m)}$  are calculated. As estimators of  $\delta_\tau$  and  $\delta_v$ , we calculate the average of the M estimations of the differences, i.e.,  $\hat{\delta}_{\tau Bay} = \frac{1}{M} \sum_{m=1}^M \delta_\tau^{(m)}$  and  $\hat{\delta}_{v Bay} = \frac{1}{M} \sum_{m=1}^M \delta_v^{(m)}$ . Finally, based on the M values of  $\delta_\tau^{(m)}$  and of  $\delta_v^{(m)}$ , we propose CIs based on quantiles. Therefore, the  $100 \times (1 - \alpha)\%$  CI for  $\delta_\tau$  is

$$\delta_\tau \in \left( q_{\tau Bay}^{(\alpha/2)}, q_{\tau Bay}^{(1-\alpha/2)} \right)$$

and the  $100 \times (1 - \alpha)\%$  CI for  $\delta_v$  is

$$\delta_v \in \left( q_{v Bay}^{(\alpha/2)}, q_{v Bay}^{(1-\alpha/2)} \right)$$

where  $q_{\tau Bay}^{(\gamma)}$  ( $q_{v Bay}^{(\gamma)}$ ) is the  $\gamma$ th quantile of the distribution of the M values  $\delta_\tau^{(m)}$  ( $\delta_v^{(m)}$ ).

### 2.2. CIs for the Ratio

Five CIs for each ratio  $\rho_\tau = \tau_1/\tau_2$  and  $\rho_v = v_1/v_2$  are studied: Wald CI, logarithmic CI, Fieller CI, bias-corrected bootstrap CI, and Monte Carlo Bayesian CI.

#### 2.2.1. Wald CI

Moskowitz and Pepe [9] have studied a Wald-type confidence interval for the ratio of the two positive (negative) predictive values. The  $100 \times (1 - \alpha)\%$  Wald CI for  $\delta_\tau$  is

$$\rho_\tau \in \hat{\rho}_\tau \pm z_{1-\alpha/2} \sqrt{\hat{Var}(\hat{\rho}_\tau)}$$

and the  $100 \times (1 - \alpha)\%$  Wald CI for  $\delta_v$  is

$$\rho_v \in \hat{\rho}_v \pm z_{1-\alpha/2} \sqrt{\hat{Var}(\hat{\rho}_v)},$$

where  $\hat{V}ar(\hat{\rho}_\tau)$  and  $\hat{V}ar(\hat{\rho}_v)$ , obtained applying the delta method, are

$$\hat{V}ar(\hat{\rho}_\tau) = \frac{\hat{\tau}_2^2 \hat{V}ar(\hat{\tau}_1) + \hat{\tau}_1^2 \hat{V}ar(\hat{\tau}_2) - 2\hat{\tau}_1 \hat{\tau}_2 \hat{C}ov(\hat{\tau}_1, \hat{\tau}_2)}{\hat{\tau}_2^4}$$

and

$$\hat{V}ar(\hat{\rho}_v) = \frac{\hat{v}_2^2 \hat{V}ar(\hat{v}_1) + \hat{v}_1^2 \hat{V}ar(\hat{v}_2) - 2\hat{v}_1 \hat{v}_2 \hat{C}ov(\hat{v}_1, \hat{v}_2)}{\hat{v}_2^4}.$$

These CIs are for  $\rho_\tau = \tau_1/\tau_2$  and  $\rho_v = v_1/v_2$ . If we want to calculate the CI for the ratio  $\tau_2/\tau_1 = 1/\rho_\tau$  and for the ratio  $v_2/v_1 = 1/\rho_v$ , then we have to divide the CI for  $\rho_\tau$  by  $\hat{\rho}_\tau^2$  and the CI for  $\rho_v$  by  $\hat{\rho}_v^2$ . For example, if  $(L_\tau, U_\tau)$  is the Wald CI for  $\tau_1/\tau_2$ , then  $(L_\tau/\hat{\rho}_\tau^2, U_\tau/\hat{\rho}_\tau^2)$  is the Wald CI for  $\tau_2/\tau_1$ .

### 2.2.2. Logarithmic CI

Assuming the asymptotic normality of the Napierian logarithm of  $\hat{\rho}_\tau$  and of  $\hat{\rho}_v$ , i.e.,  $ln(\hat{\rho}_\tau) \rightarrow N(ln(\delta_\tau), Var[ln(\delta_\tau)])$  and  $ln(\hat{\rho}_v) \rightarrow N(ln(\delta_v), Var[ln(\delta_v)])$  when  $n$  is large, an asymptotic CI for  $ln(\rho_\tau)$  is

$$ln(\hat{\rho}_\tau) \pm z_{1-\alpha/2} \sqrt{\hat{V}ar[ln(\hat{\rho}_\tau)]}$$

and an asymptotic CI for  $ln(\rho_v)$  is

$$ln(\hat{\rho}_v) \pm z_{1-\alpha/2} \sqrt{\hat{V}ar[ln(\hat{\rho}_v)]}.$$

Taking exponential in each of the previous expressions, the logarithmic CI for  $\delta_\tau$  is

$$\rho_\tau \in \hat{\rho}_\tau \times exp\left\{\pm z_{1-\alpha/2} \sqrt{\hat{V}ar[ln(\hat{\rho}_\tau)]}\right\}$$

and the logarithmic CI for  $\delta_v$  is

$$\rho_v \in \hat{\rho}_v \times exp\left\{\pm z_{1-\alpha/2} \sqrt{\hat{V}ar[ln(\hat{\rho}_v)]}\right\},$$

where  $\hat{V}ar[ln(\hat{\rho}_\tau)]$  and  $\hat{V}ar[ln(\hat{\rho}_v)]$ , obtained applying the delta method, are:

$$\hat{V}ar[ln(\hat{\rho}_\tau)] = \frac{\hat{V}ar[\hat{\tau}_1]}{\hat{\tau}_1^2} + \frac{\hat{V}ar[\hat{\tau}_2]}{\hat{\tau}_2^2} - \frac{2\hat{C}ov[\hat{\tau}_1, \hat{\tau}_2]}{\hat{\tau}_1 \hat{\tau}_2}$$

and

$$\hat{V}ar[ln(\hat{\rho}_v)] = \frac{\hat{V}ar[\hat{v}_1]}{\hat{v}_1^2} + \frac{\hat{V}ar[\hat{v}_2]}{\hat{v}_2^2} - \frac{2\hat{C}ov[\hat{v}_1, \hat{v}_2]}{\hat{v}_1 \hat{v}_2}.$$

If we want to calculate the logarithmic CI for the ratio  $\tau_2/\tau_1$ , then the CI is obtained by calculating the inverse of each boundary of CI for  $\rho_\tau = \tau_1/\tau_2$ . In a similar way, the CI for  $v_2/v_1$  is calculated.

### 2.2.3. Fieller CI

The method of Fieller [12] is a classic method used to estimate the ratio of two parameters. In order to apply this method, it is necessary to assume that  $\hat{\boldsymbol{\tau}}^T \rightarrow N(\boldsymbol{\tau}, \boldsymbol{\Sigma}_\tau)$  and that  $\hat{\mathbf{v}}^T \rightarrow N(\mathbf{v}, \boldsymbol{\Sigma}_v)$ ; i.e., it is necessary to assume that the estimators of the positive

(negative) predictive values are distributed according to a normal bivariate distribution, and where  $\hat{\tau} = (\hat{\tau}_1, \hat{\tau}_2), \hat{v} = (\hat{v}_1, \hat{v}_2),$

$$\Sigma_{\tau} = \begin{pmatrix} Var(\tau_1) & Cov(\tau_1, \tau_2) \\ Cov(\tau_1, \tau_2) & Var(\tau_2) \end{pmatrix} = \begin{pmatrix} \sigma_{\tau 11} & \sigma_{\tau 12} \\ \sigma_{\tau 21} & \sigma_{\tau 22} \end{pmatrix}$$

and

$$\Sigma_v = \begin{pmatrix} Var(v_1) & Cov(v_1, v_2) \\ Cov(v_1, v_2) & Var(v_2) \end{pmatrix} = \begin{pmatrix} \sigma_{v 11} & \sigma_{v 12} \\ \sigma_{v 21} & \sigma_{v 22} \end{pmatrix}.$$

Applying the method of Fieller, it is verified that  $\hat{\tau}_1 - \rho_{\tau} \hat{\tau}_2 \rightarrow N(0, \sigma_{\tau 11} + \rho_{\tau}^2 \sigma_{\tau 22} - 2\rho_{\tau} \sigma_{\tau 12})$  and that  $\hat{v}_1 - \rho_v \hat{v}_2 \rightarrow N(0, \sigma_{v 11} + \rho_v^2 \sigma_{v 22} - 2\rho_v \sigma_{v 12})$  when  $n$  is large. The Fieller CI for  $\rho_{\tau}$  is obtained by solving the inequality

$$\frac{(\hat{\tau}_1 - \rho_{\tau} \hat{\tau}_2)^2}{\hat{\sigma}_{\tau 11} + \rho_{\tau}^2 \hat{\sigma}_{\tau 22} - 2\rho_{\tau} \hat{\sigma}_{\tau 12}} < z_{1-\alpha/2}^2,$$

and the Fieller CI for  $\rho_v$  is obtained by solving the inequality

$$\frac{(\hat{v}_1 - \rho_v \hat{v}_2)^2}{\hat{\sigma}_{v 11} + \rho_v^2 \hat{\sigma}_{v 22} - 2\rho_v \hat{\sigma}_{v 12}} < z_{1-\alpha/2}^2.$$

Finally, the Fieller CI for  $\rho_{\tau}$  is

$$\rho_{\tau} \in \frac{\hat{\beta}_{\tau 12} \pm \sqrt{\hat{\beta}_{\tau 12}^2 - \hat{\beta}_{\tau 11} \hat{\beta}_{\tau 22}}}{\hat{\beta}_{\tau 22}}$$

where  $\hat{\beta}_{\tau ij} = \hat{\tau}_i \hat{\tau}_j - \hat{\sigma}_{\tau ij} z_{1-\alpha/2}^2$  with  $i, j = 1, 2$  and verifying that  $\hat{\beta}_{\tau 12} = \hat{\beta}_{\tau 21}$ . This CI is valid when  $\hat{\beta}_{\tau 12}^2 > \hat{\beta}_{\tau 11} \hat{\beta}_{\tau 22}$  and  $\hat{\beta}_{\tau 22} \neq 0$ . Similarly, the Fieller CI for  $\rho_v$  is

$$\rho_v \in \frac{\hat{\beta}_{v 12} \pm \sqrt{\hat{\beta}_{v 12}^2 - \hat{\beta}_{v 11} \hat{\beta}_{v 22}}}{\hat{\beta}_{v 22}}$$

where  $\hat{\beta}_{v ij} = \hat{v}_i \hat{v}_j - \hat{\sigma}_{v ij} z_{1-\alpha/2}^2$  with  $i, j = 1, 2,$  and  $\hat{\beta}_{v 12} = \hat{\beta}_{v 21}$ . This CI is valid when  $\hat{\beta}_{v 12}^2 > \hat{\beta}_{v 11} \hat{\beta}_{v 22}$  and  $\hat{\beta}_{v 22} \neq 0$ . The Fieller CI for  $\tau_2 / \tau_1$  ( $v_2 / v_1$ ) is calculated by inverting the limits of the CI for  $\rho_{\tau}$  ( $\rho_v$ ).

#### 2.2.4. Bias-Corrected Bootstrap CI

The bias-corrected bootstrap CI for  $\rho_{\tau}$  ( $\rho_v$ ) is obtained in a similar way to that of  $\delta_{\tau}$  ( $\delta_v$ ). In each sample with a replacement obtained, we calculate  $\hat{\tau}_{1b}, \hat{\tau}_{2b}, \hat{v}_{1b}, \hat{v}_{2b}, \hat{\rho}_{\tau b} = \hat{\tau}_{1b} / \hat{\tau}_{2b},$  and  $\hat{\rho}_{vb} = \hat{v}_{1b} / \hat{v}_{2b},$  with  $b = 1, \dots, B$ . Then, based on the B ratios, we estimate the average ratios as  $\hat{\rho}_{\tau B} = \frac{1}{B} \sum_{b=1}^B \hat{\rho}_{\tau b}$  and  $\hat{\rho}_{vB} = \frac{1}{B} \sum_{b=1}^B \hat{\rho}_{vb}$ . Assuming that these statistics can be transformed to a normal distribution, the bias-corrected bootstrap CI [11] for  $\rho_{\tau}$  ( $\rho_v$ ) is calculated in a similar way as the bias-corrected bootstrap CI for  $\delta_{\tau}$  ( $\delta_v$ ), considering that  $A_{\tau} = \#(\hat{\rho}_{\tau b} < \hat{\rho}_{\tau})$  and that  $A_v = \#(\hat{\rho}_{vb} < \hat{\rho}_v)$ . Finally, the bias-corrected bootstrap CI for  $\rho_{\tau}$  is

$$\rho_{\tau} \in \left( \hat{\rho}_{\tau B}^{(\alpha_1)}, \hat{\rho}_{\tau B}^{(\alpha_2)} \right),$$

where  $\hat{\rho}_{\tau B}^{(\gamma)}$  is the  $\gamma$ th quantile of the distribution of the B bootstrap estimations of  $\rho_{\tau}$ . Similarly, the bias-corrected bootstrap CI for  $\rho_v$  is

$$\rho_v \in \left( \hat{\rho}_{vB}^{(\alpha_1)}, \hat{\rho}_{vB}^{(\alpha_2)} \right),$$



where  $\hat{\rho}_{vB}^{(\gamma)}$  is the  $\gamma$ th quantile of the distribution of the B bootstrap estimations of  $\rho_v$ . The bias-corrected bootstrap CI for  $\tau_2/\tau_1$  ( $v_2/v_1$ ) is calculated by inverting the limits of the bias-corrected bootstrap CI for  $\rho_\tau$  ( $\rho_v$ ).

### 2.2.5. Monte Carlo Bayesian CI

The Monte Carlo Bayesian CI for  $\rho_\tau$  ( $\rho_v$ ) is obtained in a similar way as the Monte Carlo Bayesian CI for  $\delta_\tau$  ( $\delta_v$ ). Considering the same distributions (10) and (11), in the  $m$ th iteration, we calculate the ratios  $\rho_\tau^{(m)} = \tau_1^{(m)}/\tau_2^{(m)}$  and  $\rho_v^{(m)} = v_1^{(m)}/v_2^{(m)}$ . As estimators, we calculate  $\hat{\rho}_{\tau Bay} = \frac{1}{M} \sum_{m=1}^M \rho_\tau^{(m)}$  and  $\hat{\rho}_{v Bay} = \frac{1}{M} \sum_{m=1}^M \rho_v^{(m)}$ . Finally, we calculate the CIs based on quantiles, i.e.,

$$\rho_\tau \in \left( q_{\tau Bay}^{(\alpha/2)}, q_{\tau Bay}^{(1-\alpha/2)} \right) \text{ and } \rho_v \in \left( q_{v Bay}^{(\alpha/2)}, q_{v Bay}^{(1-\alpha/2)} \right)$$

where  $q_{\tau Bay}^{(\gamma)}$  ( $q_{v Bay}^{(\gamma)}$ ) is the  $\gamma$ th quantile of the distribution of the M values  $\rho_\tau^{(m)}$  ( $\rho_v^{(m)}$ ). The Monte Carlo Bayesian CI for  $\tau_2/\tau_1$  ( $v_2/v_1$ ) is calculated by inverting the limits of the Monte Carlo Bayesian CI for  $\rho_\tau$  ( $\rho_v$ ).

### 3. Simulation Experiments

The CIs studied in Section 2 are approximate, and therefore, it is necessary to study their asymptotic behaviors. For this, Monte Carlo simulation experiments have been carried out to study the coverage probabilities and the average lengths of the CIs studied, considering a confidence level of 95%. These experiments have consisted of generating  $N = 10,000$  random samples with multinomial distribution sized  $n = \{50, 100, 200, 500, 1000\}$ , and whose probabilities have been calculated from Equations (7) and (8). The experiments have been designed from the predictive values of both BDTs. As the value of disease prevalence, we have taken  $p = \{10\%, 25\%, 50\%, 75\%\}$ , and as values of predictive values, we have taken the values  $\tau_i, v_i = \{0.70, 0.75, \dots, 0.90, 0.95\}$ , which are realistic values in clinical practice. Next, using these values, we have calculated the maximum values of the dependence factors  $\varepsilon_1$  and  $\varepsilon_0$  (Equations (5) and (6)). As values of  $\varepsilon_1$  and  $\varepsilon_0$ , we have taken intermediate and high values, i.e., 50% of the maximum value of  $\varepsilon_i$  and 90% of the maximum value of  $\varepsilon_i$ , respectively. Finally, we have calculated the probabilities of the multinomial distributions using Equations (7) and (8). In each scenario, we have calculated all the CIs for each of the N random samples.

For the bias-corrected bootstrap CIs, for each one of the N random samples,  $B = 2000$  samples with replacement have been generated, and from these B samples, the bias-corrected bootstrap CIs have been calculated.

For the Monte Carlo Bayesian CI, we have considered a  $Beta(1, 1)$  distribution as a priori distribution for the estimators of sensitivities, specificities and prevalence. The  $Beta(1, 1)$  distribution is a non-informative distribution, which is flat for every possible value of the sensitivities, specificities, and prevalence. Therefore, the impact of the  $Beta(1, 1)$  distribution on the posteriori distributions is minimal. Moreover, for each one of the N random samples,  $M = 10,000$  random samples have been generated, and the Monte Carlo Bayesian CIs have been calculated from them.

In each of the N samples generated, all the CIs have been calculated. Furthermore, it has been checked whether each CI contains the value of the parameter (difference or ratio, depending on the type of CI). The coverage probability has been calculated by dividing the number of samples in which the CI contains the parameter by the total number of samples. For each CI, its length (upper limit minus the lower limit) has also been calculated, and finally, the average length of each CI has been calculated.

3.1. CIs for the Differences and Ratios of Positive Predictive Values

Table 2 shows some of the results obtained for the three CIs for the difference  $\delta_\tau$  for four different scenarios and for intermediate values of  $\varepsilon_1$  and  $\varepsilon_0$ . When the sample size is small ( $n = 50$ ) or moderate ( $n = 100$ ), the CIs for  $\delta_\tau$  have a coverage probability close to 1. For the difference  $\delta_\tau$ , in very general terms, the Wald CI is the interval that has a coverage probability with better fluctuations around 95%, especially when  $n$  is moderate or large ( $n \geq 200$ ). The bias-corrected bootstrap CI has a very similar behavior to the Wald CI, especially when the sample size is large. In general terms, the Monte Carlo Bayesian CI has a coverage probability greater than that of the other two intervals, even when the coverage probability of the other two intervals fluctuates around 95%.

**Table 2.** Asymptotic behaviors of the CIs for the difference of the two positive predictive values.

$\tau_1 = 0.75 \tau_2 = 0.75 \nu_1 = 0.95 \nu_2 = 0.95 \delta_\tau = 0$ $\varepsilon_1 = 0.124 \varepsilon_0 = 0.010 p = 0.10$						
n	Wald		BCB		MCB	
	CP	AL	CP	AL	CP	AL
50	1	0.749	1	0.743	1	0.787
100	1	0.538	1	0.527	1	0.683
200	0.988	0.409	1	0.397	1	0.454
500	0.953	0.261	0.990	0.254	0.998	0.363
1000	0.944	0.185	0.957	0.186	0.993	0.259
$\tau_1 = 0.90 \tau_2 = 0.85 \nu_1 = 0.95 \nu_2 = 0.90 \delta_\tau = 0.05$ $\varepsilon_1 = 0.021 \varepsilon_0 = 0.044 p = 0.50$						
n	Wald		BCB		MCB	
	CP	AL	CP	AL	CP	AL
50	0.982	0.251	1	0.242	0.999	0.326
100	0.951	0.174	0.966	0.182	0.993	0.223
200	0.954	0.122	0.952	0.126	0.991	0.156
500	0.941	0.077	0.933	0.077	0.981	0.099
1000	0.952	0.055	0.948	0.055	0.988	0.070
$\tau_1 = 0.85 \tau_2 = 0.75 \nu_1 = 0.95 \nu_2 = 0.85 \delta_\tau = 0.10$ $\varepsilon_1 = 0.037 \varepsilon_0 = 0.024 p = 0.25$						
n	Wald		BCB		MCB	
	CP	AL	CP	AL	CP	AL
50	0.998	0.513	1	0.499	1	0.602
100	0.981	0.354	1	0.343	0.999	0.445
200	0.941	0.250	0.987	0.243	0.991	0.318
500	0.956	0.158	0.959	0.159	0.989	0.204
1000	0.953	0.112	0.954	0.113	0.989	0.145

CP: coverage probability. AL: average length. Wald: Wald CI. BCB: bias-corrected bootstrap CI. MCB: Monte Carlo Bayesian CI.

Regarding the CIs for the ratio  $\rho_\tau$ , Table 3 shows the results obtained for the same scenarios as in Table 2. When the sample size is small ( $n = 50$ ), the five CIs for  $\rho_\tau$  have a coverage probability close to 1. In general terms, there is not an important difference between the coverage probabilities and the average lengths of the Wald, logarithm, and Fieller CIs, especially when  $n \geq 100$ . When the sample size is small, the logarithmic CI and the Fieller CI have an average length slightly greater than the Wald CI. The bias-corrected bootstrap CI has a very similar behavior to the Wald, logarithmic, and Fieller CIs, especially when the sample size is large. In general terms, the Monte Carlo Bayesian CI has a coverage probability greater than that of the other four intervals.

**Table 3.** Asymptotic behaviors of the CIs for the ratio of the two positive predictive values.

$\tau_1 = 0.75 \tau_2 = 0.75 \nu_1 = 0.95 \nu_2 = 0.95 \rho_\tau = 1$ $\varepsilon_1 = 0.124 \varepsilon_0 = 0.010 p = 0.10$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	1	1.326	1	1.450	1	2.046	1	1.348	1	2.183
100	0.999	0.966	1	1.018	1	1.311	1	0.973	1	1.534
200	0.989	0.630	0.992	0.643	0.994	0.652	1	0.648	1	0.978
500	0.962	0.359	0.966	0.361	0.953	0.369	0.988	0.364	0.998	0.535
1000	0.956	0.250	0.952	0.251	0.944	0.253	0.956	0.256	0.993	0.363
$\tau_1 = 0.90 \tau_2 = 0.85 \nu_1 = 0.95 \nu_2 = 0.90 \rho_\tau = 1.06$ $\varepsilon_1 = 0.021 \varepsilon_0 = 0.044 p = 0.50$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	0.986	0.326	0.994	0.328	0.993	0.334	1	0.347	1	0.448
100	0.949	0.216	0.952	0.216	0.950	0.218	0.957	0.232	0.994	0.288
200	0.952	0.151	0.955	0.151	0.954	0.152	0.953	0.157	0.990	0.196
500	0.941	0.095	0.941	0.095	0.940	0.095	0.935	0.095	0.982	0.122
1000	0.950	0.067	0.951	0.067	0.949	0.067	0.946	0.067	0.987	0.086
$\tau_1 = 0.85 \tau_2 = 0.75 \nu_1 = 0.95 \nu_2 = 0.85 \rho_\tau = 1.13$ $\varepsilon_1 = 0.037 \varepsilon_0 = 0.024 p = 0.25$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	0.997	0.950	1	0.958	0.998	0.961	1	0.992	1	1.407
100	0.972	0.591	0.983	0.596	0.979	0.636	1	0.689	0.999	0.841
200	0.941	0.390	0.943	0.392	0.940	0.396	0.988	0.398	0.989	0.528
500	0.950	0.241	0.954	0.241	0.957	0.242	0.960	0.244	0.989	0.314
1000	0.951	0.169	0.953	0.170	0.950	0.171	0.953	0.171	0.988	0.218

CP: coverage probability. AL: average length. Wald: Wald CI. Log: logarithmic CI. Fieller: Fieller CI. BCB: bias-corrected bootstrap CI. MCB: Monte Carlo Bayesian CI.

### 3.2. CIs for the Differences and Ratios of Negative Predictive Values

Table 4 shows the results for the three CIs for the difference  $\delta_v$  for the same scenarios as in Tables 2 and 3. In general terms, when the sample size is small, the three CIs have a coverage probability close to 1, although in some situations, the bias-corrected bootstrap CI may have a coverage probability well below 95%. In general terms, the Monte Carlo Bayesian CI has a coverage probability that almost always fluctuates above 95%. For the difference  $\delta_v$ , the Wald CI is the interval that has a coverage probability with better fluctuations around 95%, especially when the sample size is moderate or large.

Regarding the CIs for the ratio  $\rho_v$ , Table 5 shows the results obtained for the same scenarios as in Table 4. When the sample size is small ( $n = 50$ ), the five CIs for  $\rho_\tau$  fail or have a coverage probability close to 1. In general terms, the conclusions for the bias-corrected bootstrap CI and for the Monte Carlo Bayesian CI are very similar to those obtained for the corresponding intervals for the difference  $\delta_v$ . With respect to the other intervals, there is not an important difference between the coverage probabilities and the average lengths of the Wald, logarithm, and Fieller CIs, especially when  $n \geq 100$ . When the sample size is small, the logarithmic CI and the Fieller CI have an average length slightly greater than that of the Wald CI.

**Table 4.** Asymptotic behaviors of the CIs for the difference of the two negative predictive values.

$\tau_1 = 0.75 \tau_2 = 0.75 v_1 = 0.95 v_2 = 0.95 \delta_v = 0$ $\varepsilon_1 = 0.124 \varepsilon_0 = 0.010 p = 0.10$						
Wald		BCB		MCB		
n	CP	AL	CP	AL	CP	AL
50	1	0.127	1	0.119	1	0.154
100	0.999	0.072	1	0.069	0.999	0.095
200	0.989	0.046	1	0.042	0.999	0.063
500	0.949	0.028	0.968	0.028	0.994	0.040
1000	0.946	0.020	0.943	0.020	0.993	0.028
$\tau_1 = 0.90 \tau_2 = 0.85 v_1 = 0.95 v_2 = 0.90 \delta_v = 0.05$ $\varepsilon_1 = 0.021 \varepsilon_0 = 0.044 p = 0.50$						
Wald		BCB		MCB		
n	CP	AL	CP	AL	CP	AL
50	0.999	0.264	1	0.276	1	0.344
100	0.966	0.169	0.952	0.170	0.995	0.222
200	0.950	0.115	0.932	0.119	0.989	0.147
500	0.949	0.073	0.948	0.073	0.983	0.090
1000	0.952	0.052	0.953	0.052	0.984	0.063
$\tau_1 = 0.85 \tau_2 = 0.75 v_1 = 0.95 v_2 = 0.85 \delta_v = 0.10$ $\varepsilon_1 = 0.037 \varepsilon_0 = 0.024 p = 0.25$						
Wald		BCB		MCB		
n	CP	AL	CP	AL	CP	AL
50	0.936	0.207	0.720	0.182	0.948	0.218
100	0.938	0.142	0.874	0.133	0.960	0.151
200	0.948	0.099	0.937	0.096	0.967	0.107
500	0.957	0.062	0.961	0.062	0.975	0.068
1000	0.946	0.044	0.947	0.044	0.964	0.048

CP: coverage probability. AL: average length. Wald: Wald CI. BCB: bias-corrected bootstrap CI. MCB: Monte Carlo Bayesian CI.

Similar conclusions are obtained when  $\varepsilon_1$  and  $\varepsilon_0$  take high values. Therefore, the dependency factors  $\varepsilon_1$  and  $\varepsilon_0$  do not have an important effect on the behavior of the CIs for the difference (ratio) of the two negative predictive values.

As a conclusion, the following general rules of application can be given depending on the sample size, since the sample size is the only parameter controlled by the researcher: (a) apply the Wald CI for the difference of the positive (negative) predictive values whatever the sample size; (b) apply the Wald CI for the ratio of the two positive (negative) predictive values when the sample size is small, and apply the Wald CI, the logarithmic CI, the Fieller CI, or the bias-corrected bootstrap CI when the sample size is moderate or high.

Once some general rules of application have been established, what is better: to use a CI for the difference or a CI for the ratio? Simulation experiments have shown that the Wald CIs for the difference and the Wald CIs for the ratio have a very similar coverage probability. Furthermore, the Wald CI for the difference has a coverage probability very similar to that of the Fieller CI when the sample size is large. The Wald CIs for the difference are obtained by inverting the Wald test statistics of the tests  $H_0 : \tau_1 - \tau_2 = 0$  and  $H_0 : v_1 - v_2 = 0$ , and the Wald CIs for the ratio are obtained by inverting the Wald test statistics of the tests  $H_0 : \tau_1/\tau_2 = 1$  and  $H_0 : v_1/v_2 = 1$ . Wang et al. [4] have shown through simulation experiments that the hypothesis tests  $H_0 : \tau_1 - \tau_2 = 0$  and  $H_0 : v_1 - v_2 = 0$  have better asymptotic behavior than the tests  $H_0 : \tau_1/\tau_2 = 1$  and  $H_0 : v_1/v_2 = 1$ . Furthermore, Wang et al. recommend using the difference-based approach as it is more straightforward and more understandable for researchers. Therefore, we recommend using a CI for the difference instead of a CI for the ratio.

**Table 5.** Asymptotic behaviors of the CIs for the ratio of the two negative predictive values.

$\tau_1 = 0.75 \quad \tau_2 = 0.75 \quad v_1 = 0.95 \quad v_2 = 0.95 \quad \rho_v = 1$ $\varepsilon_1 = 0.124 \quad \varepsilon_0 = 0.010 \quad p = 0.10$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	1	0.144	1	0.144	1	0.145	1	0.128	1.000	0.173
100	0.999	0.076	1	0.079	0.999	0.080	1	0.074	0.999	0.103
200	0.988	0.046	0.991	0.047	0.992	0.048	1	0.045	0.999	0.068
500	0.950	0.030	0.950	0.030	0.949	0.030	0.971	0.029	0.994	0.042
1000	0.948	0.021	0.947	0.021	0.946	0.021	0.946	0.021	0.993	0.030
$\tau_1 = 0.90 \quad \tau_2 = 0.85 \quad v_1 = 0.95 \quad v_2 = 0.90 \quad \rho_v = 1.06$ $\varepsilon_1 = 0.021 \quad \varepsilon_0 = 0.044 \quad p = 0.50$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	1	0.324	1	0.326	0.999	0.332	1	0.349	1	0.462
100	0.954	0.201	0.964	0.201	0.962	0.202	0.964	0.219	0.995	0.274
200	0.945	0.134	0.946	0.134	0.947	0.135	0.921	0.138	0.989	0.175
500	0.946	0.085	0.945	0.085	0.946	0.085	0.947	0.085	0.982	0.105
1000	0.954	0.060	0.952	0.060	0.952	0.060	0.950	0.060	0.983	0.074
$\tau_1 = 0.85 \quad \tau_2 = 0.75 \quad v_1 = 0.95 \quad v_2 = 0.85 \quad \rho_v = 1.12$ $\varepsilon_1 = 0.037 \quad \varepsilon_0 = 0.024 \quad p = 0.25$										
n	Wald		Log		Fieller		BCB		MCB	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
50	0.936	0.271	0.934	0.272	0.933	0.275	0.724	0.283	0.933	0.291
100	0.939	0.184	0.936	0.184	0.935	0.185	0.849	0.191	0.961	0.199
200	0.945	0.129	0.947	0.129	0.945	0.129	0.923	0.125	0.963	0.140
500	0.957	0.081	0.958	0.081	0.958	0.081	0.959	0.081	0.974	0.088
1000	0.949	0.057	0.950	0.057	0.950	0.057	0.948	0.057	0.964	0.062

CP: coverage probability. AL: average length. Wald: Wald CI. Log: logarithmic CI. Fieller: Fieller CI. BCB: bias-corrected bootstrap CI. MCB: Monte Carlo Bayesian CI.

#### 4. Sample Size

The calculation of the sample size to compare parameters has great interest in Statistics. Next, we propose a procedure to determine the sample size necessary to estimate the difference between the two positive (negative) predictive values with a precision  $\phi_\tau$  ( $\phi_v$ ) and a confidence  $100(1 - \alpha)\%$ . This procedure is based on the Wald CI for the difference  $\delta_\tau$  ( $\delta_v$ ), since in general terms, this CI is the interval with the best asymptotic behavior. This procedure requires having a pilot sample (or another study) in order to estimate the predictive values and their differences. If the pilot sample is not small and the Wald CI for the difference  $\delta_\tau$  ( $\delta_v$ ) contains the value 0, then the null hypothesis of equality of the predictive values is not rejected, and it does not make sense to calculate the sample size. However, if the sample is small, it may be necessary to calculate the sample size, since the Wald CI will be very wide and may contain the value 0 even if the predictive values are different. Let us consider that  $\tau_1 \geq \tau_2$  ( $v_1 \geq v_2$ ) and therefore  $\delta_\tau \geq 0$  ( $\delta_v \geq 0$ ), and let  $\phi_\tau$  and  $\phi_v$  be the precisions set by the researcher ( $\phi$  must be a small value if the researcher wants high precision). Based on the asymptotic normality of  $\hat{\delta}_\tau = \hat{\tau}_1 - \hat{\tau}_1$  and of  $\hat{\delta}_v = \hat{v}_1 - \hat{v}_2$ , it is verified that

$$\hat{\delta}_\tau \in \delta_\tau \pm z_{1-\alpha/2} \sqrt{Var(\hat{\delta}_\tau)} \text{ and } \hat{\delta}_v \in \delta_v \pm z_{1-\alpha/2} \sqrt{Var(\hat{\delta}_v)},$$

i.e., the probability of obtaining an estimator  $\hat{\delta}_\tau$  ( $\hat{\delta}_v$ ) is in this interval with a probability  $100(1 - \alpha)\%$ .

For positive predictive values, the method is as follows. Setting a precision  $\phi_\tau$ , the sample size is calculated from the equation

$$\phi_\tau = z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\delta}_\tau)}, \tag{13}$$

where the variance is

$$\text{Var}(\hat{\delta}_\tau) = \frac{pqQ_2\tau_1\bar{\tau}_1 + pqQ_1\tau_2\bar{\tau}_2 - 2(pq^2\tau_1\tau_2\varepsilon_0 + p^2q\bar{\tau}_1\bar{\tau}_2\varepsilon_1 + \tau_1\tau_2\bar{\tau}_1\bar{\tau}_2Q_1Q_2)}{npqQ_1Q_2}. \tag{14}$$

The proof can be seen in the Appendix A. This variance depends on the positive predictive values ( $\tau_i$ ), on the disease prevalence ( $p$ ), on the probability of a positive result of each test ( $Q_i$ ), on the dependency factors ( $\varepsilon_i$ ), and on the sample size  $n$ . Substituting in Equation (13) the parameters with their estimators and clearing  $n$ , the sample size to estimate the difference  $\delta_\tau$  with precision  $\phi_\tau$  and a confidence  $100(1 - \alpha)\%$  is

$$n_\tau = n = \frac{z_{1-\alpha/2}^2}{\phi_\tau^2} \times \frac{\hat{p}\hat{q}\hat{\tau}_1\hat{\tau}_1\hat{Q}_2 + \hat{p}\hat{q}\hat{\tau}_2\hat{\tau}_2\hat{Q}_1 - 2(\hat{p}\hat{q}^2\hat{\tau}_1\hat{\tau}_2\hat{\varepsilon}_0 + \hat{p}^2\hat{q}\hat{\tau}_1\hat{\tau}_2\hat{\varepsilon}_1 + \hat{\tau}_1\hat{\tau}_2\hat{\tau}_1\hat{\tau}_2\hat{Q}_1\hat{Q}_2)}{\hat{p}\hat{q}\hat{Q}_1\hat{Q}_2}. \tag{15}$$

Once the equation for the sample size is obtained, the method to calculate the sample size consists of the following steps:

- (1) Take pilot samples sized  $n'_\tau$ , and from this sample, calculate  $\hat{\tau}_i, \hat{v}_i, \hat{\varepsilon}_i, \hat{p}, \hat{Q}_i$  and the Wald CI for the difference  $\delta_\tau$ . If the Wald CI has a precision  $\phi_\tau$ , then with the pilot sample, the precision has been reached, and the process ends. In this situation, the difference  $\delta_\tau$  has been estimated with a precision  $\phi_\tau$  to a confidence  $100(1 - \alpha)\%$ . Otherwise, go to the next step.
- (2) From the estimates obtained with the pilot sample, calculate the sample size  $n_\tau$  applying Equation (15).
- (3) Take the sample sized  $n_\tau$  ( $n_\tau - n'_\tau$  individuals are added to the initial pilot sample), and from this sample, calculate all the estimators and the Wald CI for the difference  $\delta_\tau$ . If the Wald CI has a precision  $\phi_\tau$ , then the process ends (the precision has been reached with the new sample). If the Wald CI does not have the precision  $\phi_\tau$ , then this sample is considered as a pilot sample and go to step 1.

This method to calculate the sample size  $n$  is an iterative method, which depends on the initial pilot sample and therefore does not guarantee that the difference between the positive predictive values will be estimated with the precision  $\phi_\tau$ .

Sample size to estimate the difference  $\delta_v$  is calculated in a similar way. In this case,

$$\text{Var}(\hat{\delta}_v) = \frac{pqv_1\bar{v}_1\bar{Q}_2 + pqv_2\bar{v}_2\bar{Q}_1 - 2(pq^2\bar{v}_1\bar{v}_2\varepsilon_0 + p^2qv_1v_2\varepsilon_1 + v_1v_2\bar{v}_1\bar{v}_2\bar{Q}_1\bar{Q}_2)}{npq\bar{Q}_1\bar{Q}_2} \tag{16}$$

and the sample size  $n_v$  to estimate the difference  $\delta_v$  with precision  $\phi_v$  and a confidence  $100(1 - \alpha)\%$  is

$$n_v = \frac{z_{1-\alpha/2}^2}{\phi_v^2} \times \frac{\hat{p}\hat{q}\hat{v}_1\hat{v}_1\hat{\bar{Q}}_2 + \hat{p}\hat{q}\hat{v}_2\hat{v}_2\hat{\bar{Q}}_1 - 2(\hat{p}\hat{q}^2\hat{v}_1\hat{v}_2\hat{\varepsilon}_0 + \hat{p}^2\hat{q}\hat{v}_1\hat{v}_2\hat{\varepsilon}_1 + \hat{v}_1\hat{v}_2\hat{v}_1\hat{v}_2\hat{\bar{Q}}_1\hat{\bar{Q}}_2)}{\hat{p}\hat{q}\hat{\bar{Q}}_1\hat{\bar{Q}}_2}. \tag{17}$$

If the researcher wants to estimate  $\delta_\tau$  with precision  $\phi_\tau$  and also wants to estimate  $\delta_v$  with precision  $\phi_v$ , at the same level of confidence, then the final sample size is  $n = \text{Max}(n_\tau, n_v)$ . Using the largest sample size, can guarantee that the CI for the difference of the two positive predictive values and that the CI for the difference of the two negative predictive values verify the set precision for each of them.

The method for calculating the sample size depends on the values of the estimators obtained from the pilot sample. As the values of the estimators depend on each sample

(and therefore vary from one sample to another), it is necessary to study how the values of the estimators affect the calculation of the sample size. Therefore, we have carried out simulation experiments to study the effect of the values of the estimators on the calculation of the sample size. These simulation experiments consisted of the following steps:

- (1) Calculate the sample size  $n_\tau$  ( $n_v$ ), Equations (14) and (16), from the values of the parameters in the scenarios considered (Tables 2 and 4). Therefore, these equations have been applied using the values of the parameters instead of the values of the estimators.
- (2) Generate  $N = 10,000$  multinomial random samples sized  $n_\tau$  ( $n_v$ ) and whose probabilities have been calculated from Equations (7) and (8), using the parameters of the scenarios considered, and as values  $\varepsilon_i$ , intermediate (50%) and high (90%) values have been considered. From each one of the  $N$  random samples, all the estimators ( $\hat{\tau}_i$ ,  $\hat{v}_i$ ,  $\hat{\varepsilon}_i$ ,  $\hat{p}$  and  $\hat{Q}_i$ ) have been calculated, and then, the sample size  $n'_\tau$  ( $n'_v$ ) has been calculated applying Equation (14) (Equation (16)).
- (3) In each scenario considered, the average sample size and relative bias have been calculated.

Table 6 shows the results obtained for different precision values (2.5% and 5%, which are values that can be considered as high precision values) and  $1 - \alpha = 0.95$ . The relative biases are very small, the equations of the sample sizes provide robust values, and therefore, the pilot sample has little effect on the calculation of the sample sizes.

**Table 6.** Sample size for estimated the difference between the positive (negative) predictive values.

Positive Predictive Values				
	$\tau_1 = 0.90 \ \tau_2 = 0.85 \ v_1 = 0.95 \ v_2 = 0.90$ $\delta_\tau = 0.05 \ \varepsilon_1 = 0.021 \ \varepsilon_0 = 0.044 \ p = 0.50$		$\tau_1 = 0.85 \ \tau_2 = 0.75 \ v_1 = 0.95 \ v_2 = 0.85$ $\delta_\tau = 0.10 \ \varepsilon_1 = 0.037 \ \varepsilon_0 = 0.024 \ p = 0.25$	
	$\phi_\tau = 0.025$	$\phi_\tau = 0.05$	$\phi_\tau = 0.025$	$\phi_\tau = 0.05$
Sample size	1203	301	5048	1262
Average sample size	1204	302	5054	1267
Relative bias (%)	0.17	0.33	0.12	0.40
Negative predictive values				
	$\tau_1 = 0.90 \ \tau_2 = 0.85 \ v_1 = 0.95 \ v_2 = 0.90$ $\delta_v = 0.05 \ \varepsilon_1 = 0.021 \ \varepsilon_0 = 0.044 \ p = 0.50$		$\tau_1 = 0.85 \ \tau_2 = 0.75 \ v_1 = 0.95 \ v_2 = 0.85$ $\delta_v = 0.10 \ \varepsilon_1 = 0.037 \ \varepsilon_0 = 0.024 \ p = 0.25$	
	$\phi_v = 0.025$	$\phi_v = 0.05$	$\phi_v = 0.025$	$\phi_v = 0.05$
Sample size	1079	270	782	196
Average sample size	1080	272	783	198
Relative bias (%)	0.09	0.74	0.13	1.02

### 5. Program Cipvbd

We have written a program in R [13] to solve the problems raised in this manuscript. The program is called “cicpvbd” (confidence intervals to compare the predictive values of binary diagnostic tests), and it allows calculating all CIs and sample sizes. The program runs with the command “cicpvbd( $s_{11}, s_{10}, s_{01}, s_{00}, r_{11}, r_{10}, r_{01}, r_{00}, \delta_\tau, \delta_v$ )”. By default, the level of confidence is 95%. The program does not calculate the sample sizes when  $\delta_\tau = 0$  and  $\delta_v = 0$ , and only calculates the sample sizes when  $\delta_\tau > 0$  and/or  $\delta_v > 0$ . In this last situation, the program checks if the set precision is reached. The program checks that all values are valid (e.g., that there are no negative observed frequencies, etc.). The program also checks that all the parameters and their variances–covariances can be estimated. For the bias-corrected bootstrap CIs, 2000 samples with replacement are generated, and for the Monte Carlo Bayesian CIs, 10,000 random samples are generated. The results obtained are saved in a file called “results\_cicpvbt.txt” in the folder from where the program is run. The program is available as Supplemental Material of this manuscript.

### 6. Example

The results obtained have been applied to a study on the diagnosis of colorectal cancer, using two diagnostic tests: Fecal Immunochemical Testing (FIT) and Fecal Occult Blood Testing (FOBT). The GS for the diagnosis of colorectal cancer is the biopsy. Table 7 shows the observed frequencies when the two BDTs and the GS have been applied to a sample of 168 adult men suspected of having colorectal cancer. Using the program “cicpvbdt” with the command “cicpvbdt(68, 18, 1, 13, 4, 1, 2, 61, 0, 0)”; all the results shown in Table 7 are obtained.

**Table 7.** Observed frequencies and CIs.

Observed Frequencies					
Biopsy	FIT: positive		FIT: negative		Total
	FOBT: positive	FOBT: negative	FOBT: positive	FOBT: negative	
Cancer	68	18	1	13	100
Normal	4	1	2	61	68
Total	72	19	3	74	168
Results					
	Positive predictive value			Negative predictive value	
FIT	0.945 ± 0.024			0.818 ± 0.044	
FOBT	0.920 ± 0.031			0.667 ± 0.049	
<i>p</i>	$\epsilon_1$	$\epsilon_0$	$Q_1$	$Q_2$	
0.595	0.087	0.052	0.542	0.446	
CIs for $\delta_\tau = \tau_1 - \tau_2$					
	Wald		BCB		MCB
	(-0.016, 0.066)		(-0.013, 0.073)		(-0.045, 0.105)
CIs for $\rho_\tau = \tau_1/\tau_2$					
	Wald	Log	Fieller	BCB	MCB
	(0.981, 1.073)	(0.982, 1.074)	(0.983, 1.076)	(0.985, 1.084)	(0.952, 1.124)
CIs for $\delta_v = v_1 - v_2$					
	Wald		BCB		MCB
	(0.081, 0.222)		(0.089, 0.231)		(0.049, 0.248)
CIs for $\rho_v = v_1/v_2$					
	Wald	Log	Fieller	BCB	MCB
	(1.101, 1.353)	(1.108, 1.350)	(1.112, 1.368)	(1.121, 1.382)	(1.069, 1.420)

Wald: Wald CI. Log: logarithmic CI. Fieller: Fieller CI. BCB: bias-corrected bootstrap CI. MCB: Monte Carlo Bayesian CI.

The estimated positive predictive values of FIT and of FOBT are 94.5% and 92.0%, and the estimated negative predictive values are 81.8% and 66.7%, respectively. Using the recommendations given in Section 3, the 95% Wald CI for the difference between the two positive predictive values contains the value zero, and therefore (with  $\alpha = 5\%$ ), the equality of the two positive predictive values is not rejected.

Regarding the negative predictive values, the 95% Wald CI does not contain the value zero, and therefore, we reject the equality of both negative predictive values. Therefore, negative predictive value of FIT is significantly greater than the negative predictive value of FOBT. The negative predictive value of FIT is (with a confidence of 95%) a value between 8.1% and 22.2% greater than the negative predictive value of FOBT. The same conclusions are obtained using the other CIs.



To illustrate the method of calculating the sample size, we are going to suppose that the clinician is interested in calculating the sample size necessary to estimate the difference between the two negative predictive values with a precision  $\phi_v = 0.05$  and  $1 - \alpha = 0.95$ . The 95% Wald CI for  $\delta_v = v_1 - v_2$  is (0.081, 0.222), and the precision is 0.0705 ( $= (0.222 - 0.081)/2$ ). Since  $\phi_v = 0.05 < 0.0705$ , with the sample of 168 individuals, the precision has not been reached, and therefore, the sample size must be calculated. Using the sample of 168 patients as a pilot sample and executing the command “cicpvbdt(68, 18, 1, 13, 4, 1, 2, 61, 0, 0.05)”, it is obtained that  $n_v = 338$ . A sample of 338 patients is necessary to estimate the difference between the two negative predictive values with a precision  $\phi_v = 0.05$  and a confidence of 95%. To the sample of 168 patients, another 170 new patients must be added. The two BDTs and the biopsy should be applied to new patients. Finally, it is necessary to recalculate the CIs with the sample of 338 patients and check if the set precision is verified.

## 7. Discussion

Comparison of the predictive values of two medical tests is a topic of interest in biostatistics. There are several articles that have studied hypothesis tests to solve these problems; however, the comparison of predictive values through confidence intervals has been little studied. In this manuscript, we have studied confidence intervals for the difference and for the ratio of the positive (negative) predictive values of two diagnostic tests under a paired design. We have carried out simulation experiments to study the asymptotic behaviors of the CIs, and we have given general rules of application. These rules are based on the sample size, since this is the only parameter that is set by the researcher, and also on the practical interpretation of the CIs. As a general conclusion, we recommend using the Wald interval for the difference of the two positive (negative) predictive values.

We have also proposed a method, based on the Wald CI for the difference, to calculate the sample size to estimate the difference between the two positive (negative) predictive values with a determined precision and confidence. This method starts from an initial pilot sample, and then the sample size is calculated from the estimators obtained with the initial sample. This method depends on the estimators of the pilot sample, so we have carried out simulation experiments to study the effect of the pilot sample on the sample size. The results obtained in these experiments have shown that the pilot sample does not have any important effect on the calculation of the sample size, and that therefore, the method has practical validity.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/math9131462/s1>.

**Author Contributions:** The two authors have collaborated equally in the realization of this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank the Academic Editor and the anonymous referees for their helpful comments that improved the quality of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

Let  $\boldsymbol{\pi} = (p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00})^T$  be the vector of probabilities of the multinomial distribution; then, the variance–covariance matrix of  $\hat{\boldsymbol{\pi}}$  is  $\Sigma_{\hat{\boldsymbol{\pi}}} = \{\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T\}/n$ . In terms of  $\boldsymbol{\omega}$ , the predictive values are written as

$$\tau_1 = \frac{p_{10} + p_{11}}{Q_1}, \tau_2 = \frac{p_{01} + p_{11}}{Q_2}, v_1 = \frac{q_{00} + q_{01}}{\bar{Q}_1} \text{ and } v_2 = \frac{q_{00} + q_{10}}{\bar{Q}_2},$$

where

$$\begin{aligned} Q_1 &= P(T_1 = 1) = p_{10} + p_{11} + q_{10} + q_{11} = pSe_1 + q(1 - Sp_1), \\ \bar{Q}_1 &= 1 - Q_1 = P(T_1 = 0) = p_{00} + p_{01} + q_{00} + q_{01} = p(1 - Se_1) + qSp_1, \\ Q_2 &= P(T_2 = 1) = p_{01} + p_{11} + q_{01} + q_{11} = pSe_2 + q(1 - Sp_2) \end{aligned}$$

and

$$\bar{Q}_2 = 1 - Q_2 = P(T_2 = 0) = p_{00} + p_{10} + q_{00} + q_{10} = p(1 - Se_2) + qSp_2.$$

Let  $\boldsymbol{\omega} = (\tau_1, \tau_2, v_1, v_2)^T$  be the vector of predictive values; then, applying the delta method, the matrix of the asymptotic variances–covariances of  $\hat{\boldsymbol{\omega}}$  is

$$\Sigma_{\hat{\boldsymbol{\omega}}} = \left( \frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\pi}} \right) \Sigma_{\hat{\boldsymbol{\pi}}} \left( \frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\pi}} \right)^T.$$

Performing the algebraic operations, it is obtained that

$$\begin{aligned} \text{Var}(\hat{\tau}_1) &= \frac{(p_{10} + p_{11})(q_{10} + q_{11})}{nQ_1^3} = \frac{\tau_1 \bar{\tau}_1}{nQ_1}, \\ \text{Var}(\hat{\tau}_2) &= \frac{(p_{01} + p_{11})(q_{01} + q_{11})}{nQ_2^3} = \frac{\tau_2 \bar{\tau}_2}{nQ_2}, \\ \text{Var}(\hat{v}_1) &= \frac{(q_{01} + q_{00})(p_{01} + p_{00})}{nQ_1^3} = \frac{v_1 \bar{v}_1}{nQ_1}, \\ \text{Var}(\hat{v}_2) &= \frac{(q_{00} + q_{10})(p_{00} + p_{10})}{nQ_2^3} = \frac{v_2 \bar{v}_2}{nQ_2}, \\ \text{Cov}(\hat{\tau}_1, \hat{\tau}_2) &= \frac{pq^2 \tau_1 \tau_2 \varepsilon_0 + p^2 q \bar{\tau}_1 \bar{\tau}_2 \varepsilon_1 + \tau_1 \tau_2 \bar{\tau}_1 \bar{\tau}_2 Q_1 Q_2}{npqQ_1 Q_2} \end{aligned}$$

and

$$\text{Cov}(\hat{v}_1, \hat{v}_2) = \frac{pq^2 \bar{v}_1 \bar{v}_2 \varepsilon_0 + p^2 q v_1 v_2 \varepsilon_1 + v_1 v_2 \bar{v}_1 \bar{v}_2 \bar{Q}_1 \bar{Q}_2}{npq\bar{Q}_1 \bar{Q}_2}$$

where  $\bar{\tau}_i = 1 - \tau_i$  and  $\bar{v}_i = 1 - v_i$ , with  $i = 1, 2$ . The estimated variances–covariances are obtained by substituting the parameters for their estimators. Equations (15) and (16) are obtained by substituting in equations

$$\text{Var}(\hat{\delta}_\tau) = \text{Var}(\hat{\tau}_1) + \text{Var}(\hat{\tau}_2) - 2\text{Cov}(\hat{\tau}_1, \hat{\tau}_2)$$

and

$$\text{Var}(\hat{\delta}_v) = \text{Var}(\hat{v}_1) + \text{Var}(\hat{v}_2) - 2\text{Cov}(\hat{v}_1, \hat{v}_2)$$

the variances–covariances by their corresponding expressions obtained previously.

### References

1. Bennett, B.M. On comparison of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics* **1972**, *28*, 793–800. [[CrossRef](#)] [[PubMed](#)]
2. Bennett, B.M. On tests for equality of predictive values for t diagnostic procedures. *Stat. Med.* **1985**, *4*, 535–539. [[CrossRef](#)] [[PubMed](#)]
3. Leisenring, W.; Alonzo, T.; Pepe, M.S. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics* **2000**, *56*, 345–351. [[CrossRef](#)] [[PubMed](#)]
4. Wang, W.; Davis, C.S.; Soong, S.J. Comparison of predictive values of two diagnostic tests from the same sample of subjects using weighted least squares. *Stat. Med.* **2006**, *25*, 2215–2229. [[CrossRef](#)] [[PubMed](#)]

5. Kosinski, A.S. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. *Stat. Med.* **2013**, *32*, 964–977. [[CrossRef](#)] [[PubMed](#)]
6. Tsou, T.S. A new likelihood approach to inference about predictive values of diagnostic tests in paired designs. *Stat. Methods Med. Res.* **2018**, *27*, 541–548. [[CrossRef](#)] [[PubMed](#)]
7. Takahashi, K.; Yamamoto, K. An exact test for comparing two predictive values in small-size clinical trials. *Pharm. Stat.* **2020**, *19*, 31–43. [[CrossRef](#)] [[PubMed](#)]
8. Roldán-Nofuentes, J.A.; Luna del Castillo, J.D.; Montero-Alonso, M.A. Global hypothesis test to simultaneously compare the predictive values of two binary diagnostic tests. *Comput. Stat. Data Anal.* **2012**, *56*, 1161–1173. [[CrossRef](#)]
9. Moskowitz, C.S.; Pepe, M.S. Comparing the predictive values of diagnostic tests: Sample size and analysis for paired study designs. *Clin. Trials* **2006**, *3*, 272–279. [[CrossRef](#)] [[PubMed](#)]
10. Vacek, P.M. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **1985**, *41*, 959–968. [[CrossRef](#)] [[PubMed](#)]
11. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall: Binghamton, NY, USA, 1993.
12. Fieller, E.C. The biological standardization of insulin. *J. R. Stat. Soc.* **1940**, *7*, 1–64. [[CrossRef](#)]
13. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2016; Available online: <https://www.R-project.org/> (accessed on 11 June 2021).