

Article

Global Hypothesis Test to Compare the Predictive Values of Diagnostic Tests Subject to a Case-Control Design

Saad Bouh Regad ^{1,*} and José Antonio Roldán-Nofuentes ² 

¹ Department of Epidemiology and Public Health Research Unit and URMCD, University of Nouakchott Alaasriya, Nouakchott BP 880, Mauritania

² Department of Statistics (Biostatistics), School of Medicine, University of Granada, 18016 Granada, Spain; jaroldan@ugr.es

* Correspondence: regad@una.mr

Abstract: Use of a case-control design to compare the accuracy of two binary diagnostic tests is frequent in clinical practice. This design consists of applying the two diagnostic tests to all of the individuals in a sample of those who have the disease and in another sample of those who do not have the disease. This manuscript studies the comparison of the predictive values of two diagnostic tests subject to a case-control design. A global hypothesis test, based on the chi-square distribution, is proposed to compare the predictive values simultaneously, as well as other alternative methods. The hypothesis tests studied require knowing the prevalence of the disease. Simulation experiments were carried out to study the type I errors and the powers of the hypothesis tests proposed, as well as to study the effect of a misspecification of the prevalence on the asymptotic behavior of the hypothesis tests and on the estimators of the predictive values. The proposed global hypothesis test was extended to the situation in which there are more than two diagnostic tests. The results have been applied to the diagnosis of coronary disease.

Keywords: binary diagnostic test; case-control design; predictive values; type I bivariate binomial distribution



Citation: Regad, S.B.; Roldán-Nofuentes, J.A. Global Hypothesis Test to Compare the Predictive Values of Diagnostic Tests Subject to a Case-Control Design. *Mathematics* **2021**, *9*, 658. <https://doi.org/10.3390/math9060658>

Academic Editor: Maria Laura Manca

Received: 17 February 2021

Accepted: 17 March 2021

Published: 19 March 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/>).

1. Introduction

The main parameters to assess and compare the accuracy of binary diagnostic tests (BDTs) are sensitivity and specificity. The sensitivity (Se) is the probability of the result of the BDT being positive when the individual has the disease, and the specificity (Sp) is the probability of the result of the BDT being negative when the individual does not have the disease. Other parameters that are used to assess and compare two BDTs are the predictive values (PVs). The positive predictive value (PPV) is the probability of an individual having the disease when the result of the BDT is positive, and the negative predictive value (NPV) is the probability of an individual not having the disease when the result of the BDT is negative. The PVs 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. The PVs depend on Se , Sp and on the disease prevalence (p), and are easily calculated applying Bayes Theorem, i.e.,

$$PPV = \frac{p \times Se}{p \times Se + (1 - p) \times (1 - Sp)} \text{ and } NPV = \frac{(1 - p) \times Sp}{p \times (1 - Se) + (1 - p) \times Sp} \quad (1)$$

whereas the Se and the Sp quantify how well the BDT reflects the true disease status (present or absent), the PVs quantify the clinical value of the BDT, since both the individual and the clinician are more interested in knowing how probable it is to have the disease given a BDT result.

The comparison of the performance of two binary diagnostic tests is a topic of special importance in the study of statistical methods for the diagnosis of diseases. This comparison

is made through a paired-design or through a case-control design. The paired design consists of applying the two BDTs and the gold standard to all of the individuals in a single sample. The case-control design consists of applying the two BDTs to all of the individuals in two samples, one made up of individuals who have the disease (case sample) and another made up of individuals who do not have the disease (control sample). The advantages and disadvantages of the case-control design over the paired design can be seen in the book by Pepe [1]. Summarizing, the case-control design has some advantages over the paired design: (a) the case-control design is more efficient in terms of sample size requirements, (b) case-control studies allow for the exploration of subject-related characteristics of the test. Nevertheless, the case-control design has the disadvantage is that by using it we cannot estimate the prevalence of the disease.

In paired designs, the comparison of PVs has been the subject of several studies. Bennett [2,3], Leisenring et al. [4], Wang et al. [5] and Kosinski [6] studied hypothesis tests to independently compare the *PPVs* and the *NPVs* of two BDTs. Moskowitz and Pepe [7] studied the estimation of the PVs through a confidence region. Roldán-Nofuentes et al. [8] studied the joint comparison of the *PPVs* and *NPVs* of two *BDTs*, and proposed a global hypothesis test based on the chi-square distribution to simultaneously compare the PVs of two BDTs.

In a case-control design, Mercado et al. [9] have studied the estimation of the PVs of a BDT, assuming that the prevalence of the disease (*p*) is known. The prevalence can be known from other studies, such as population studies of health services, cohort studies, etc. Mercado et al. have verified through simulation experiments that the confidence interval with the best asymptotic behavior is the logit interval, whose equations are:

$$PPV \in \left(\frac{\exp\left\{\text{logit}(\widehat{PPV}) - z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{PPV}))}\right\}}{1 + \exp\left\{\text{logit}(\widehat{PPV}) - z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{PPV}))}\right\}}, \frac{\exp\left\{\text{logit}(\widehat{PPV}) + z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{PPV}))}\right\}}{1 + \exp\left\{\text{logit}(\widehat{PPV}) + z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{PPV}))}\right\}} \right);$$

and

$$NPV \in \left(\frac{\exp\left\{\text{logit}(\widehat{NPV}) - z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{NPV}))}\right\}}{1 + \exp\left\{\text{logit}(\widehat{NPV}) - z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{NPV}))}\right\}}, \frac{\exp\left\{\text{logit}(\widehat{NPV}) + z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{NPV}))}\right\}}{1 + \exp\left\{\text{logit}(\widehat{NPV}) + z_{1-\alpha/2} \sqrt{\widehat{Var}(\text{logit}(\widehat{NPV}))}\right\}} \right);$$

where \widehat{PPV} and \widehat{NPV} are the estimators of the PVs calculated from Equation (1), $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of the normal standard distribution, and the variances are:

$$\widehat{Var}(\text{logit}(\widehat{PPV})) = \frac{1 - \widehat{Se}}{n_1 \widehat{Se}} + \frac{\widehat{Sp}}{n_2(1 - \widehat{Sp})} \text{ and } \widehat{Var}(\text{logit}(\widehat{NPV})) = \frac{\widehat{Se}}{n_1(1 - \widehat{Se})} + \frac{1 - \widehat{Sp}}{n_2 \widehat{Sp}},$$

where \widehat{Se} and \widehat{Sp} are the estimators of sensitivity and specificity, n_1 is the size of the case sample and n_2 is the size of the control sample.

In this article, we extended the study of Mercado et al. [9] to the case of two BDTs, studying different hypothesis tests to compare the PVs of the two BDTs subject to a case-control design. Subject to a case-control design, the two BDTs are applied to all of the individual in two samples, one of n_1 individuals who have the disease (case sample) and another with n_2 individuals who do not have the disease (control sample). In this design, the sample sizes n_1 and n_2 are set by the researcher. 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. As the PVs depend on the disease prevalence and subject to a case-control design the quotient $n_1 / (n_1 + n_2)$ is not an estimator of the prevalence, in order to estimate and compare the PVs subject to this design it is necessary to know the value of the prevalence of the disease. This value can be obtained from health surveys or from previous studies. Consequently, the methods of comparison of the PVs subject to a paired design cannot be applied when there is a case-control design. In Section 2, we study hypothesis tests to simultaneously compare the PVs of two BDTs subject to a case-control design. A global hypothesis test is studied to simultaneously compare the PVs of the two BDTs, i.e.,

$$H_0 : (PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2),$$

and simultaneous comparison is also studied from individual hypothesis tests, i.e.,

$$H_0 : PPV_1 = PPV_2 \text{ and } H_0 : NPV_1 = NPV_2$$

each of them to the α error and also applying multiple comparison methods. In Section 3, simulation experiments are carried out to study the type I errors and the powers of the hypothesis tests proposed in Section 2, and we study the effect of the misspecification of the prevalence on the asymptotic behavior of the hypothesis tests proposed in Section 2 and on the estimators of the PVs. In Section 4, the results are applied to a real example on the diagnosis of coronary heart disease. In Section 5, the model proposed in Section 2 was extended to the situation in which we compare the PVs of more than two BDTs, and in Section 6 the results are discussed.

2. Global Hypothesis Test

Let us consider two BDTs, Test 1 and Test 2, which are applied to all of the individuals in two samples, one of n_1 individuals who have the disease (case sample) and another of n_2 individuals who do not have it (control sample). Let T_1 and T_2 be two binary variables that model the results of each BDT, in such a way that $T_i = 1$ when the result of the corresponding BDT is positive and $T_i = 0$ when it is negative. In Table 1, we can see the probabilities associated to the application of both BDTs to both types of individuals (cases and controls), as well as the frequencies observed.

Table 1. Probabilities and observed frequencies subject to case-control design.

Probabilities							
Case				Control			
	$T_2 = 1$	$T_2 = 0$	Total		$T_2 = 1$	$T_2 = 0$	Total
$T_1 = 1$	ξ_{111}	ξ_{110}	Se_1	$T_1 = 1$	ξ_{211}	ξ_{210}	$1 - Sp_1$
$T_1 = 0$	ξ_{101}	ξ_{100}	$1 - Se_1$	$T_1 = 0$	ξ_{201}	ξ_{200}	Sp_1
Total	Se_2	$1 - Se_2$	1	Total	$1 - Sp_2$	Sp_2	1
Observed Frequencies							
Case				Control			
	$T_2 = 1$	$T_2 = 0$	Total		$T_2 = 1$	$T_2 = 0$	Total
$T_1 = 1$	n_{111}	n_{110}	$n_{11\cdot}$	$T_1 = 1$	n_{211}	n_{210}	$n_{21\cdot}$
$T_1 = 0$	n_{101}	n_{100}	$n_{10\cdot}$	$T_1 = 0$	n_{201}	n_{200}	$n_{20\cdot}$
Total	$n_{1\cdot 1}$	$n_{1\cdot 0}$	n_1	Total	$n_{2\cdot 1}$	$n_{2\cdot 0}$	n_2

Using the conditional dependence model of Vacek [10], the probabilities given in the table are written as:

$$\xi_{1jk} = Se_1^j (1 - Se_1)^{1-j} Se_2^k (1 - Se_2)^{1-k} + \delta_{jk} \epsilon_1,$$

and

$$\xi_{2jk} = Sp_1^{1-j} (1 - Sp_1)^j Sp_2^{1-k} (1 - Sp_2)^k + \delta_{jk} \epsilon_2,$$

with $j, k = 0, 1$. The parameter ε_1 (ε_2) is the covariance between the two BDTs in cases (controls), where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = -1$ if $j \neq k$, and it is verified that $0 \leq \varepsilon_1 \leq \text{Min}\{Se_1(1 - Se_2), Se_2(1 - Se_1)\}$ and $0 \leq \varepsilon_2 \leq \text{Min}\{Sp_1(1 - Sp_2), Sp_2(1 - Sp_1)\}$. If $\varepsilon_i = 0$ then the two BDTs are conditionally independent on the disease status. In practice, the assumption of the conditional independence is not realistic, and therefore $\varepsilon_1 > 0$ and/or $\varepsilon_2 > 0$. In terms of the probabilities ζ_{ijk} , the sensitivities are written as:

$$Se_1 = \zeta_{111} + \zeta_{110} \text{ and } Se_2 = \zeta_{111} + \zeta_{101},$$

and the specificities are written as:

$$Sp_1 = \zeta_{201} + \zeta_{200} \text{ and } Sp_2 = \zeta_{210} + \zeta_{200}.$$

The estimators of sensitivities are $\hat{Se}_1 = n_{11}./n_1$ and $\hat{Se}_2 = n_{1.1}/n_1$, and the estimators of specificities are $\hat{Sp}_1 = n_{20}./n_2$ and $\hat{Sp}_2 = n_{2.0}/n_2$. The estimators of their variances are $\hat{Var}(\hat{Se}_1) = \hat{Se}_1(1 - \hat{Se}_1)/n_1$, $\hat{Var}(\hat{Se}_2) = \hat{Se}_2(1 - \hat{Se}_2)/n_1$, $\hat{Var}(\hat{Sp}_1) = \hat{Sp}_1(1 - \hat{Sp}_1)/n_2$ and $\hat{Var}(\hat{Sp}_2) = \hat{Sp}_2(1 - \hat{Sp}_2)/n_2$. Therefore, the sensitivities and the specificities are estimated as proportions of marginal totals. In this way, in the case sample we are interested in the marginal frequencies $n_{11}.$ and $n_{1.1}$, and therefore these frequencies are the product of a type I bivariate binomial distribution [11]. In an analogous way, from the control sample, the marginal frequencies $n_{20}.$ and $n_{2.0}$ are the product of a type I bivariate binomial distribution. In the individuals with the disease, the type I bivariate binomial distribution is characterized [11] by Se_1, Se_2 and the correlation coefficient (ρ_1) between T_1 and T_2 . In an analogous way, in the individuals who do not have the disease, the type I bivariate binomial distribution is characterized by Sp_1, Sp_2 and the correlation coefficient (ρ_2) between T_1 and T_2 . Therefore, the proposed model is a parametric model based on the distribution of the marginal frequencies in each 2×2 table. In the individuals with the disease (cases), the correlation coefficient between the two BDTs is:

$$\rho_1 = \frac{\zeta_{111} - Se_1Se_2}{\sqrt{Se_1(1 - Se_1)Se_2(1 - Se_2)}} = \frac{\varepsilon_1}{\sqrt{Se_1(1 - Se_1)Se_2(1 - Se_2)'}}$$

and in the individuals who do not have the disease (controls), the correlation coefficient between the two BDTs is:

$$\rho_2 = \frac{\zeta_{200} - Sp_1Sp_2}{\sqrt{Sp_1(1 - Sp_1)Sp_2(1 - Sp_2)}} = \frac{\varepsilon_2}{\sqrt{Sp_1(1 - Sp_1)Sp_2(1 - Sp_2)'}}$$

It is easy to show that $\hat{\varepsilon}_1 = (n_1n_{111} - n_{11}n_{1.1})/n_1^2$, $\hat{\varepsilon}_2 = (n_2n_{200} - n_{20}n_{2.0})/n_2^2$, $\hat{Cov}(\hat{Se}_1, \hat{Se}_2) = \hat{\varepsilon}_1/n_1$ and $\hat{Cov}(\hat{Sp}_1, \hat{Sp}_2) = \hat{\varepsilon}_2/n_2$. All of the other covariances are zero, since the two samples are independent. The estimators of ρ_1 and ρ_2 are $\hat{\rho}_1 = \frac{n_1n_{111} - n_{11}n_{1.1}}{\sqrt{n_{11}.(n_1 - n_{11})n_{1.1}(n_1 - n_{1.1})}}$ and $\hat{\rho}_2 = \frac{n_2n_{200} - n_{20}n_{2.0}}{\sqrt{n_{20}.(n_2 - n_{20})n_{2.0}(n_2 - n_{2.0})}}$. Assuming that the disease prevalence p is known, the estimators of the predictive values are:

$$\widehat{PPV}_1 = \frac{pn_2n_{11.}}{pn_2n_{11.} + qn_1(n_2 - n_{20.})} \text{ and } \widehat{NPV}_1 = \frac{qn_1n_{20.}}{pn_2(n_1 - n_{11.}) + qn_1n_{20.}'},$$

for Test 1, and

$$\widehat{PPV}_2 = \frac{pn_2n_{1.1}}{pn_2n_{1.1} + qn_1(n_2 - n_{2.0})} \text{ and } \widehat{NPV}_2 = \frac{qn_1n_{2.0}}{pn_2(n_1 - n_{1.1}) + qn_1n_{2.0}'}$$

for Test 2, where $q = 1 - p$. Let the variance-covariance matrixes be defined as:

$$\Sigma_{\hat{se}} = \begin{pmatrix} \text{Var}(\hat{Se}_1) & \text{Cov}(\hat{Se}_1, \hat{Se}_2) \\ \text{Cov}(\hat{Se}_1, \hat{Se}_2) & \text{Var}(\hat{Se}_2) \end{pmatrix}$$

and

$$\Sigma_{\hat{s}_p} = \begin{pmatrix} Var(\hat{S}_{p1}) & Cov(\hat{S}_{p1}, \hat{S}_{p2}) \\ Cov(\hat{S}_{p1}, \hat{S}_{p2}) & Var(\hat{S}_{p2}) \end{pmatrix}.$$

Let $\theta = (Se_1, Se_2, Sp_1, Sp_2)^T$ be a vector whose components are the sensitivities and the specificities, and let $\omega = (PPV_1, PPV_2, NPV_1, NPV_2)^T$ be a vector whose components are the PVs. The variance-covariance matrix of $\hat{\theta}$ is:

$$\Sigma_{\hat{\theta}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \Sigma_{\hat{s}_e} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \Sigma_{\hat{s}_p} \tag{2}$$

where \otimes is the Kronecker product. Applying the delta method, the matrix of variances-covariances of $\hat{\omega}$ is:

$$\Sigma_{\hat{\omega}} = \left(\frac{\partial \omega}{\partial \theta} \right) \Sigma_{\hat{\theta}} \left(\frac{\partial \omega}{\partial \theta} \right)^T. \tag{3}$$

Expressions of the variances-covariances of the PVs can be seen in Appendix A. The PVs of each BDT depend on the same parameters, the sensitivity and the specificity of the test and disease prevalence, and therefore they are related parameters. Consequently, the PVs of the two BDTs can be compared simultaneously. The global hypothesis test to simultaneously compare the PVs of the two BDTs is:

$$\begin{aligned} H_0 &: PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2 \\ H_1 &: \text{at least one equality is not true,} \end{aligned}$$

which is equivalent to the hypothesis test:

$$H_0 : \mathbf{A}\omega = 0 \text{ vs } H_1 : \mathbf{A}\omega \neq 0 \tag{4}$$

where \mathbf{A} is a complete range matrix sized 2×4 whose elements are known constants, i.e.:

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

As the vector $\hat{\omega}$ is distributed asymptotically according to a multivariate normal distribution, i.e., $\sqrt{n_1 + n_2} \left(\hat{\omega} - \omega \right) \xrightarrow{n_1 + n_2 \rightarrow \infty} N(0, \Sigma_{\omega})$, then the test statistic for the global hypothesis test (4) is:

$$Q^2 = \hat{\omega}^T \mathbf{A}^T \left(\mathbf{A} \hat{\Sigma}_{\hat{\omega}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\omega}, \tag{5}$$

which is distributed asymptotically according to Hotelling’s T-squared distribution with a dimension 2 and $n_1 + n_2$ degrees of freedom, where 2 is the dimension of the vector $\mathbf{A}\hat{\omega}$. When $n_1 + n_2$ is large, the statistic Q^2 is approximately distributed according to a central chi-square distribution with 2 degrees of freedom when the null hypothesis is true.

On the other hand, the individual comparison of the positive (negative) predictive values is solved with the hypothesis test:

$$H_0 : PV_1 = PV_2 \text{ vs } H_1 : PV_1 \neq PV_2,$$

where PV is PPV or NPV. Based on the asymptotic normality of the estimators, the test statistic for this hypothesis test is:

$$z = \frac{\widehat{PV}_1 - \widehat{PV}_2}{\sqrt{\hat{V}ar(\widehat{PV}_1) + \hat{V}ar(\widehat{PV}_2) - 2\hat{C}ov(\widehat{PV}_1, \widehat{PV}_2)}}, \tag{6}$$

which is distributed asymptotically according to a normal standard distribution, and where the variances-covariances are obtained from the Equation (3) (see Appendix A).

The global hypothesis test $H_0 : \mathbf{A}\boldsymbol{\omega} = 0$ simultaneously compares the PPVs and the NPVs of the two BDTs. Some alternative methods to this global hypothesis test, based on the individual hypothesis tests, are: (1) testing the hypotheses $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$ (Equation (6)) each one to an α error; (2) testing the hypotheses $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$ (Equation (6)) and applying a multiple comparison method such as the Bonferroni method [12] or the Holm method [13], which are methods that are very easy to apply based on the p -values. Bonferroni method [12] consists of solving each individual hypothesis test to an error equal to $\alpha/2$. The Holm method is a step-down method which is based on Bonferroni method but is more conservative. In Appendix B, the Holm method [13] is summarized.

3. Simulation Experiments

Simulation experiments were carried out to study the type I errors and the powers of the four methods proposed to simultaneously compare the predictive values: the global hypothesis test based on the chi-square distribution (Equation (5)), the individual hypothesis tests each one to an α error (Equation (6)), the individual hypothesis tests (Equation (6)) applying the Bonferroni method and the individual hypothesis tests (Equation (6)) applying the Holm method. We have also studied the effect of a misspecification of the prevalence on the asymptotic behavior of these methods and on the estimators of the PVs.

The experiments were designed setting the values of the PVs. For each BDT, we took as PVs the values 0.60, 0.65, . . . , 0.90, 0.95, and as disease prevalence we took the values 10%, 25% and 50%. Based on the PVs and the prevalence, Se and Sp of each BDT were calculated from the Equation (1), only considering those cases in which the solutions are between 0 and 1. As values of the correlation coefficients ρ_1 and ρ_2 we took low values (25% of the maximum value), intermediate (50% of the maximum value) and high (75% of the maximum value), where the maximum value of each correlation coefficient is: $\max(\rho_1) = \frac{\min\{Se_1(1-Se_2), (1-Se_1)Se_2\}}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}}$ and $\max(\rho_2) = \frac{\min\{Sp_1(1-Sp_2), (1-Sp_1)Sp_2\}}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}}$ respectively. As sample sizes, we took the values $n_i = (50, 75, 100, 200, 500)$. The simulation experiments were carried out with R [14], using the “bindata” package [15] to generate the samples of each type I bivariate binomial distribution.

Regarding the random samples, these were generated in the following way. Firstly, once the values of the PVs and of the prevalence were set, we calculated the sensitivities, the specificities and the maximum values of the coefficients ρ_1 and ρ_2 . We then generated 10,000 random samples from a type I bivariate binomial distribution with a sample size n_1 , probabilities Se_1 and Se_2 , and correlation coefficient ρ_1 . Similarly, we generated another 10,000 random samples from a type I bivariate binomial distribution with a sample size n_2 , probabilities Sp_1 and Sp_2 , and correlation coefficient ρ_2 . In this way, we obtained the marginal frequencies n_{11} . and $n_{1.1}$ (n_{20} . and $n_{2.0}$) of each one of the 10,000 case (control) samples. The rest of the marginal frequencies were easily calculated: n_{10} . = $n_1 - n_{11}$., $n_{1.0}$ = $n_1 - n_{1.1}$, n_{21} . = $n_2 - n_{20}$. and $n_{2.1}$ = $n_2 - n_{2.0}$. In order to construct the 2×2 table of each case sample, we generated a random value n_{111} from a doubly truncated binomial distribution of parameters n_1 and $\zeta_{111} = Se_1Se_2 + \varepsilon_1$, with n_{11} . + $n_{1.1} - n_1 \leq n_{111} \leq \min(n_{11}$., $n_{1.1}$.) This is necessary so that the sum of the frequencies leads to the marginal totals randomly generated through the type I bivariate binomial distribution. In the same way, in order to construct the 2×2 table of each control sample, we generated a random value n_{200} from a doubly truncated binomial distribution of parameters n_2 and $\zeta_{200} = Sp_1Sp_2 + \varepsilon_2$, with n_{20} . + $n_{2.0} - n_2 \leq n_{200} \leq \min(n_{20}$., $n_{2.0}$.) For each one of the 10,000 case (control) samples, once we have generated the values n_{11} ., $n_{1.1}$ and n_{111} (n_{20} ., $n_{2.0}$ and n_{200}) it is easy to construct the complete 2×2 table. Thus, n_{110} = $n_1 - n_{11}$., n_{101} = $n_{1.1} - n_{111}$ and n_{100} = n_{10} . - n_{101} for the case samples, and n_{201} = n_{20} . - n_{200} ., n_{210} = $n_{2.0} - n_{200}$ and n_{211} = n_{21} . - n_{210} for the control samples. For the experiments $\alpha = 5\%$ was set. Moreover, all of the samples were generated in such a way that in all

of them the parameters and the variances-covariances can be estimated. If in a random sample it is obtained that $n_{i10} = n_{i01} = 0$, with $i = 1, 2$, then $\hat{S}e_i = \hat{S}p_i = 1$ and $\hat{V}ar(\hat{S}e_i) = \hat{V}ar(\hat{S}p_i) = 0$, and therefore the test statistic $Q^2 = \hat{\omega}^T \mathbf{A}^T (\mathbf{A} \hat{\Sigma}_{\hat{\omega}} \mathbf{A}^T)^{-1} \mathbf{A} \hat{\omega}$ cannot be calculated since $\mathbf{A} \hat{\Sigma}_{\hat{\omega}} \mathbf{A}^T$ is a non-singular matrix. This problem occurs mainly when the sample size is small or moderate. In this situation, the sample has been discarded and another is generated in its place until the 10,000 samples are obtained.

3.1. Type I Errors and Powers

In Tables 2 and 3, we can see some results obtained for the type I errors of the global test and of the alternative methods proposed in Section 2. In these tables, we can only see the results for the global test, the individual comparisons with $\alpha = 5\%$ and with the Bonferroni method. The results obtained with the Holm method are not shown as they are practically the same as those obtained with the Bonferroni method. From the results obtained we can draw the following conclusions. In general terms, the type I error of the global hypothesis test fluctuates around the nominal error, especially in the case of samples sized $n_i \geq 100$, depending on the prevalence and the correlations between the two BDTs. For samples with smaller sizes ($n_i \leq 75$), the type I error of the global test is lower than $\alpha = 5\%$. The correlations between the two BDTs have an important effect on the type I error of the global test, with a decrease in the type I error when there is an increase in the correlation coefficients.

Table 2. Type I errors for $PPV_1 = PPV_2 = 0.70$ and $NPV_1 = NPV_2 = 0.95$.

$Se_1 = 0.5385, Sp_1 = 0.9744, Se_2 = 0.5385, Sp_2 = 0.9744, p = 10\%$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.031	0.051	0.029	0.027	0.048	0.027	0.004	0.013	0.004
50	75	0.029	0.059	0.029	0.025	0.051	0.026	0.004	0.017	0.005
50	100	0.028	0.063	0.030	0.029	0.061	0.028	0.008	0.018	0.007
75	75	0.023	0.061	0.026	0.031	0.056	0.028	0.015	0.034	0.017
100	100	0.027	0.063	0.029	0.023	0.052	0.024	0.020	0.043	0.019
200	200	0.044	0.086	0.045	0.032	0.063	0.031	0.025	0.050	0.026
500	500	0.055	0.107	0.056	0.058	0.102	0.057	0.040	0.077	0.039
$Se_1 = 0.8615, Sp_1 = 0.8769, Se_2 = 0.8615, Sp_2 = 0.8769, p = 25\%$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.048	0.094	0.046	0.018	0.047	0.018	0.001	0.007	0.002
50	75	0.053	0.100	0.051	0.025	0.063	0.026	0.002	0.012	0.003
50	100	0.053	0.106	0.057	0.034	0.076	0.032	0.008	0.023	0.008
75	75	0.059	0.105	0.055	0.039	0.087	0.037	0.007	0.016	0.006
100	100	0.059	0.117	0.059	0.056	0.102	0.054	0.011	0.040	0.010
200	200	0.058	0.099	0.057	0.048	0.094	0.049	0.044	0.090	0.042
500	500	0.052	0.098	0.053	0.051	0.101	0.052	0.049	0.090	0.048
$Se_1 = 0.9692, Sp_1 = 0.5846, Se_2 = 0.9692, Sp_2 = 0.5846, p = 50\%$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.026	0.049	0.026	0.025	0.061	0.026	0.006	0.017	0.006
50	75	0.020	0.049	0.023	0.019	0.052	0.024	0.007	0.028	0.010
50	100	0.019	0.043	0.023	0.016	0.045	0.019	0.010	0.034	0.014
75	75	0.024	0.065	0.027	0.020	0.051	0.027	0.012	0.038	0.017
100	100	0.028	0.066	0.029	0.021	0.052	0.025	0.012	0.042	0.019
200	200	0.047	0.088	0.044	0.034	0.074	0.032	0.021	0.058	0.026
500	500	0.052	0.099	0.052	0.050	0.097	0.049	0.037	0.077	0.034

Global: global hypothesis test based on the chi-square distribution. $\alpha = 5\%$: individual hypothesis tests each one to an error of 5%. Bonf.: Bonferroni method.

Regarding the method based on the individual hypothesis tests $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$ to an error $\alpha = 5\%$ each one of them, the type I error may clearly overwhelm the nominal error (a situation that we have considered when the type I error is greater than 7%), especially when the correlations are not high. Consequently, this method may lead to erroneous results (false significances) and, therefore, should not be used. As for solving the global test from the individual tests applying the Bonferroni (Holm) method, the type I error has a very similar behavior to that of the global hypothesis test.

Table 3. Type I errors for $PPV_1 = PPV_2 = 0.85$ and $NPV_1 = NPV_2 = 0.95$.

$Se_1 = 0.5312, Sp_1 = 0.9896, Se_2 = 0.5312, Sp_2 = 0.9896, p = 10\%, 0 \leq \rho_1 \leq 1, 0 \leq \rho_2 \leq 1$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.033	0.056	0.034	0.020	0.051	0.024	0.004	0.014	0.004
50	75	0.024	0.049	0.024	0.026	0.050	0.025	0.005	0.019	0.006
50	100	0.032	0.057	0.033	0.030	0.056	0.030	0.004	0.016	0.004
75	75	0.034	0.054	0.033	0.025	0.052	0.026	0.014	0.036	0.015
100	100	0.027	0.055	0.026	0.027	0.055	0.026	0.017	0.041	0.017
200	200	0.033	0.059	0.031	0.025	0.050	0.024	0.022	0.055	0.021
500	500	0.046	0.087	0.049	0.031	0.068	0.033	0.018	0.050	0.024
$Se_1 = 0.85, Sp_1 = 0.95, Se_2 = 0.85, Sp_2 = 0.95, p = 25\%, 0 \leq \rho_1 \leq 1, 0 \leq \rho_2 \leq 1$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.023	0.058	0.022	0.005	0.030	0.007	0.001	0.005	0.001
50	75	0.037	0.077	0.036	0.014	0.039	0.015	0.001	0.008	0.001
50	100	0.049	0.092	0.048	0.022	0.056	0.022	0.001	0.007	0.002
75	75	0.042	0.087	0.041	0.025	0.055	0.025	0.004	0.014	0.004
100	100	0.048	0.095	0.043	0.028	0.066	0.027	0.005	0.025	0.005
200	200	0.033	0.059	0.031	0.025	0.050	0.024	0.022	0.055	0.021
500	500	0.048	0.097	0.046	0.056	0.101	0.051	0.050	0.099	0.049
$Se_1 = 0.9562, Sp_1 = 0.8312, Se_2 = 0.9562, Sp_2 = 0.8312, p = 50\%, 0 \leq \rho_1 \leq 1, 0 \leq \rho_2 \leq 1$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.25$			$\rho_1 = 0.50 \rho_2 = 0.50$			$\rho_1 = 0.75 \rho_2 = 0.75$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.031	0.072	0.031	0.014	0.041	0.015	0.001	0.007	0.001
50	75	0.032	0.069	0.033	0.022	0.049	0.022	0.005	0.015	0.005
50	100	0.025	0.057	0.026	0.025	0.064	0.026	0.008	0.025	0.008
75	75	0.038	0.081	0.037	0.027	0.054	0.025	0.006	0.017	0.006
100	100	0.039	0.084	0.038	0.031	0.073	0.030	0.008	0.030	0.009
200	200	0.033	0.059	0.031	0.025	0.050	0.024	0.022	0.055	0.021
500	500	0.051	0.099	0.049	0.050	0.097	0.047	0.043	0.087	0.042

Global: global hypothesis test based on the chi-square distribution. $\alpha = 5\%$: individual hypothesis tests, each one to an error of 5%. Bonf.: Bonferroni method.

Regarding the powers of the hypothesis tests, in Tables 4 and 5 we can see some of the results obtained for the global test and other alternative methods. The results obtained with the Holm method are not shown as they are practically the same as those obtained with the Bonferroni method. The power of the global hypothesis test is calculated as the proportion of samples in which it is accepted that $PPV_1 \neq PPV_2$ or $NPV_1 \neq NPV_2$ (being true that $PPV_1 \neq PPV_2$ or $NPV_1 \neq NPV_2$). From the results, the following conclusions are obtained. The disease prevalence has an important effect on the power of each one of the methods to solve the global test, and the power increases with an increase in the prevalence. Regarding the correlations ρ_1 and ρ_2 , these do not have a clear effect on the power, and the power increases sometimes and decreases other times when the correlations increase. In general terms, when the prevalence is small ($p = 10\%$) we need large samples ($n_i > 500$) so that the power of the global hypothesis test is greater than 80%; for a prevalence of 25% with sample sizes $n_i \geq 200$ we obtain a power greater than 80%; and for a very large

prevalence ($p = 50\%$) with sample sizes $n_i \geq 50$ we obtain a very higher power, greater than 80%–90%, depending on the difference between the PVs.

Table 4. Powers for $PPV_1 = 0.75, NPV_1 = 0.95, PPV_2 = 0.60$ and $NPV_2 = 0.95$.

$Se_1 = 0.5357, Sp_1 = 0.9802, Se_2 = 0.5455, Sp_2 = 0.9596, p = 10\%, 0 \leq \rho_1 \leq 0.9805, 0 \leq \rho_2 \leq 0.6933$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.17$			$\rho_1 = 0.49 \rho_2 = 0.35$			$\rho_1 = 0.74 \rho_2 = 0.52$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.025	0.056	0.031	0.023	0.049	0.024	0.005	0.019	0.007
50	75	0.037	0.077	0.036	0.029	0.063	0.030	0.010	0.030	0.011
50	100	0.054	0.103	0.052	0.042	0.084	0.038	0.019	0.046	0.016
75	75	0.038	0.078	0.038	0.032	0.066	0.033	0.018	0.042	0.018
100	100	0.053	0.098	0.047	0.044	0.081	0.037	0.031	0.063	0.026
200	200	0.199	0.276	0.180	0.208	0.286	0.181	0.168	0.252	0.138
500	500	0.495	0.575	0.462	0.591	0.668	0.556	0.720	0.785	0.678
$Se_1 = 0.8571, Sp_1 = 0.9048, Se_2 = 0.8727, Sp_2 = 0.8061, p = 25\%, 0 \leq \rho_1 \leq 0.9354, 0 \leq \rho_2 \leq 0.6614$										
n_1	n_2	$\rho_1 = 0.23 \rho_2 = 0.17$			$\rho_1 = 0.47 \rho_2 = 0.33$			$\rho_1 = 0.70 \rho_2 = 0.50$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.259	0.335	0.230	0.254	0.345	0.230	0.195	0.334	0.210
50	75	0.409	0.496	0.378	0.454	0.543	0.424	0.467	0.606	0.470
50	100	0.505	0.584	0.462	0.598	0.677	0.556	0.683	0.776	0.675
75	75	0.416	0.498	0.382	0.469	0.557	0.436	0.501	0.608	0.476
100	100	0.528	0.606	0.488	0.625	0.699	0.579	0.718	0.793	0.685
200	200	0.822	0.862	0.790	0.891	0.923	0.873	0.974	0.983	0.964
500	500	0.996	0.999	0.996	1	1	1	1	1	1
$Se_1 = 0.9643, Sp_1 = 0.6786, Se_2 = 0.9818, Sp_2 = 0.3455, p = 50\%, 0 \leq \rho_1 \leq 0.7071, 0 \leq \rho_2 \leq 0.50$										
n_1	n_2	$\rho_1 = 0.18 \rho_2 = 0.13$			$\rho_1 = 0.35 \rho_2 = 0.25$			$\rho_1 = 0.53 \rho_2 = 0.38$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.890	0.939	0.893	0.935	0.969	0.941	0.977	0.989	0.978
50	75	0.978	0.990	0.977	0.995	0.997	0.993	0.999	0.999	0.999
50	100	0.995	0.998	0.995	0.999	0.998	0.999	1	1	1
75	75	0.984	0.992	0.983	0.995	0.999	0.994	0.999	1	0.999
100	100	0.998	0.999	0.998	1	1	0.999	1	1	1
200	200	1	1	1	1	1	1	1	1	1
500	500	1	1	1	1	1	1	1	1	1

Global: global hypothesis test based on the chi-square distribution. $\alpha = 5\%$: individual hypothesis tests, each one to an error of 5%. Bonf.: Bonferroni method.

The power of the method based on the individual hypothesis tests to an error $\alpha = 5\%$ is greater than that of the global test based on the chi-square distribution due to the fact that its type I error is also greater. Regarding the hypothesis test based on the individual tests with the Bonferroni method, in general terms, its power is very similar to that of the global test when the sample sizes are large. When the sample sizes are small or moderate, in general terms and depending on prevalence and correlations, the power of the global test is slightly greater than that of the individual tests with the Bonferroni method. The same conclusions are obtained when the Holm method is applied (whose results are almost identical to those of the Bonferroni method).

Table 5. Powers for $PPV_1 = 0.95$ $NPV_1 = 0.95$, $PPV_2 = 0.75$ and $NPV_2 = 0.95$.

$Se_1 = 0.5278, Sp_1 = 0.9969, Se_2 = 0.5357, Sp_2 = 0.9802, p = 10\%, 0 \leq \rho_1 \leq 0.9841, 0 \leq \rho_2 \leq 0.3910$										
n_1	n_2	$\rho_1 = 0.25 \rho_2 = 0.10$			$\rho_1 = 0.49 \rho_2 = 0.19$			$\rho_1 = 0.74 \rho_2 = 0.29$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.030	0.059	0.030	0.019	0.048	0.020	0.007	0.019	0.008
50	75	0.031	0.063	0.032	0.023	0.054	0.023	0.009	0.024	0.009
50	100	0.033	0.064	0.033	0.030	0.063	0.030	0.010	0.031	0.009
75	75	0.033	0.057	0.032	0.025	0.055	0.025	0.015	0.036	0.015
100	100	0.034	0.067	0.033	0.026	0.059	0.026	0.026	0.054	0.026
200	200	0.123	0.182	0.094	0.122	0.177	0.095	0.108	0.168	0.078
500	500	0.666	0.770	0.662	0.669	0.781	0.667	0.699	0.811	0.696
$Se_1 = 0.8444, Sp_1 = 0.9852, Se_2 = 0.8571, Sp_2 = 0.9048, p = 25\%, 0 \leq \rho_1 \leq 0.9511, 0 \leq \rho_2 \leq 0.3779$										
n_1	n_2	$\rho_1 = 0.24 \rho_2 = 0.09$			$\rho_1 = 0.48 \rho_2 = 0.19$			$\rho_1 = 0.71 \rho_2 = 0.28$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.172	0.247	0.142	0.154	0.234	0.127	0.107	0.198	0.097
50	75	0.422	0.537	0.396	0.398	0.521	0.378	0.365	0.541	0.367
50	100	0.627	0.734	0.615	0.641	0.755	0.638	0.674	0.779	0.653
75	75	0.434	0.549	0.400	0.432	0.555	0.410	0.402	0.552	0.391
100	100	0.635	0.753	0.634	0.655	0.774	0.656	0.666	0.796	0.683
200	200	0.965	0.981	0.964	0.977	0.987	0.974	0.989	0.994	0.988
500	500	1	1	1	1	1	1	1	1	1
$Se_1 = 0.95, Sp_1 = 0.95, Se_2 = 0.9643, Sp_2 = 0.6786, p = 50\%, 0 \leq \rho^+ \leq 0.8388, 0 \leq \rho^- \leq 0.3333$										
n_1	n_2	$\rho_1 = 0.21 \rho_2 = 0.08$			$\rho_1 = 0.42 \rho_2 = 0.17$			$\rho_1 = 0.63 \rho_2 = 0.25$		
		Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	0.929	0.969	0.942	0.954	0.983	0.966	0.965	0.992	0.978
50	75	0.994	0.998	0.995	0.999	0.999	0.999	0.999	1	0.999
50	100	1	1	1	1	1	1	1	1	1
75	75	0.995	0.998	0.995	0.997	0.999	0.998	1	1	1
100	100	1	1	1	1	1	1	1	1	1
200	200	1	1	1	1	1	1	1	1	1
500	500	1	1	1	1	1	1	1	1	1

Global: global hypothesis test based on the chi-square distribution. $\alpha = 5\%$: individual hypothesis tests, each one to an error of 5%. Bonf.: Bonferroni method.

The graphs in Figure 1 show the powers of the three methods when $PPV_1 = 0.90$, $NPV_1 = (0.80, 0.85, 0.90, 0.95)$, $PPV_2 = 0.85$ and $NPV_2 = 0.90$, for different sample sizes $n_1 = n_2 = (50, 100, 200)$, $p = (25\%, 50\%)$ and values intermediate of the correlation coefficients. These graphs show that when NPV_1 varies and the rest of the PVs are constant, the powers decrease when the prevalence increases. Similarly, the graphs in Figure 2 show the powers of the three methods when $PPV_1 = (0.80, 0.85, 0.90, 0.95)$, $NPV_1 = 0.95$, $PPV_2 = 0.60$ and $NPV_2 = 0.95$, for different sample sizes $n_1 = n_2 = (50, 100, 200)$, $p = (10\%, 25\%)$ and values intermediate of the correlation coefficients. These graphs show that when the PPV_1 varies and the rest of the PVs are constant, the power of each method increases when the prevalence increases.

As conclusions of the results obtained in the simulation experiments, the global hypothesis test based on the chi-square distribution behaves well in terms of the type I error (it does not overwhelm the nominal error of 5%), the same as the individual tests along with the Bonferroni (Holm) method. The method based on the individual tests to a global error $\alpha = 5\%$ should not be used as it may clearly overwhelm the nominal error.

In the simulation experiments, the proportion of times that $PPV_1 \neq PPV_2$ and that $NPV_1 \neq NPV_2$ are correctly concluded has also been studied. This issue is of special interest when the alternative hypothesis of the global test is true, as it can be a valid method to investigate the causes of significance. The study was carried out by applying the individual hypothesis tests together with the Bonferroni (Holm) method. Individual tests to an α error have not been considered as they have a type I error that can exceed the nominal

error. If it is verified that $PV_1 \neq PV_2$, then this study is equivalent to studying the power of the individual test $H_0 : PV_1 = PV_2$ to an $\alpha/2$ error (since the Bonferroni method has been applied), where PV_i is PPV_i or NPV_i . If it is verified that $PV_1 = PV_2$ then this study is equivalent to studying the type I error of the individual test $H_0 : PV_1 = PV_2$ to an $\alpha/2$ error. In the scenarios considered in Tables 4 and 5 it is verified that $PPV_1 \neq PPV_2$ and that $NPV_1 = NPV_2$. Therefore, for these two scenarios, the power of the test $H_0 : PPV_1 = PPV_2$ and the type I error of the test $H_0 : NPV_1 = NPV_2$ have been studied, each with an error equal to $\alpha/2 = 2.5\%$. Tables 6 and 7 show the results obtained applying the Bonferroni method. The results obtained with the Holm method are not shown as they are practically the same as those obtained with the Bonferroni method.

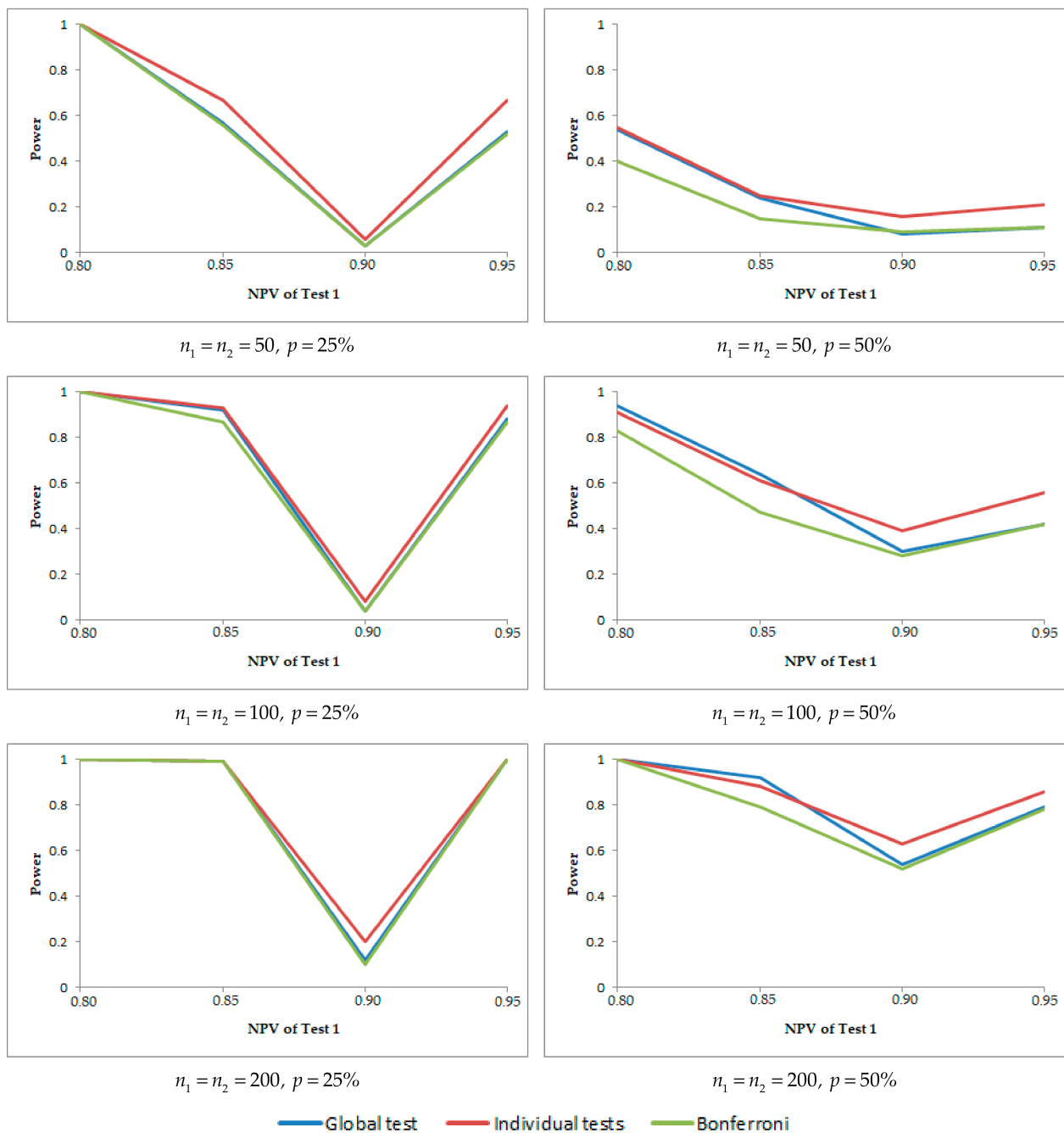


Figure 1. Powers of the three methods when the negative predictive value (NPV) of a binary diagnostic test varies and the rest of the PVs are constant.

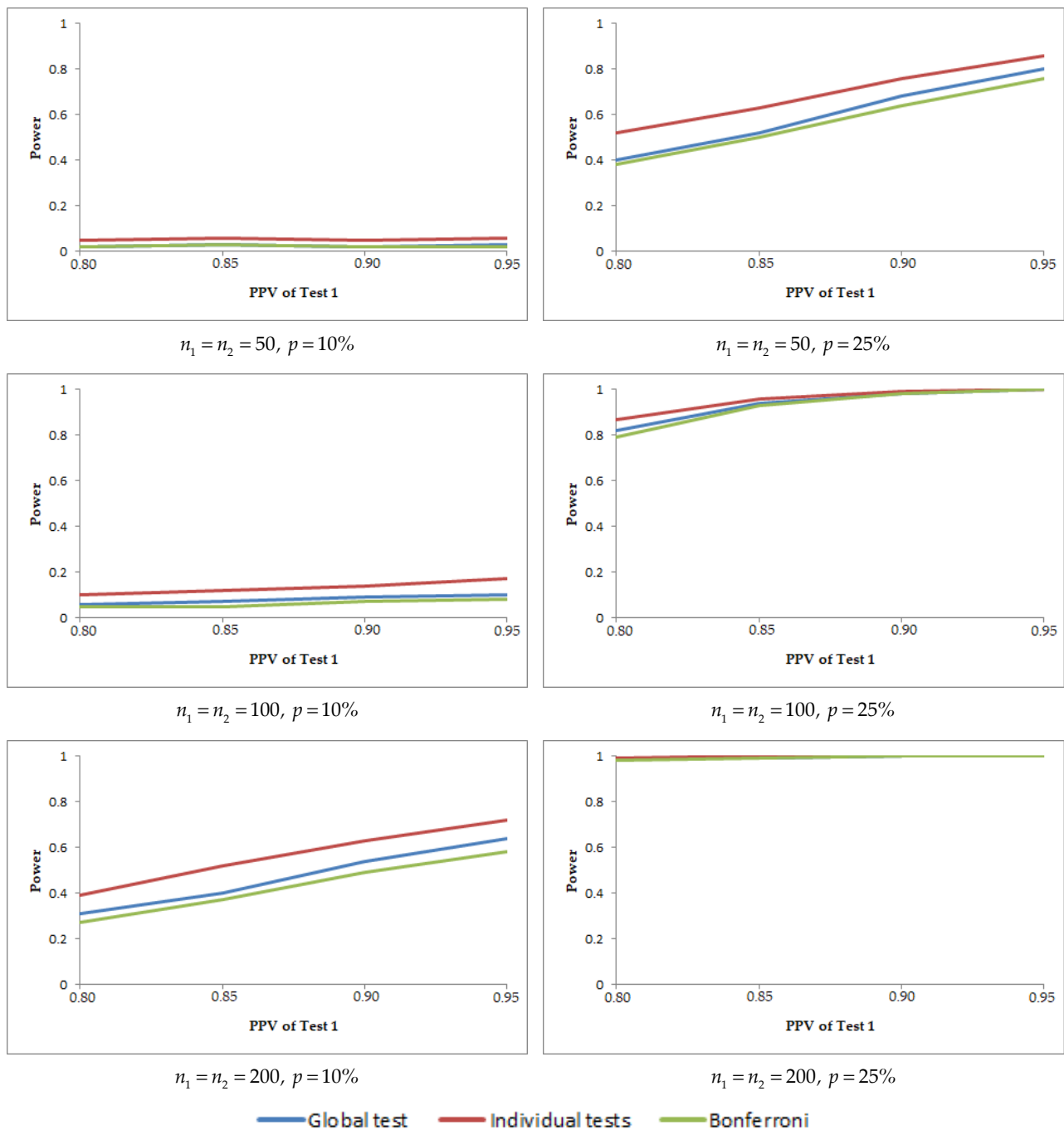


Figure 2. Powers of the three methods when the positive predictive value (PPV) of a binary diagnostic test varies and the rest of the PVs are constant.

In general terms, the hypothesis test $H_0 : PPV_1 = PPV_2$ has a high power when the sample sizes are moderate or high, depending on the prevalence and the correlation coefficients. Its behavior is very similar to that of the global hypothesis test. With respect to the test $H_0 : NPV_1 = NPV_2$, its type I error fluctuates around the nominal error (2.5%) when the sample sizes are moderate or large, depending on the prevalence of the correlation coefficients. In general terms, the hypothesis tests $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$ have a good asymptotic behavior, both in terms of power and type I error.

Table 6. Power of the test $H_0 : PPV_1 = PPV_2$ and type I error of the test $H_0 : NPV_1 = NPV_2$ when $PPV_1 = 0.75$, $NPV_1 = 0.95$, $PPV_2 = 0.60$ and $NPV_2 = 0.95$.

$Se_1 = 0.5357, Sp_1 = 0.9802, Se_2 = 0.5455, Sp_2 = 0.9596, p = 10\%, 0 \leq \rho_1 \leq 0.9805, 0 \leq \rho_2 \leq 0.6933$							
n_1	n_2	$\rho_1 = 0.25 \quad \rho_2 = 0.17$		$\rho_1 = 0.49 \quad \rho_2 = 0.35$		$\rho_1 = 0.74 \quad \rho_2 = 0.52$	
		Power	Type I error	Power	Type I error	Power	Type I error
50	50	0.001	0.030	0.001	0.024	0.001	0.007
50	75	0.007	0.029	0.004	0.027	0.002	0.001
50	100	0.021	0.029	0.013	0.027	0.008	0.009
75	75	0.008	0.031	0.005	0.030	0.003	0.018
100	100	0.026	0.026	0.017	0.026	0.010	0.021
200	200	0.160	0.024	0.159	0.028	0.119	0.024
500	500	0.449	0.026	0.544	0.023	0.672	0.020
$Se_1 = 0.8571, Sp_1 = 0.9048, Se_2 = 0.8727, Sp_2 = 0.8061, p = 25\%, 0 \leq \rho_1 \leq 0.9354, 0 \leq \rho_2 \leq 0.6614$							
n_1	n_2	$\rho_1 = 0.23 \quad \rho_2 = 0.17$		$\rho_1 = 0.47 \quad \rho_2 = 0.33$		$\rho_1 = 0.70 \quad \rho_2 = 0.50$	
		Power	Type I error	Power	Type I error	Power	Type I error
50	50	0.211	0.025	0.223	0.011	0.210	0.002
50	75	0.346	0.021	0.420	0.008	0.470	0.002
50	100	0.448	0.026	0.551	0.010	0.675	0.001
75	75	0.352	0.030	0.435	0.020	0.479	0.005
100	100	0.472	0.027	0.559	0.025	0.683	0.010
200	200	0.785	0.023	0.870	0.027	0.961	0.019
500	500	0.996	0.024	0.999	0.026	1	0.027
$Se_1 = 0.9643, Sp_1 = 0.6786, Se_2 = 0.9818, Sp_2 = 0.3455, p = 50\%, 0 \leq \rho_1 \leq 0.7071, 0 \leq \rho_2 \leq 0.50$							
n_1	n_2	$\rho_1 = 0.18 \quad \rho_2 = 0.13$		$\rho_1 = 0.35 \quad \rho_2 = 0.25$		$\rho_1 = 0.53 \quad \rho_2 = 0.38$	
		Power	Type I error	Power	Type I error	Power	Type I error
50	50	0.893	0.003	0.941	0.002	0.980	0.001
50	75	0.977	0.003	0.991	0.002	0.999	0.001
50	100	0.990	0.002	0.999	0.002	1	0.001
75	75	0.983	0.005	0.993	0.002	1	0.003
100	100	0.997	0.005	1	0.004	1	0.002
200	200	1	0.012	1	0.006	1	0.007
500	500	1	0.024	1	0.024	1	0.019

From the results obtained in the simulation experiments, we propose the following method to compare the PVs of two BDTs subject to a case-control design: (1) Applying the global hypothesis test based on the chi-square distribution (Equation (5)) to an α error; (2) If the global hypothesis test is not significant, the equality hypothesis of the PVs is not rejected; if the global hypothesis test is significant to an α error, the investigation of the causes of the significance is made by testing the individual tests (Equation (6)) and applying the Bonferroni method or the Holm method to an α error. Therefore, if the global test is significant, the investigation of the significance consists in solving the individual hypothesis tests $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$, each of them to an $\alpha/2$ error (Bonferroni method) or applying Holm method.

This method to simultaneously compare the PVs is very similar to other methods used in other statistical models, such as the analysis of variance: first the global test is resolved to an α error, and if it is significant then the causes of significance are investigated from pairwise comparisons and the application of a multiple comparison method.

Table 7. Power of the test $H_0 : PPV_1 = PPV_2$ and type I error of the test $H_0 : NPV_1 = NPV_2$ $PPV_1 = 0.95, NPV_1 = 0.95, PPV_2 = 0.75$ and $NPV_2 = 0.95$.

$Se_1 = 0.5278, Sp_1 = 0.9969, Se_2 = 0.5357, Sp_2 = 0.9802, p = 10\%, 0 \leq \rho_1 \leq 0.9841, 0 \leq \rho_2 \leq 0.3910$								
n_1	n_2	$\rho_1 = 0.25$ Power	$\rho_2 = 0.10$ Type I error	$\rho_1 = 0.49$ Power	$\rho_2 = 0.35$ Type I error	$\rho_1 = 0.74$ Power	$\rho_2 = 0.29$ Type I error	
50	50	0.001	0.030	0.001	0.024	0	0.006	
50	75	0.002	0.032	0.002	0.027	0.001	0.008	
50	100	0.005	0.031	0.003	0.027	0.002	0.008	
75	75	0.002	0.032	0.001	0.026	0.001	0.017	
100	100	0.010	0.030	0.006	0.022	0.005	0.023	
200	200	0.071	0.027	0.068	0.028	0.056	0.022	
500	500	0.654	0.025	0.678	0.028	0.693	0.025	
$Se_1 = 0.8444, Sp_1 = 0.9852, Se_2 = 0.8571, Sp_2 = 0.9048, p = 25\%, 0 \leq \rho_1 \leq 0.9511, 0 \leq \rho_2 \leq 0.3779$								
n_1	n_2	$\rho_1 = 0.24$ Power	$\rho_2 = 0.09$ Type I error	$\rho_1 = 0.48$ Power	$\rho_2 = 0.19$ Type I error	$\rho_1 = 0.71$ Power	$\rho_2 = 0.28$ Type I error	
50	50	0.120	0.025	0.118	0.012	0.097	0.001	
50	75	0.378	0.028	0.371	0.012	0.376	0.001	
50	100	0.604	0.026	0.634	0.012	0.652	0.001	
75	75	0.382	0.027	0.396	0.022	0.388	0.005	
100	100	0.622	0.031	0.644	0.026	0.679	0.012	
200	200	0.963	0.025	0.974	0.024	0.987	0.026	
500	500	1	0.028	1	0.024	1	0.024	
$Se_1 = 0.95, Sp_1 = 0.95, Se_2 = 0.9643, Sp_2 = 0.6786, p = 50\%, 0 \leq \rho_1 \leq 0.8388, 0 \leq \rho_2 \leq 0.3333$								
n_1	n_2	$\rho_1 = 0.21$ Power	$\rho_2 = 0.08$ Type I error	$\rho_1 = 0.42$ Power	$\rho_2 = 0.17$ Type I error	$\rho_1 = 0.63$ Power	$\rho_2 = 0.25$ Type I error	
50	50	0.942	0.002	0.965	0.001	0.978	0	
50	75	0.995	0.003	0.999	0	0.997	0	
50	100	1	0.002	1	0.002	1	0	
75	75	0.996	0.006	0.997	0.003	0.999	0	
100	100	1	0.010	1	0.006	1	0.002	
200	200	1	0.029	1	0.015	1	0.010	
500	500	1	0.026	1	0.024	1	0.024	

3.2. Effect of the Prevalence

The estimation and comparison of the PVs of two BDTs subject to a case-control design requires knowledge of the disease prevalence. To study the effect of a misspecification of the prevalence on the comparison of the PVs and on the estimators of the PVs, we carried out simulation experiments similar to those made to study the type I errors and the powers. For this purpose, we took as the prevalence for the inference a misspecification equal to 10% and to 20% of the value of the prevalence set, and we have studied the type I errors and the powers of the global test and of the Bonferroni and Holm methods, and the relative root mean square error (RRMSE) of the estimator of each PVs. Thus, for each estimator we calculated the relative root mean square error (RRMSE), i.e.,

$$RRMSE(\widehat{PV}_i) = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N (\widehat{PV}_{ik} - PV_i)^2}}{PV_i},$$

where PV_i is the PPV or the NPV of the i th BDT ($i = 1, 2$) and \widehat{PV}_{ik} is its estimator calculated from the k th sample ($k = 1, \dots, N$), and $N = 10,000$. For the values of the parameters we took as prevalence $p = (10\%, 25\%, 50\%)$ respectively, and to estimate the PVs we took as prevalence $p' = p \pm d \times p$ with $d = (10\%, 20\%)$. A value $d = 10\%$ ($d = 20\%$) can be considered as a small (moderate) value of the relative deviation.

Table 8 shows some of the results obtained for the type I errors and the powers of the global test and the Bonferroni method (the results of the Holms method are not shown as they are practically identically to those obtained with the Bonferroni method). In this Table we show the results when there is no misspecification of the prevalence ($p' = p$) and

when there is a misspecification of the prevalence ($p' < p$ or $p' > p$). From the results of these experiments, it is verified that the type I errors of the methods studied do not overwhelm the nominal error ($\alpha = 5\%$). In general terms there are no important differences between the type I errors when there is a misspecification of the prevalence and when there is not. Regarding the powers, the conclusions are also very similar, there are no important differences between the powers when there is a misspecification of the prevalence and when there is not.

Regarding the estimators, Table 8 shows some of the results obtained for the RRMSEs (in %) of the estimators of the PVs of Test 1 (the results for Test 2 are identical). In general terms, the difference between the RRMSEs is small (around 5% or less, in absolute value) when the two sample sizes are moderate ($n_i = 100$) or large ($n_i \geq 200$) and the relative deviation is small (10%) or moderate (20%). Therefore, a small or moderate misspecification of the prevalence ($p' < p$ or $p' > p$) does not have an important effect on the estimators of the PVs when the samples are moderate or large. Additionally, there is not an important difference between the RRMSEs when the sample sizes are small ($n_i \leq 75$) and the relative deviation is small. However, the difference between the RRMSEs is larger when the sample sizes are small and the relative deviation is moderate. In this situation, a misspecification of the prevalence has an important effect on the estimators of the PVs.

Table 8. Effect of a misspecification of the prevalence.

Type I errors											
$PPV_1 = PPV_2 = 0.90, NPV_1 = NPV_2 = 0.80$ $Se_1 = 0.2571, Sp_1 = 0.9905, Se_2 = 0.2571, Sp_2 = 0.9905, p = 25\%, \rho_1 = 0.75, \rho_2 = 0.75$											
n_1	n_2	$p' = p = 25\%$		$p' = 20\%$		$p' = 22.50\%$		$p' = 27.50\%$		$p' = 30\%$	
		Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.
50	50	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001
50	75	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
50	100	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
75	75	0.003	0.007	0.003	0.008	0.003	0.008	0.003	0.007	0.003	0.007
100	100	0.006	0.010	0.006	0.011	0.006	0.011	0.006	0.009	0.006	0.009
200	200	0.010	0.020	0.010	0.020	0.010	0.020	0.010	0.019	0.010	0.019
500	500	0.020	0.024	0.020	0.024	0.020	0.024	0.020	0.023	0.020	0.023

Powers											
$PPV_1 = 0.95, PPV_2 = 0.75, NPV_1 = 0.95, NPV_2 = 0.95$ $Se_1 = 0.8444, Sp_1 = 0.9852, Se_2 = 0.8571, Sp_2 = 0.9048, p = 25\%, \rho_1 = 0.71, \rho_2 = 0.28$											
n_1	n_2	$p' = p = 25\%$		$p' = 20\%$		$p' = 22.50\%$		$p' = 27.50\%$		$p' = 30\%$	
		Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.
50	50	0.172	0.142	0.145	0.139	0.143	0.135	0.142	0.121	0.139	0.117
50	75	0.422	0.396	0.381	0.382	0.385	0.382	0.374	0.380	0.375	0.379
50	100	0.647	0.615	0.600	0.587	0.627	0.604	0.616	0.614	0.598	0.596
75	75	0.434	0.400	0.414	0.396	0.415	0.396	0.397	0.394	0.390	0.390
100	100	0.635	0.634	0.618	0.604	0.632	0.612	0.639	0.620	0.623	0.615
200	200	0.965	0.964	0.941	0.931	0.951	0.952	0.958	0.956	0.942	0.933
500	500	1	1	1	1	1	1	1	1	1	1

RRMSEs of the estimators of PVs of Test 1											
$PPV_1 = PPV_2 = 0.90, NPV_1 = NPV_2 = 0.80$ $Se_1 = 0.2571, Sp_1 = 0.9905, Se_2 = 0.2571, Sp_2 = 0.9905, p = 25\%, \rho_1 = 0.75, \rho_2 = 0.75$											
n_1	n_2	$p' = p = 25\%$		$p' = 20\%$		$p' = 22.50\%$		$p' = 27.50\%$		$p' = 30\%$	
		\widehat{PPV}_i	\widehat{NPV}_i	\widehat{PPV}_i	\widehat{NPV}_i	\widehat{PPV}_i	\widehat{NPV}_i	\widehat{PPV}_i	\widehat{NPV}_i	\widehat{PPV}_i	\widehat{NPV}_i
50	50	27.4	1.8	36.6	5.5	31.8	2.6	29.3	3.8	34.5	6.4
50	75	20.5	1.7	27.0	5.1	23.5	2.8	21.9	3.6	25.5	6.1
50	100	16.1	1.7	21.9	5.2	18.8	2.9	17.8	3.5	21.1	6.0
75	75	20.1	1.4	26.5	5.1	23.1	2.6	21.4	3.5	24.1	6.1
100	100	15.4	1.2	19.2	4.9	18.1	2.5	16.1	3.3	18.6	5.9
200	200	8.0	0.9	11.7	4.7	9.9	2.4	8.7	3.0	10.2	5.6
500	500	4.4	0.5	7.1	4.1	5.4	2.2	4.9	2.8	5.7	5.5

Global: global hypothesis test based on the chi-square distribution. Bonf.: Bonferroni method.

4. Example

The results obtained have been applied to the diagnosis of coronary heart disease, using an electrocardiogram and an echocardiography as diagnostic tests. Both tests have been applied to a sample of 105 older men with coronary heart disease (case sample) and to another sample of 120 older men without this disease (control sample). In Table 9 we can see the frequencies obtained, where the random variable T_1 models the result of the electrocardiogram and the variable T_2 models the result of the echocardiography. In order to illustrate the method proposed, we are going to consider that the prevalence of the disease in older men is 5%. The objective is to compare the clinical accuracy (PVs) of both BDTs in the population whose prevalence of coronary heart disease is 5%. The comparison of the PVs will be made with $\alpha = 5\%$.

Table 9. Diagnosis of coronary heart disease.

		Observed Frequencies					
		Case			Control		
		$T_2 = 1$	$T_2 = 0$	Total	$T_2 = 1$	$T_2 = 0$	Total
$T_1 = 1$		77	10	87	$T_1 = 1$	4	6
$T_1 = 0$		6	12	18	$T_1 = 0$	13	114
Total		83	22	105	Total	17	120

From the case sample, the estimates of the two sensitivities (and their standard errors, SE) and of the correlation coefficient between them are $\hat{S}e_1 \pm SE = 0.829 \pm 0.037$, $\hat{S}e_2 \pm SE = 0.790 \pm 0.040$ and $\hat{\rho}_1 = 0.511$. From the control sample, the estimates of the two specificities and of the correlation coefficient between them are $\hat{S}p_1 \pm SE = 0.950 \pm 0.020$, $\hat{S}p_2 \pm SE = 0.858 \pm 0.032$ and $\hat{\rho}_2 = 0.345$. Assuming that the prevalence of coronary heart disease is 5%, the estimates of the PVs are:

$$\begin{aligned} \widehat{PPV}_1 &= \frac{0.05 \times 0.829}{0.05 \times 0.829 + (1 - 0.05) \times (1 - 0.950)} = 0.466, \\ \widehat{PPV}_2 &= \frac{0.05 \times 0.790}{0.05 \times 0.790 + (1 - 0.05) \times (1 - 0.858)} = 0.227, \\ \widehat{NPV}_1 &= \frac{(1 - 0.05) \times 0.950}{0.05 \times (1 - 0.829) + (1 - 0.05) \times 0.950} = 0.991 \end{aligned}$$

and

$$\widehat{NPV}_2 = \frac{(1 - 0.05) \times 0.858}{0.05 \times (1 - 0.790) + (1 - 0.05) \times 0.858} = 0.987.$$

Applying the delta method (Equation (3)), the estimated variance-covariance matrix of the estimators of the PVs is:

$$\hat{\Sigma}_{\hat{\omega}} = \begin{pmatrix} 0.009926 & 0.001398 & 0.000041 & 0.000029 \\ 0.001398 & 0.001632 & 0.000012 & 0.000039 \\ 0.000041 & 0.000012 & 0.000004 & 0.000002 \\ 0.000029 & 0.000039 & 0.000002 & 0.000006 \end{pmatrix}.$$

Applying Equation (5), the value of the test statistic for the global test:

$$\begin{aligned} H_0 : PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2 \\ H_1 : \text{at least one equality is not true,} \end{aligned}$$

is $Q^2 = 7.516$ and the p -value is 0.023, and therefore the null hypothesis of the global test is rejected. Testing the individual hypothesis tests it is found that the value of the test statistic for $H_0 : PPV_1 = PPV_2$ is equal to 2.552 (two sided p -value is 0.011) and that the value of the test statistic for $H_0 : NPV_1 = NPV_2$ is equal to 1.469 (two sided p -value is 0.142). Applying the Bonferroni (or Holm) method, the hypothesis of equality of the positive predictive values is rejected and the hypothesis of equality of the negative predictive values is not rejected. Therefore, in a population in which the prevalence of coronary heart disease is

5%, the positive predictive value of electrocardiogram is significantly greater than that of the echocardiography (95% confidence interval for the difference: 0.056 to 0.422), while there are no significant differences between the two negative predictive values.

5. More Than Two BDTs

Let us consider that J BDTs ($J \geq 3$) are applied to all of the individuals in the case sample and the control sample. For each BDT we define the random variable T_j in a similar way to how this was done in Section 2. Let Se_j and Sp_j be the sensitivity and the specificity of the j th BDT, with $j = 1, \dots, J$. Let $n_{1i_1 \dots i_J}$ be the number of individuals with the disease for whom $T_1 = i_1, \dots, T_J = i_J$, with $i_j = 1$ when the result of the j th BDT is positive and $i_j = 0$ when it is negative. In a similar way, $n_{2i_1 \dots i_J}$ is the number of without the disease for whom $T_1 = i_1, \dots, T_J = i_J$. Let us consider the probabilities $\xi_{hi_1 \dots i_J} = P(T_1 = i_1, T_2 = i_2, \dots, T_J = i_J)$, with $h = 1, 2$. Thus, for example for three BDTs, using the dependence model of Torrance–Rynard and Walter [16], these probabilities are:

$$\xi_{1i_1 i_2 i_3} = \prod_{j=1}^3 Se_j^{i_j} (1 - Se_j)^{1-i_j} + \sum_{j,k,j < k}^3 (-1)^{|i_j - i_k|} \varepsilon_{1jk}$$

and

$$\xi_{2i_1 i_2 i_3} = \prod_{j=1}^3 Sp_j^{1-i_j} (1 - Sp_j)^{i_j} + \sum_{j,k,j < k}^3 (-1)^{|i_j - i_k|} \varepsilon_{2jk}$$

with $i_j = 0, 1$, $i_k = 0, 1$ and $j, k = 1, 2, 3$, and where ε_{1jk} (ε_{2jk}) is the covariance between the j th BDT and the k th BDT for individuals with the disease (without the disease). The estimators of these probabilities are $\hat{\xi}_{hi_1 \dots i_J} = n_{hi_1 \dots i_J} / n_h$, with $h = 1, 2$. The sensitivity and the specificity of the j th BDT are:

$$Se_j = \frac{\sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 1}}^1 \xi_{1i_1 \dots i_J}}{\sum_{i_j = 1}^1 \xi_{1i_1 \dots i_J}} \text{ and } Sp_j = \frac{\sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 0}}^1 \xi_{2i_1 \dots i_J}}{\sum_{i_j = 0}^1 \xi_{2i_1 \dots i_J}}$$

and its estimators are:

$$\hat{Se}_j = \frac{\sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 1}}^1 n_{1i_1 \dots i_J}}{n_1} \text{ and } \hat{Sp}_j = \frac{\sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 0}}^1 n_{2i_1 \dots i_J}}{n_2}$$

The estimators of the variances-covariances of these estimators are $\hat{V}ar(\hat{Se}_j) = \hat{Se}_j(1 - \hat{Se}_j) / n_1$, $\hat{V}ar(\hat{Sp}_j) = \hat{Sp}_j(1 - \hat{Sp}_j) / n_2$, $\hat{C}ov(\hat{Se}_j, \hat{Se}_k) = \varepsilon_{1jk} / n_1$ and $\hat{C}ov(\hat{Sp}_j, \hat{Sp}_k) = \varepsilon_{2jk} / n_2$, and the rest of the covariances are equal to zero. The estimators of the PVs of the j th BDT are:

$$\widehat{PPV}_j = \frac{pn_2 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 1}}^1 n_{1i_1 \dots i_J}}{pn_2 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 1}}^1 n_{1i_1 \dots i_J} + qn_1 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 1}}^1 n_{2i_1 \dots i_J}}$$

and

$$\widehat{NPV}_j = \frac{qn_1 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 0}}^1 n_{2i_1, \dots, i_J}}{qn_1 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 0}}^1 n_{2i_1, \dots, i_J} + pn_2 \sum_{\substack{i_1, \dots, i_J = 0 \\ i_j = 0}}^1 n_{1i_1, \dots, i_J}},$$

where p is the disease prevalence and $q = 1 - p$.

Let $\theta = (Se_1, \dots, Se_J, Sp_1, \dots, Sp_J)^T$ be the vector whose components are the sensitivities and the specificities, and let $\omega = (PPV_1, \dots, PPV_J, NPV_1, \dots, NPV_J)^T$ be the vector whose components are the PVs. The variance-covariance matrix of the vector $\hat{\theta}$, with a dimension $2J \times 2J$, is similar to that given in Equation (2), where Σ_{δ_e} and Σ_{δ_p} are matrixes with a dimension $J \times J$. Applying the delta method, the variance-covariance matrix of $\hat{\omega}$, with a dimension $2J \times 2J$, has an expression similar to that given in Equation (3). The PVs of each one of the J BDTs depend on the same parameters (the sensitivity and the specificity of the j th diagnostic test) and, therefore, these parameters can be compared simultaneously. The global hypothesis test to simultaneously compare the PVs of the J BDTs is:

$$H_0 : PPV_1 = PPV_2 = \dots = PPV_J \text{ and } NPV_1 = NPV_2 = \dots = NPV_J$$

$$H_1 : \text{at least one equality is not true,}$$

which is equivalent to the hypothesis test:

$$H_0 : \mathbf{A}\omega = 0 \text{ vs } H_1 : \mathbf{A}\omega \neq 0,$$

where \mathbf{A} is a matrix with a dimension $2(J - 1) \times 2J$, i.e.,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_0 \\ \mathbf{A}_0 & \mathbf{A}_1 \end{pmatrix}.$$

\mathbf{A}_0 is a matrix with a dimension $(J - 1) \times J$ whose elements are all equal to 0, and \mathbf{A}_1 is a matrix with a dimension $(J - 1) \times J$ where each component (i, i) is equal to 1, each element $(i, i + 1)$ is equal to -1 for $i = 1, \dots, J - 1$, and the rest of the elements in this matrix are equal to 0. Applying the multivariate central limit theorem it is verified that $\sqrt{n_1 + n_2}(\hat{\omega} - \omega) \xrightarrow{n_1 + n_2 \rightarrow \infty} N_{2J}(0, \Sigma_\omega)$. Then, the statistic $Q^2 = (\hat{\omega}\mathbf{A})^T (\mathbf{A}\hat{\Sigma}_\omega \mathbf{A}^T)^{-1} \mathbf{A}\hat{\omega}$ is distributed according to Hotelling's T-squared distribution with a dimension $2(J - 1)$ and $n_1 + n_2$ degrees of freedom, where $2(J - 1)$ is the dimension of the vector $\mathbf{A}\hat{\omega}$. When $n_1 + n_2$ is large, the statistic Q^2 is distributed according to a central chi-squared distribution with $2(J - 1)$ degrees of freedom when the null hypothesis is true, i.e.,

$$Q^2 = \left(\hat{\omega}\mathbf{A}\right)^T \left(\mathbf{A}\hat{\Sigma}_\omega \mathbf{A}^T\right)^{-1} \mathbf{A}\hat{\omega} \xrightarrow{n \rightarrow \infty} \chi_{2(J-1)}^2.$$

Finally, the method to compare the PVs of the J BDTs would consist of the following steps: (1) Solve the global hypothesis test to an α error calculating the statistic $Q^2 = (\hat{\omega}\mathbf{A})^T (\mathbf{A}\hat{\Sigma}_\omega \mathbf{A}^T)^{-1} \mathbf{A}\hat{\omega}$ based on the chi-squared distribution; (2) if the global test is not significant to an α error then we do not reject the homogeneity of the J PVs, but if the hypothesis test is significant then the causes of significance are investigated comparing the $PPVs$ ($NPVs$) in pairs (Equation (6)) and applying a multiple comparison method (e.g., Bonferroni or Holm).

6. Discussion

The comparison of the positive and negative predictive values of two binary diagnostic tests is an important topic in the study of statistical methods in diagnostic medicine. Subject to a paired design, this topic has been subject to different studies. In this article we studied the simultaneous comparison of the predictive values of two diagnostic tests subject to a case-control design, analyzing and comparing several methods. These methods consisted of a global test based on the chi-square distribution, a method based on the individual comparisons each one to an α error, and other two methods based on individual comparisons along with a multiple comparison method. The multiple comparison methods that were used were the Bonferroni method and the Holm method, which are methods based on the p -values of the individual hypothesis tests and are very easy to apply. The methods studied to compare the predictive values require knowing the prevalence of the disease. The prevalence can be known from other studies, such as population studies of health services, cohort studies, etc. If the researcher has a great uncertainty about the value of the prevalence, the problem can be solved by using several values for the prevalence and then analyzing and comparing the results obtained.

Simulation experiments were carried out to study the type I errors and the powers of the four methods proposed. These experiments were based on the generation of random samples with type I bivariate binomial distributions, which are the distributions that are inherent to case-control design, since proportions of marginal totals are estimated from these samples. The results have shown that the global hypothesis test based on the chi-square distribution behaves well in terms of type I error, and does not overwhelm the nominal error. Regarding its power, in general this strongly depends on the disease prevalence, and it is necessary to have very large samples when the prevalence is small and relatively small sample sizes when the prevalence is high, so that the power will be high.

Based on the results of the simulation experiments, a method has been proposed to compare the predictive values of two diagnostic tests subject to a case-control design. This method, which is similar to that proposed by Roldán-Nofuentes et al. [8], consists of the following steps: (1) Simultaneously comparing the predictive values applying the global hypothesis test based on the chi-square distribution to an α error; (2) if the global hypothesis test is not significant, then the equality hypothesis of the PVs is not rejected. If the global hypothesis test is significant to an α error, then the causes of the significance are studied solving the individual hypothesis tests and applying the Bonferroni method or the Holm method to an α error. This procedure that we propose is similar to the analysis of variance: firstly, the global test is solved and, if this is significant, then the causes of the significance are studied starting with paired comparisons along with some multiple comparison method.

Simulation experiments were carried out to study the effect of a misspecification of the prevalence in the asymptotic behavior of the global hypothesis test based on the chi-square distribution and on the methods based on multiple comparisons. In general terms, we can conclude that a small or moderate misspecification of the prevalence do not have an important effect on the behavior of these hypothesis tests, especially when the sample sizes are moderate or large.

The global hypothesis test was extended to the situation in which we simultaneously compare the PVs of more than two BDTs, and for this we propose a method which is similar to that proposed for two BDTs. To be able to calculate the global test statistic $Q^2 = \left(\hat{\omega}A\right)^T \left(A\hat{\Sigma}_{\omega}A^T\right)^{-1} A\hat{\omega}$ it is necessary that $A\hat{\Sigma}_{\omega}A^T$ to be non-singular. For two BDTs, this matrix is non-singular when it is verified that $n_{i10} + n_{i01} > 0$, with $i = 1, 2$. If $n_{i10} = n_{i01} = 0$ then the method proposed to compare the PVs cannot be applied. A solution to this problem consists in adding the value 0.5 to all the observed frequencies (the sample size increases by two units), a very frequent solution in the analysis of 2×2 tables. Simulation experiments have been carried out to study the type I errors and the powers of the hypothesis tests proposed in Section 2, using this solution when in a sample

it is verified that $n_{i10} = n_{i01} = 0$. These experiments have been designed in a similar way to those performed in Section 3. Table 10 shows some results (type I errors and powers) for some of the scenarios considered in Section 3, as well as the average proportions (of the three correlation scenarios) of case (control) samples in which the value 0.5 has been added. Obviously, the proportion of samples in which the value 0.5 has been added is greater when the sample size is small. In general terms, the conclusions are the same as those obtained in the simulation experiments presented in Section 3, although the powers of all hypothesis tests are slightly lower when the sample sizes are small. Therefore, adding 0.5 to all the observed frequencies of a sample in which $n_{i10} = n_{i01} = 0$ is an adequate solution to be able to apply the PVs comparison method.

Table 10. Type I errors and powers when 0.5 is added to the samples in which $n_{i10} = n_{i01} = 0$.

Type I Errors												
$PPV_1 = PPV_2 = 0.70, NPV_1 = NPV_2 = 0.95$ $Se_1 = 0.5385, Sp_1 = 0.9744, Se_2 = 0.5385, Sp_2 = 0.9744, p = 10\%$ $\rho_1 = 0.25\rho_2 = 0.25$ $\rho_1 = 0.50\rho_2 = 0.50$ $\rho_1 = 0.75\rho_2 = 0.75$												
n_1	n_2	\bar{P}_1	\bar{P}_2	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	3.3	78.1	0.028	0.065	0.030	0.023	0.056	0.025	0.012	0.040	0.015
50	75	3.1	63.7	0.038	0.075	0.039	0.026	0.060	0.026	0.012	0.041	0.014
50	100	2.7	51.9	0.047	0.096	0.046	0.035	0.074	0.037	0.011	0.041	0.015
75	75	0.4	64.3	0.037	0.074	0.037	0.026	0.062	0.027	0.015	0.045	0.019
100	100	0.1	51.8	0.040	0.085	0.038	0.032	0.090	0.034	0.021	0.048	0.023
200	200	0	22.7	0.061	0.104	0.060	0.045	0.088	0.042	0.024	0.055	0.025
500	500	0	2.8	0.052	0.099	0.0524	0.056	0.101	0.048	0.044	0.094	0.045

Powers												
$PPV_1 = 0.75, PPV_2 = 0.60, NPV_1 = 0.95, NPV_2 = 0.95$ $Se_1 = 0.8571, Sp_1 = 0.9048, Se_2 = 0.8727, Sp_2 = 0.8061, p = 25\%, 0 \leq \rho_1 \leq 0.9354, 0 \leq \rho_2 \leq 0.6614$ $\rho_1 = 0.25\rho_2 = 0.25$ $\rho_1 = 0.50\rho_2 = 0.50$ $\rho_1 = 0.75\rho_2 = 0.75$												
n_1	n_2	\bar{P}_1	\bar{P}_2	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.	Global	$\alpha = 5\%$	Bonf.
50	50	14.8	18.1	0.297	0.369	0.269	0.342	0.427	0.306	0.369	0.484	0.340
50	75	14.5	9.6	0.417	0.501	0.379	0.492	0.576	0.449	0.604	0.688	0.568
50	100	14.2	5.5	0.515	0.590	0.476	0.629	0.693	0.576	0.739	0.798	0.704
75	75	5.8	9.7	0.424	0.506	0.392	0.502	0.584	0.456	0.622	0.705	0.579
100	100	2.5	5.5	0.525	0.603	0.486	0.624	0.693	0.583	0.761	0.819	0.728
200	200	0.1	0.7	0.817	0.863	0.786	0.910	0.930	0.876	0.975	0.983	0.965
500	500	0	0	0.996	0.999	0.997	1	1	1	1	1	1

\bar{P}_1 : average proportion (in %) of case samples in which 0.5 has been added. \bar{P}_2 : average proportion (in %) of control samples in which 0.5 has been added. Global: global hypothesis test based on the chi-square distribution. $\alpha = 5\%$: individual hypothesis tests, each one to an error of 5%. Bonf.: Bonferroni method.

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

Funding: This research was supported by the Spanish Ministry of Economy, Grant Number MTM2016-76938-P, and by the University of Nouakchott Alaasriya.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the three referees for their helpful comments that improved the quality of this manuscript. We thank Professor Cheikh Saad Bouh Camara, the president of the University of Nouakchott, for his help to finance the publication of this manuscript.

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

Appendix A

Performing algebraic operations in Equation (3) it is found that:

$$\begin{aligned} \text{Var}(\widehat{PPV}_1) &= \left(\frac{p^2Se_1 + pQ_1}{Q_1^2}\right)^2 \text{Var}(\hat{Se}_1) + \left(\frac{pqSe_1}{Q_1^2}\right)^2 \text{Var}(\hat{Sp}_1), \\ \text{Var}(\widehat{PPV}_2) &= \left(\frac{p^2Se_2 + pQ_2}{Q_2^2}\right)^2 \text{Var}(\hat{Se}_2) + \left(\frac{pqSe_2}{Q_2^2}\right)^2 \text{Var}(\hat{Sp}_2), \\ \text{Var}(\widehat{NPV}_1) &= \left(\frac{pqSp_1}{(1-Q_1)^2}\right)^2 \text{Var}(\hat{Se}_1) + \left(\frac{q(1-Q_1) - q^2Sp_1}{(1-Q_1)^2}\right)^2 \text{Var}(\hat{Sp}_1), \\ \text{Var}(\widehat{NPV}_2) &= \left(\frac{pqSp_2}{(1-Q_2)^2}\right)^2 \text{Var}(\hat{Se}_2) + \left(\frac{q(1-Q_2) - q^2Sp_2}{(1-Q_2)^2}\right)^2 \text{Var}(\hat{Sp}_2), \\ \text{Cov}(\widehat{PPV}_1, \widehat{PPV}_2) &= \\ &\left(\frac{pQ_1 - p^2Se_1}{Q_1^2}\right) \left(\frac{pQ_2 - p^2Se_2}{Q_2^2}\right) \text{Cov}(\hat{Se}_1, \hat{Se}_2) + \frac{p^2q^2Se_1Se_2}{Q_1^2Q_2^2} \text{Cov}(\hat{Sp}_1, \hat{Sp}_2), \\ \text{Cov}(\widehat{PPV}_1, \widehat{NPV}_1) &= \\ \frac{pq}{Q_1^2(1-Q_1)^2} [(pQ_1 - p^2Se_1)Sp_1 \text{Var}(\hat{Se}_1) + \{q(1-Q_1) - q^2Sp_1\}Se_1 \text{Var}(\hat{Sp}_1)], \\ \text{Cov}(\widehat{PPV}_1, \widehat{NPV}_2) &= \\ \frac{pq(pQ_1 - p^2Se_1)}{Q_1^2(1-Q_2)^2} Sp_2 \text{Cov}(\hat{Se}_1, \hat{Se}_2) + \frac{pq\{q(1-Q_2) - q^2Sp_2\}}{Q_1^2(1-Q_2)^2} Se_1 \text{Cov}(\hat{Sp}_1, \hat{Sp}_2), \\ \text{Cov}(\widehat{PPV}_2, \widehat{NPV}_1) &= \\ \frac{pq}{Q_2^2(1-Q_1)^2} [(pQ_2 - p^2Se_2)Sp_1 \text{Cov}(\hat{Se}_1, \hat{Se}_2) + \{q(1-Q_1) - q^2Sp_1\}Se_2 \text{Cov}(\hat{Sp}_1, \hat{Sp}_2)], \\ \text{Cov}(\widehat{PPV}_2, \widehat{NPV}_2) &= \\ \frac{pq}{Q_2^2(1-Q_2)^2} [(pQ_2 - p^2Se_2)Sp_2 \text{Var}(\hat{Se}_2) + \{q(1-Q_2) - q^2Sp_2\}Se_2 \text{Var}(\hat{Sp}_2)] \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(\widehat{NPV}_1, \widehat{NPV}_2) &= \\ \frac{p^2q^2Sp_1Sp_2}{(1-Q_1)^2(1-Q_2)^2} \text{Cov}(\hat{Se}_1, \hat{Se}_2) + \left(\frac{q(1-Q_1) - q^2Sp_1}{(1-Q_1)^2}\right) \left(\frac{q(1-Q_2) - q^2Sp_2}{(1-Q_2)^2}\right) \text{Cov}(\hat{Sp}_1, \hat{Sp}_2), \end{aligned}$$

and where $q = 1 - p$ and $Q_i = p \times Se_i + q \times (1 - Sp_i)$.

Appendix B

Let us assume that we are going to solve K hypothesis test H_{0k} vs. H_{1k} with $k = 1, \dots, K$. Let $p_{[1]} \leq p_{[2]} \leq \dots \leq p_{[K]}$ be the p -values obtained ordered from the lowest to the highest, and therefore $p_{[k]}$ is the p -value that corresponds to the hypothesis test $H_{0[k]}$ vs. $H_{1[k]}$. The Holm method [13] consists of the following steps:

Step 1. If $p_{[1]} \leq \alpha/K$ then hypothesis $H_{0[1]}$ is rejected and we go to the next step; if $p_{[1]} > \alpha/K$ then no null hypothesis is rejected and the process finishes.

Step 2. If $p_{[2]} \leq \alpha/(K - 1)$ then $H_{0[2]}$ is rejected and we go to the next step; if $p_{[2]} > \alpha/(K - 1)$ we do not reject the null hypotheses $H_{0[k]}$ with $k = 2, \dots, K$ and the process finishes

Step K. If $p_{[K]} \leq \alpha$ then $H_{0[K]}$ is rejected and the process finishes; and if $p_{[K]} > \alpha$ then $H_{0[K]}$ is not rejected and the process finishes.

References

1. Pepe, M.S. *The Statistical Evaluation of Medical Tests for Classification and Prediction*; Oxford University Press: New York, NY, USA, 2003.
2. Bennett, B.M. On comparison of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics* **1972**, *28*, 793–800. [[CrossRef](#)] [[PubMed](#)]
3. Bennett, B.M. On tests for equality of predictive values for t diagnostic procedures. *Stat. Med.* **1985**, *4*, 535–539. [[CrossRef](#)] [[PubMed](#)]
4. 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](#)]
5. 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](#)]
6. Kosinski, A.S. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. *Stat. Med.* **2013**, *32*, 964–977. [[CrossRef](#)] [[PubMed](#)]
7. 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](#)]
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. Mercaldo, N.D.; Kit, F.L.; Zhou, X.H. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat. Med.* **2007**, *26*, 2170–2183. [[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. Kocherlakota, S.; Kocherlakota, K. *Bivariate Discrete Distributions*; Marcel Dekker INC: New York, NY, USA, 1992.
12. Bonferroni, C.E. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **1936**, *8*, 3–62.
13. Holm, S. A simple sequential rejective multiple testing procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
14. RC Team. *R: A Language and Environment for Statistical Computing*; RC Team: Vienna, Austria, 2016; Available online: <https://www.R-project.org/> (accessed on 17 March 2021).
15. Leisch, F.; Weingessel, A.; Hornik, K. Bindata Package. Available online: <https://cran.r-project.org/web/packages/bindata/> (accessed on 17 March 2021).
16. Torrance-Rynard, V.L.; Walter, S.D. Effects of dependent errors in the assessment of diagnostic test performance. *Stat. Med.* **1997**, *16*, 2157–2175. [[CrossRef](#)]