Comparison of the predictive values of two binary diagnostic tests subject to a case-control sampling

Jose Antonio Roldán-Nofuentes (✉ jaroldan@ugr.es)

University of Granada <https://orcid.org/0000-0003-0251-5588>

Saad Bouh Sidaty-Regad

Universite de Nouakchott

Technical advance

Keywords: Binary diagnostic test, Chi-square distribution, Positive and negative predictive values.

Posted Date: August 30th, 2019

DOI: <https://doi.org/10.21203/rs.2.12286/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Comparison of the predictive values of two binary diagnostic tests subject to a case-control sampling

José Antonio Roldán-Nofuentes¹, Saad Bouh Sidaty-Regad²

¹Biostatistics, School of Medicine, University of Granada, Spain. Email: jaroldan@ugr.es

²Public Health and Epidemiology, School of Medicine, University of Nouakchott, Mauritania.

Email: sidaty_saad@yahoo.com

Abstract

Background: The main parameters to compare the accuracy of binary diagnostic tests are the sensitivity and the specificity. Case-control sampling 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. In this design, the sensitivities are compared from the case sample applying the McNemar's test, and the specificities from the control sample. Other parameters of binary diagnostic tests are the positive and negative predictive values. The predictive values of a diagnostic test represent the clinical accuracy of a binary diagnostic test when it is applied to the individuals in a population with a determined disease prevalence.

Methods: This article studies the comparison of the predictive values of two diagnostic tests subject to a case-control sampling. A global hypothesis test, based on the chi-square distribution, is proposed to compare the predictive values simultaneously. The comparison of the predictive values is also studied individually. The hypothesis tests studied require knowledge of an estimation of the disease prevalence.

Results: 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 behaviour of the hypothesis tests and on the estimators of the predictive values. The results obtained were applied to a real example on the diagnosis of the Human African Trypanosomiasis.

Conclusions: A method has been proposed to compare the predictive values of two diagnostic tests subject to a case-control design. This method consists of the following steps:

1) Simultaneously comparing the predictive values applying the global hypothesis test based on the chi-square distribution to an error α ; 2) If the global hypothesis test is not significant, then the equality hypothesis of the predictive values is not rejected. If the global hypothesis test is significant to an error α , then the causes of the significance are studied solving the individual hypothesis tests and applying Bonferroni's method (or Holm's method) to an error α .

Keywords: Binary diagnostic test, Chi-square distribution, Positive and negative predictive values.

Background

A diagnostic test is a medical test that is applied to an individual to determine the presence or absence of a disease. When the result of a diagnostic test is positive or negative, the diagnostic test is called a binary diagnostic test (BDT). The PSA for the diagnosis of prostate cancer is an example of BDT. The accuracy of a BDT is assessed in relation to a gold standard (GS), which is a medical test used to objectively diagnose the presence or absence of the disease. From this point on, it is assumed that the GS is an error-free test, and therefore it objectively determines if an individual does or does not have the disease. A biopsy for prostate is an example of GS. 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 the Se and the Sp of the BDT and on the disease prevalence (p), and are easily calculated applying Bayes' Theorem i.e.

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

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

74 clinician are more interested in knowing how probable it is to have the disease given a BDT
75 result.

76 The comparison of the performance of two binary diagnostic tests is a topic of special
77 importance in the study of statistical methods for the diagnosis of diseases. This comparison
78 is made through a cross-sectional sampling or through a case-control sampling. Cross-
79 sectional sampling consists of applying the two BDTs and the GS to all of the individuals in a
80 single sample. Case-control design consists of applying the two BDTs to all of the individuals
81 in two samples, one made up of individuals who have the disease (case sample) and another
82 made up of individuals who do not have the disease (control sample). The advantages and
83 disadvantages of case-control design over the cross-sectional can be seen in the book by Pepe
84 [1]. Summarizing, case-control design has some advantages over paired design: a) case-
85 control design is more efficient in terms of sample size requirements, b) case-control studies
86 allow for the exploration of subject-related characteristics of the test. Nevertheless, the case-
87 control design has the disadvantage is that by using it we cannot estimate the prevalence of
88 the disease.

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

96 In a case-control sampling, Mercaldo et al [9] have studied the estimation of the PVs of a
97 BDT, assuming that an estimation of the prevalence of the disease is known. In this article,
98 we extended the study of Mercaldo et al [9] to the case of two BDTs and we adapt the method

of Roldán-Nofuentes et al [8], studying the comparison of the PVs of the two BDTs subject to a case-control sampling. Subject to a case-control sampling, 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_0 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 have an estimation of the disease prevalence. This estimation can be obtained from health surveys or from previous studies. Consequently, the methods of comparison of the PVs subject to a cross-sectional sampling cannot be applied when there is a case-control design. In the next section, we study hypothesis tests to jointly and individually compare the PVs of two BDTs subject to case-control sampling. Then, simulation experiments are carried out to study the type I errors and the powers of the hypothesis tests proposed in the previous section, and we study the effect of the misspecification of the prevalence on the asymptotic behaviour of these hypothesis tests and on the estimators of the PVs. The results are applied to a real example of the diagnosis of Human African Trypanosomiasis, and finally the results are discussed.

Methods

Let us consider two BDTs 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. Using the conditional dependence model of Vacek [10], the probabilities given in the table are written as

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

and

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

with $j, k = 0, 1$. The parameter ε^+ (ε^-) 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 $0 \leq \varepsilon^+ \leq \text{Min}\{Se_1(1 - Se_2), Se_2(1 - Se_1)\}$ and $0 \leq \varepsilon^- \leq \text{Min}\{Sp_1(1 - Sp_2), Sp_2(1 - Sp_1)\}$. If $\varepsilon^+ = \varepsilon^- = 0$ then the two BDTs are conditionally independent on the disease status. In practice, the assumption of the conditional independence is not realistic, and therefore $\varepsilon^+ > 0$ and/or $\varepsilon^- > 0$.

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

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

139 In terms of the probabilities ξ_{ijk} , the sensitivities are written as

$$140 \quad Se_1 = \xi_{111} + \xi_{110} \quad \text{and} \quad Se_2 = \xi_{111} + \xi_{101},$$

141 and the specificities as

$$142 \quad Sp_1 = \xi_{201} + \xi_{200} \quad \text{and} \quad Sp_2 = \xi_{210} + \xi_{200}.$$

143 From the case (control) samples Se_1 and Se_2 (Sp_1 and Sp_2) are estimated i.e.

$$144 \quad \hat{Se}_1 = \frac{n_{11\cdot}}{n_1} \quad \text{and} \quad \hat{Se}_2 = \frac{n_{1\cdot 1}}{n_1},$$

145 and

$$146 \quad \hat{Sp}_1 = \frac{n_{20\cdot}}{n_2} \quad \text{and} \quad \hat{Sp}_2 = \frac{n_{2\cdot 0}}{n_2},$$

147 and the estimators of their variances are $\hat{Var}(\hat{Se}_1) = \hat{Se}_1(1 - \hat{Se}_1)/n_1$,

148 $\hat{Var}(\hat{Se}_2) = \hat{Se}_2(1 - \hat{Se}_2)/n_1$, $\hat{Var}(\hat{Sp}_1) = \hat{Sp}_1(1 - \hat{Sp}_1)/n_2$ and $\hat{Var}(\hat{Sp}_2) = \hat{Sp}_2(1 - \hat{Sp}_2)/n_2$.

149 Therefore, the sensitivities and the specificities are estimated as proportions of marginal

150 totals. In this way, in the case sample we are interested in the marginal frequencies $n_{11\cdot}$ and

151 $n_{1\cdot 1}$, and therefore these frequencies are the product of a type I bivariate binomial distribution

152 [11]. In an analogous way, from the control sample, the marginal frequencies $n_{20\cdot}$ and $n_{2\cdot 0}$

153 are the product of a type I bivariate binomial distribution. In the individuals with the disease,

154 the type I bivariate binomial distribution is characterized [11] by the two probabilities Se_1 and

155 Se_2 and by the correlation coefficient (ρ^+) between T_1 and T_2 . In an analogous way, in the

156 individuals who do not have the disease, the type I bivariate binomial distribution is

157 characterized by Sp_1 , Sp_2 and the correlation coefficient (ρ^-) between T_1 and T_2 . In the

158 individuals with the disease (cases), the correlation coefficient between the two BDTs is

$$\rho^+ = \frac{\xi_{111} - Se_1 Se_2}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}} = \frac{\varepsilon^+}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}},$$

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

$$\rho^- = \frac{\xi_{200} - Sp_1 Sp_2}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}} = \frac{\varepsilon^-}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}}.$$

It is easy to check that

$$\hat{\varepsilon}^+ = \frac{n_1 n_{111} - n_{11\cdot} n_{1\cdot 1}}{n_1^2} \quad \text{and} \quad \hat{\varepsilon}^- = \frac{n_2 n_{200} - n_{20\cdot} n_{2\cdot 0}}{n_2^2},$$

$$\hat{Cov}(\hat{Se}_1, \hat{Se}_2) = (\hat{\xi}_{111} - \hat{Se}_1 \hat{Se}_2) / n_1 = \hat{\varepsilon}^+ / n_1$$

and

$$\hat{Cov}(\hat{Sp}_1, \hat{Sp}_2) = (\hat{\xi}_{200} - \hat{Sp}_1 \hat{Sp}_2) / n_2 = \hat{\varepsilon}^- / n_2.$$

All of the other covariances are zero, since the two samples are independent. The estimators of ρ^+ and ρ^- are

$$\hat{\rho}^+ = \frac{n_1 n_{111} - n_{11\cdot} n_{1\cdot 1}}{\sqrt{n_{11\cdot} (n_1 - n_{11\cdot}) n_{1\cdot 1} (n_1 - n_{1\cdot 1})}} \quad \text{and} \quad \hat{\rho}^- = \frac{n_2 n_{200} - n_{20\cdot} n_{2\cdot 0}}{\sqrt{n_{20\cdot} (n_2 - n_{20\cdot}) n_{2\cdot 0} (n_2 - n_{2\cdot 0})}}.$$

Assuming that an estimation p of the disease prevalence is known, the estimators of the predictive values are

$$\widehat{PPV}_1 = \frac{pn_2 n_{11\cdot}}{pn_2 n_{11\cdot} + qn_1 (n_2 - n_{20\cdot})} \quad \text{and} \quad \widehat{NPV}_1 = \frac{qn_1 n_{20\cdot}}{pn_2 (n_1 - n_{11\cdot}) + qn_1 n_{20\cdot}}$$

for *BDT* 1, and

$$\widehat{PPV}_2 = \frac{pn_2 n_{1\cdot 1}}{pn_2 n_{1\cdot 1} + qn_1 (n_2 - n_{2\cdot 0})} \quad \text{and} \quad \widehat{NPV}_2 = \frac{qn_1 n_{2\cdot 0}}{pn_2 (n_1 - n_{1\cdot 1}) + qn_1 n_{2\cdot 0}}$$

176 for *BDT 2*, where $q = 1 - p$. Let $\boldsymbol{\theta} = (Se_1, Se_2, Sp_1, Sp_2)^T$ be a vector whose components are the
 177 sensitivities and the specificities, and let $\boldsymbol{\omega} = (PPV_1, PPV_2, NPV_1, NPV_2)^T$ be a vector whose
 178 components are the predictive values. The estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$179 \quad \hat{\Sigma}_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \hat{Var}(\hat{Se}_1) & \hat{Cov}(\hat{Se}_1, \hat{Se}_2) & 0 & 0 \\ \hat{Cov}(\hat{Se}_1, \hat{Se}_2) & \hat{Var}(\hat{Se}_2) & 0 & 0 \\ 0 & 0 & \hat{Var}(\hat{Sp}_1) & \hat{Cov}(\hat{Sp}_1, \hat{Sp}_2) \\ 0 & 0 & \hat{Cov}(\hat{Sp}_1, \hat{Sp}_2) & \hat{Var}(\hat{Sp}_2) \end{pmatrix},$$

180 Applying the delta method, the estimated matrix of variances- covariances of $\hat{\boldsymbol{\omega}}$ is

$$181 \quad \hat{\Sigma}_{\hat{\boldsymbol{\omega}}} = \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \hat{\Sigma}_{\hat{\boldsymbol{\theta}}} \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}^T. \quad (2)$$

182 Performing algebraic operations in equation (2), the elements of the estimated variance-
 183 covariance matrix of the estimators of predictive values are

$$184 \quad \hat{Var}(\widehat{PPV_1}) = \left(\frac{p^2 \hat{Se}_1 + p \hat{Q}_1}{\hat{Q}_1^2} \right)^2 \hat{Var}(\hat{Se}_1) + \left(\frac{pq \hat{Se}_1}{\hat{Q}_1^2} \right)^2 \hat{Var}(\hat{Sp}_1),$$

$$185 \quad \hat{Var}(\widehat{PPV_2}) = \left(\frac{p^2 \hat{Se}_2 + p \hat{Q}_2}{\hat{Q}_2^2} \right)^2 \hat{Var}(\hat{Se}_2) + \left(\frac{pq \hat{Se}_2}{\hat{Q}_2^2} \right)^2 \hat{Var}(\hat{Sp}_2),$$

$$186 \quad \hat{Var}(\widehat{NPV_1}) = \left(\frac{pq \hat{Sp}_1}{(1 - \hat{Q}_1)^2} \right)^2 \hat{Var}(\hat{Se}_1) + \left(\frac{q(1 - \hat{Q}_1) - q^2 \hat{Sp}_1}{(1 - \hat{Q}_1)^2} \right)^2 \hat{Var}(\hat{Sp}_1),$$

$$187 \quad \hat{Var}(\widehat{NPV_2}) = \left(\frac{pq \hat{Sp}_2}{(1 - \hat{Q}_2)^2} \right)^2 \hat{Var}(\hat{Se}_2) + \left(\frac{q(1 - \hat{Q}_2) - q^2 \hat{Sp}_2}{(1 - \hat{Q}_2)^2} \right)^2 \hat{Var}(\hat{Sp}_2),$$

$$188 \quad \begin{aligned} & \hat{Cov}(\widehat{PPV_1}, \widehat{PPV_2}) = \\ & \left(\frac{p \hat{Q}_1 - p^2 \hat{Se}_1}{\hat{Q}_1^2} \right) \left(\frac{p \hat{Q}_2 - p^2 \hat{Se}_2}{\hat{Q}_2^2} \right) \hat{Cov}(\hat{Se}_1, \hat{Se}_2) + \frac{p^2 q^2 \hat{Se}_1 \hat{Se}_2}{\hat{Q}_1^2 \hat{Q}_2^2} \hat{Cov}(\hat{Sp}_1, \hat{Sp}_2), \end{aligned}$$

$$\begin{aligned}
& \hat{Cov}(\widehat{PPV_1}, \widehat{NPV_1}) = \\
189 \quad & \frac{pq}{\hat{Q}_1^2(1-\hat{Q}_1)^2} \left[(p\hat{Q}_1 - p^2\hat{S}e_1) \hat{S}p_1 \hat{V}ar(\hat{S}e_1) + \{q(1-\hat{Q}_1) - q^2\hat{S}p_1\} \hat{S}e_1 \hat{V}ar(\hat{S}p_1) \right], \\
& \hat{Cov}(\widehat{PPV_1}, \widehat{NPV_2}) = \\
190 \quad & \frac{pq(p\hat{Q}_1 - p^2\hat{S}e_1)}{\hat{Q}_1^2(1-\hat{Q}_2)^2} \hat{S}p_2 \hat{Cov}(\hat{S}e_1, \hat{S}e_2) + \frac{pq\{q(1-\hat{Q}_2) - q^2\hat{S}p_2\}}{\hat{Q}_1^2(1-\hat{Q}_2)^2} \hat{S}e_1 \hat{Cov}(\hat{S}p_1, \hat{S}p_2), \\
& \hat{Cov}(\widehat{PPV_2}, \widehat{NPV_1}) = \\
191 \quad & \frac{pq}{\hat{Q}_2^2(1-\hat{Q}_1)^2} \left[(p\hat{Q}_2 - p^2\hat{S}e_2) \hat{S}p_1 \hat{Cov}(\hat{S}e_1, \hat{S}e_2) + \{q(1-\hat{Q}_1) - q^2\hat{S}p_1\} \hat{S}e_2 \hat{Cov}(\hat{S}p_1, \hat{S}p_2) \right], \\
& \hat{Cov}(\widehat{PPV_2}, \widehat{NPV_2}) = \\
192 \quad & \frac{pq}{\hat{Q}_2^2(1-\hat{Q}_2)^2} \left[(p\hat{Q}_2 - p^2\hat{S}e_2) \hat{S}p_2 \hat{V}ar(\hat{S}e_2) + \{q(1-\hat{Q}_2) - q^2\hat{S}p_2\} \hat{S}e_2 \hat{V}ar(\hat{S}p_2) \right]
\end{aligned}$$

193 and

$$\begin{aligned}
& \hat{Cov}(\widehat{NPV_1}, \widehat{NPV_2}) = \\
194 \quad & \frac{p^2q^2\hat{S}p_1\hat{S}p_2}{(1-\hat{Q}_1)^2(1-\hat{Q}_2)^2} \hat{Cov}(\hat{S}e_1, \hat{S}e_2) + \left(\frac{q(1-\hat{Q}_1) - q^2\hat{S}p_1}{(1-\hat{Q}_1)^2} \right) \left(\frac{q(1-\hat{Q}_2) - q^2\hat{S}p_2}{(1-\hat{Q}_2)^2} \right) \hat{Cov}(\hat{S}p_1, \hat{S}p_2),
\end{aligned}$$

195 where $q = 1 - p$ and $\hat{Q}_i = p \times \hat{S}e_i + q \times (1 - \hat{S}p_i)$ with $i = 1, 2$.

196 Then, we study the joint comparison and the individual comparison of the PVs of the two
197 BDTs. In both cases, and as has been explained above, it is assumed that there is an
198 estimation of the disease prevalence based on a health survey or other studies.

199

200 Global hypothesis test

201 The PVs of each BDT depend on the same parameters, the sensitivity and the specificity of
202 the test and disease prevalence, and therefore they are parameters that depend on each other.

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

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

which is equivalent to the hypothesis test

$$H_0 : \mathbf{A}\boldsymbol{\omega} = \mathbf{0} \text{ vs } H_1 : \mathbf{A}\boldsymbol{\omega} \neq \mathbf{0}, \quad (3)$$

where $\mathbf{A}$ is a complete range design matrix, i.e.

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

As the vector $\hat{\boldsymbol{\omega}}$ is distributed asymptotically according to a multivariate normal distribution, i.e. $\sqrt{n_1 + n_2}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{n_1 + n_2 \rightarrow \infty} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$, then the statistic for the global hypothesis test (3) is

$$Q^2 = \hat{\boldsymbol{\omega}}^T \mathbf{A}^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}}, \quad (4)$$

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 $\mathbf{A}\hat{\boldsymbol{\omega}}$.

When $n_1 + n_2$ is large, the statistic Q^2 is distributed according to a central chi-square distribution with 2 degrees of freedom when the null hypothesis is true.

Individual hypothesis tests

The hypothesis test to individually compare the two PPVs (NPVs) is

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

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

$$z = \frac{|\widehat{PV}_1 - \widehat{PV}_2|}{\sqrt{\widehat{Var}(\widehat{PV}_1) + \widehat{Var}(\widehat{PV}_2) - 2\widehat{Cov}(\widehat{PV}_1, \widehat{PV}_2)}}, \quad (5)$$

which is distributed asymptotically according to a normal standard distribution, and where the variances-covariances is obtained from the equation (2).

Alternative methods to the global test

The global hypothesis test (3) 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 $H_0 : NPV_1 = NPV_2$, each one to an error α ; 2) Testing the hypotheses $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$, and applying a multiple comparison method such as Bonferroni's method [12] or Holm's method [13], which are methods that are very easy to apply based on the p-values. Bonferroni's method consists of solving each individual hypothesis test to an error $\alpha/2$; and Holm's method is a step-down method which is based on Bonferroni's method but is more conservative.

Simulation experiments

Simulation experiments were carried out to study the type I errors and the powers of the four methods proposed to solve the global hypothesis test: the hypothesis test based on the chi-square (equation (4)), the individual hypothesis tests each one to an error α , and the individual hypothesis tests applying Bonferroni's method and Holm's method. We have also studied the effect of a misspecification of the prevalence on the asymptotic behaviour of the global hypothesis test 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 $\{0.70, 0.75, \dots, 0.90, 0.95\}$, and as disease prevalence we took the values 10%, 25% and 50%. Based on the PVs and the prevalence, the Se and the Sp of each BDT were calculated from the equations (1), i.e.

$$Se_i = \frac{PPV_i(NPV_i - q)}{p(PPV_i + NPV_i - 1)} \text{ and } Sp_i = \frac{NPV_i(PPV_i - p)}{q(PPV_i + NPV_i - 1)}$$

with $i = 1, 2$, $q = 1 - p$ and only considering those cases in which the solutions are between 0 and 1. As values of the correlation coefficients ρ^+ and ρ^- we took low values (25% of the maximum value), intermediate ones (50% of the maximum value) and high ones (75% of the maximum value), and the maximum value of each correlation coefficient is:

$$\max(\rho^+) = \frac{\min\{Se_1(1 - Se_2), (1 - Se_1)Se_2\}}{\sqrt{Se_1(1 - Se_1)Se_2(1 - Se_2)}} \text{ and } \max(\rho^-) = \frac{\min\{Sp_1(1 - Sp_2), (1 - Sp_1)Sp_2\}}{\sqrt{Sp_1(1 - Sp_1)Sp_2(1 - Sp_2)}}.$$

As sample sizes, we took the values $n_i = \{50, 75, 100, 200, 500\}$. The simulation experiments were carried out with R, using the “bindata” package 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 and the specificities and maximum values of the coefficients ρ^+ and ρ^- . We then generated 10,000 samples with a type I bivariate binomial distribution with a sample size n_1 , probabilities Se_1 and Se_2 and correlation coefficient ρ^+ , and another 10,000 samples with a type I bivariate binomial distribution with a sample size n_0 , probabilities Sp_1 and Sp_2 and correlation coefficient ρ^- . 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 $n_{2.1} = n_2 - n_{2.0}$. Then and in order to construct the 2×2 table of each case simple, we generated a random valor n_{111} of a doubly truncated binomial distribution of parameters n_1 and $\xi_{111} = Se_1Se_2 + \varepsilon^+$ with $n_{11.} + n_{1.1} - n_1 \leq n_{111} \leq \text{Min}\{n_{11.}, n_{1.1}\}$. 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} of a doubly truncated binomial distribution of parameters n_2 and $\xi_{200} = Sp_1Sp_2 + \varepsilon^-$ with $n_{20\cdot} + n_{2\cdot0} - n_2 \leq n_{200} \leq \text{Min}\{n_{20\cdot}, n_{2\cdot0}\}$. For each one of the 10,000 case (control) samples, once we have generated the values $n_{11\cdot}$, $n_{1\cdot1}$ and n_{111} ($n_{20\cdot}$, $n_{2\cdot0}$ and n_{200}) it is easy to construct the complete 2×2 table. Thus, $n_{110} = n_{1\cdot} - n_{11\cdot}$, $n_{101} = n_{1\cdot1} - n_{111}$ and $n_{100} = n_{10\cdot} - n_{101}$ for the case samples, and $n_{201} = n_{20\cdot} - n_{200}$, $n_{210} = n_{2\cdot0} - n_{200}$ and $n_{211} = n_{21\cdot} - n_{210}$. For the experiments, the error $\alpha = 5\%$ 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.

Type I errors and powers

In Table 2, we can see some results obtained for the type I errors of the global test and of the alternative methods previously proposed. In this table, we can only see the results for the global test, the individual comparisons with $\alpha = 5\%$ and with Bonferroni's method. The results obtained with Holm's method are not shown as they are practically the same as those obtained with Bonferroni's 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 error $\alpha = 5\%$, 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 the error $\alpha = 5\%$. 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 $PPV_1 = PPV_2 = 0.85$ and $NPV_1 = NPV_2 = 0.95$.

$Se_1 = 0.5312, Sp_1 = 0.9896, Se_2 = 0.5312, Sp_2 = 0.9896$										
$0 \leq \rho^+ \leq 1, 0 \leq \rho^- \leq 1$										
$p = 10\%$										
		$\rho^+ = 0.25 \quad \rho^- = 0.25$			$\rho^+ = 0.50 \quad \rho^- = 0.50$			$\rho^+ = 0.75 \quad \rho^- = 0.75$		
n_1	n_2	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
$Se_1 = 0.85, Sp_1 = 0.95, Se_2 = 0.85, Sp_2 = 0.95$										
$0 \leq \rho^+ \leq 1, 0 \leq \rho^- \leq 1$										
$p = 25\%$										
		$\rho^+ = 0.25 \quad \rho^- = 0.25$			$\rho^+ = 0.50 \quad \rho^- = 0.50$			$\rho^+ = 0.75 \quad \rho^- = 0.75$		
n_1	n_2	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
$Se_1 = 0.9562, Sp_1 = 0.8312, Se_2 = 0.9562, Sp_2 = 0.8312$										
$0 \leq \rho^+ \leq 1, 0 \leq \rho^- \leq 1$										
$p = 50\%$										
		$\rho^+ = 0.25 \quad \rho^- = 0.25$			$\rho^+ = 0.50 \quad \rho^- = 0.50$			$\rho^+ = 0.75 \quad \rho^- = 0.75$		
n_1	n_2	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 $\alpha = 5\%$.

Bonf.: Bonferroni's 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 6.5%), 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 Bonferroni's (Holm's) method, the type I error has a very similar behaviour to that of the global hypothesis test.

Regarding the power of the hypothesis tests, in Table 3 we can see some of the results obtained for the global test and other alternative methods. The results obtained with Holm's method are not shown as they are practically the same as those obtained with Bonferroni's method. 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^+ and ρ^- , these do not have a clear effect on the power, and the power increases sometimes and decreases other times when the correlations increase. In very general terms, when the prevalence is relatively small ($p = 10\%$) we need large samples ($n_i > 500$) so that the power of the global hypothesis test (equation (4)) 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.

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 tests based on the individual tests with Bonferroni's method and Holm's method, their corresponding power is practically the same, and is also very similar to the power of the global test based on the chi-square distribution.

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 Bonferroni's method or Holm's 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.

From the results obtained, we propose the following method to compare the PVs of two BDTs subject to a case-control design: 1) Applying the hypothesis test based on the chi-square distribution (equation (4)) to an error α , 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 error α , the investigation of the causes of the significance is made by testing the individual tests (equation (5)) and applying Bonferroni's method or Holm's method to an error α .

Table 3. Powers for $PPV_1 = 0.95$, $NPV_1 = 0.95$, $PPV_2 = 0.75$ and $NPV_2 = 0.95$.

$Se_1 = 0.5278$, $Sp_1 = 0.9969$, $Se_2 = 0.5357$, $Sp_2 = 0.9802$										
$0 \leq \rho^+ \leq 0.9841$, $0 \leq \rho^- \leq 0.3910$										
$p = 10\%$										
$\rho^+ = 0.25$ $\rho^- = 0.10$			$\rho^+ = 0.49$ $\rho^- = 0.19$			$\rho^+ = 0.74$ $\rho^- = 0.29$				
n_1	n_2	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
$Se_1 = 0.8444$, $Sp_1 = 0.9852$, $Se_2 = 0.8571$, $Sp_2 = 0.9048$										
$0 \leq \rho^+ \leq 0.9511$, $0 \leq \rho^- \leq 0.3779$										
$p = 25\%$										
$\rho^+ = 0.24$ $\rho^- = 0.09$			$\rho^+ = 0.48$ $\rho^- = 0.19$			$\rho^+ = 0.71$ $\rho^- = 0.28$				
n_1	n_2	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
$Se_1 = 0.95$, $Sp_1 = 0.95$, $Se_2 = 0.9643$, $Sp_2 = 0.6786$										
$0 \leq \rho^+ \leq 0.8388$, $0 \leq \rho^- \leq 0.3333$										
$p = 50\%$										
$\rho^+ = 0.21$ $\rho^- = 0.08$			$\rho^+ = 0.42$ $\rho^- = 0.17$			$\rho^+ = 0.63$ $\rho^- = 0.25$				
n_1	n_2	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 $\alpha = 5\%$.

Bonf.: Bonferroni's method.

Effect of the prevalence

The estimation and comparison of the PVs of two BDTs subject to a case-control design requires knowledge of an estimation of the disease prevalence obtained from another study,

e.g. a health survey. To study the effect of a misspecification of the prevalence in the comparison of the PVs of two BDTs and in 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 an overestimation (and an underestimation) equal to 5% and to 10% 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*) defined as

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

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 prevalences $p = \{10\%, 25\%, 50\%\}$ respectively, and to estimate the PVs we took as prevalences $p' = p \pm d \times p$ with $d = \{5\%, 10\%\}$. In Table 4, we show 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 the prevalence is underestimated ($p' < p$) and overestimated ($p' > p$). From the results of these experiments, it is verified that the type I errors of the methods studied do not overwhelm the nominal error $\alpha = 5\%$, and 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, in Table 4 we show some of the results obtained for the RRMSEs (in %) of the estimators of the PVs of the two BDTs. There is no important difference between the RRMSEs when there is a misspecification of the prevalence ($p' < p$ or $p' > p$) and the RRMSEs when there is no misspecification of the prevalence ($p' = p$). In general terms, this difference is not usually over 5% when the samples are small, and this is even lower when the samples are large.

Consequently, misspecifications (5% or 10%) of the disease prevalence do not have any important effect on the type I errors and on the powers of the global hypothesis test and on the alternative methods (Bonferroni and Holm), and nor do they have an important effect on the estimators of the PVs.

Table 4. Effect of a misspecification of the prevalence.

Type I errors											
$PPV_1 = PPV_2 = 0.90$, $NPV_1 = NPV_2 = 0.80$											
$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.2571$, $Sp_2 = 0.9905$, $\rho^+ = 0.75$, $\rho^- = 0.75$											
$p = 25\%$											
n_1	n_2	$p' = p = 25\%$		$p' = 22.50\%$		$p' = 23.75\%$		$p' = 26.25\%$		$p' = 27.50\%$	
		Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.
50	50	0.001	0.001	0.001	0.002	0.001	0.001	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	0.010	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	0.024	0.020	0.023
Powers											
$PPV_1 = 0.90$, $PPV_2 = 0.70$, $NPV_1 = 0.80$, $NPV_2 = 0.90$											
$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.70$, $Sp_2 = 0.90$, $\rho^+ = 0.29$, $\rho^- = 0.22$											
$p = 25\%$											
n_1	n_2	$p' = p = 25\%$		$p' = 22.50\%$		$p' = 23.75\%$		$p' = 26.25\%$		$p' = 27.50\%$	
		Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.	Global	Bonf.
50	50	0.999	0.992	0.999	0.993	0.999	0.992	0.999	0.992	0.999	0.991
50	75	1	0.996	1	0.995	1	0.996	1	0.996	1	0.995
50	100	1	0.998	1	0.997	1	0.997	1	0.998	1	0.998
75	75	1	1	1	1	1	1	1	1	1	1
100	100	1	1	1	1	1	1	1	1	1	1
200	200	1	1	1	1	1	1	1	1	1	1
500	500	1	1	1	1	1	1	1	1	1	1
RRMSEs of the estimators of PVs of BDT 1											
$PPV_1 = 0.90$, $PPV_2 = 0.70$, $NPV_1 = 0.80$, $NPV_2 = 0.90$											
$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.70$, $Sp_2 = 0.90$, $\rho^+ = 0.29$, $\rho^- = 0.22$											
$p = 25\%$											
n_1	n_2	$p' = p = 25\%$		$p' = 22.50\%$		$p' = 23.75\%$		$p' = 26.25\%$		$p' = 27.50\%$	
		$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$
50	50	26.8	1.8	31.1	2.7	29.3	2.1	28.3	2.1	29.7	3.2
50	75	20.0	1.7	22.9	2.9	21.6	2.0	20.9	2.3	21.8	3.1
50	100	15.6	1.6	18.3	3.0	16.8	1.9	15.9	2.2	17.8	3.0
75	75	18.4	1.4	22.3	2.5	21.2	1.7	19.3	2.0	21.9	2.8
100	100	14.2	1.2	18.0	2.2	16.0	1.6	14.9	1.9	17.1	2.5
200	200	8.0	0.9	9.4	1.8	8.9	1.5	8.5	1.6	8.9	1.9
500	500	4.1	0.5	5.4	1.5	4.6	1.2	4.5	1.2	5.1	1.6
RRMSEs of the estimators of PVs of BDT 2											
n_1	n_2	$p' = p = 25\%$		$p' = 22.50\%$		$p' = 23.75\%$		$p' = 26.25\%$		$p' = 27.50\%$	
		$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$	$\widehat{PPV}_1$	$\widehat{NPV}_1$
50	50	10.8	2.3	14.3	2.4	13.0	2.3	11.4	2.6	13.6	2.7
50	75	10.3	2.2	12.4	2.2	11.1	2.2	10.8	2.5	11.8	2.6
50	100	9.1	2.0	11.0	2.1	9.8	2.1	9.6	2.5	11.3	2.5
75	75	10.0	1.8	12.1	2.0	10.9	1.9	10.5	2.1	12.4	2.3
100	100	8.9	1.6	10.6	1.8	9.5	1.7	9.3	1.9	10.9	2.1
200	200	6.7	1.1	8.0	1.6	6.9	1.2	7.0	1.4	8.1	1.7
500	500	4.5	0.7	5.5	1.3	4.7	0.9	4.7	1.0	5.5	1.5

Global: Global hypothesis test based on the chi-square distribution.

Bonf.: Bonferroni's method.

Example

The results obtained were applied to the diagnosis of Human African Trypanosomiasis (HAT). HAT, also known as sleeping sickness, is a parasitic disease caused by protozoa belonging to the genus *Trypanosoma*, and it is transmitted to human beings by a bite from the tsetse fly (genus *Glossina*) infected by other people or animals that host human pathogenic parasites. The HAT threatens millions of people in sub-Saharan Africa countries. According to the World Health Organization [14], during the most recent periods of epidemic, in some areas of these countries the prevalence of this disease reached 50%, being the first or second cause of mortality, even ahead of HIV/AIDS. Matovu et al [15] applied two diagnostic tests to a sample of 75 cases and another sample of 65 controls in Uganda. In Table 7 (observed frequencies) we can see the frequencies obtained (constructed from the data provided by Matovu et al) when applying the *PCR-Oligochromatography* (*PCR-OC*, variable T_1) test and the *NASBA-Oligochromatography* (*NASBA-OC*, variable T_2) test to both samples of individuals. In order to illustrate the method proposed in this article, two values were considered for the prevalence of HAT in epidemic periods: 10% and 50%. In Table 5, we can also see the estimations of the sensitivities and the specificities (and their standard errors SE) of the BDTs and the correlations.

Table 5. Study by Matuovo et al.

Observed frequencies					
Case			Control		
	$T_2 = 1$	$T_2 = 0$	Total		
$T_1 = 1$	57	4	61	$T_1 = 1$	1
$T_1 = 0$	6	8	14	$T_1 = 0$	4
Total	63	12	75	Total	60
Sensitivities, specificities and correlations					
$\hat{S}e_1 \pm SE$	$\hat{S}e_2 \pm SE$	$\hat{\rho}^+$	$\hat{S}p_1 \pm SE$	$\hat{S}p_2 \pm SE$	$\hat{\rho}^+$
0.813 ± 0.045	0.84 ± 0.042	0.538	0.923 ± 0.033	0.985 ± 0.015	0.433

429 For a prevalence value equal to 10%, the estimations of the PVs are $\widehat{PPV}_1 = 0.540$,
 430 $\widehat{PPV}_2 = 0.858$, $\widehat{NPV}_1 = 0.978$ and $\widehat{NPV}_2 = 0.982$, and the estimated variance and covariance
 431 matrix of the estimators of the PVs is

$$432 \quad \widehat{\Sigma}_{\hat{\omega}} = \begin{pmatrix} 0.01158 & 0.00562 & 0.00015 & 0.00005 \\ 0.00562 & 0.01456 & 0.00006 & 0.00006 \\ 0.00015 & 0.00006 & 0.00003 & 0.00001 \\ 0.00005 & 0.00006 & 0.00001 & 0.00002 \end{pmatrix}.$$

433 The value of the test statistic for the test

$$434 \quad \begin{aligned} H_0 : PPV_1 &= PPV_2 \text{ and } NPV_1 = NPV_2 \\ H_1 : &\text{at least one equality is not true,} \end{aligned}$$

435 is $Q^2 = 6.954$ (P -value = 0.031) and therefore null hypothesis of the global test is rejected.

436 Testing the individual hypothesis tests it is found that the value of the test statistic for the

437 $H_0 : PPV_1 = PPV_2$ is equal to 2.606 (two sided p-value = 0.009), and that the value of the test

438 statistic for the test $H_0 : NPV_1 = NPV_2$ is equal to 0.886 (two sided p-value = 0.375).

439 Applying Bonferroni's (Holm's) method the equality hypothesis of the negative predictive

440 values is not rejected and the equality hypothesis of the two positive predictive values is

441 rejected. The positive predictive value of the *NASBA-OC* test is significantly greater than that

442 of the *PCR-OC* test (95% CI: 0.079 to 0.558).

443 For a HAT prevalence equal to 50%, the estimations of the PVs are $\widehat{PPV}_1 = 0.914$,

444 $\widehat{PPV}_2 = 0.982$, $\widehat{NPV}_1 = 0.834$ and $\widehat{NPV}_2 = 0.860$, and estimated variance and covariance

445 matrix of the estimators of the PVs is

$$446 \quad \widehat{\Sigma}_{\hat{\omega}} = \begin{pmatrix} 0.00117 & 0.00026 & 0.00032 & 0.00010 \\ 0.00026 & 0.00031 & 0.00005 & 0.00006 \\ 0.00032 & 0.00005 & 0.00116 & 0.00058 \\ 0.00010 & 0.00006 & 0.00058 & 0.00102 \end{pmatrix}.$$

447 The value of the test statistic for the test

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

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

is $Q^2 = 5.048$ ($P\text{-value} = 0.080$), and therefore with an error $\alpha = 5\%$ we do not reject the equality of the positive predictive values and the negative predictive values of both diagnostic tests, although there are signs of significance and, therefore, an increase in the two sample sizes may be necessary. Solving the individual hypothesis tests it is found that the value of the test statistic for the test $H_0 : PPV_1 = PPV_2$ is equal to 2.21 (two sided p-value = 0.027), and that the value of the test statistic for the $H_0 : NPV_1 = NPV_2$ is equal to 0.892 (two sided p-value = 0.373). If each individual hypothesis test is solved to an error $\alpha = 5\%$, then $H_0 : PPV_1 = PPV_2$ is rejected and $H_0 : NPV_1 = NPV_2$ is not rejected, and the result is contradictory to that obtained with the global test.

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, analysing 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 a nominal error α , and another three methods based on individual comparisons along with a multiple comparison method. The multiple comparison methods that were used were Bonferroni's method and Holm's method, which are methods based on the p-values of the individual hypothesis tests and are very easy to apply.

Simulation experiments were carried out to study the type I errors and the power of the four methods proposed. These experiments were based on the generation samples with type I bivariate binomial distributions, which are the distributions that are inherent to case-control design, since from these samples proportions of marginal totals are estimated. 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 $\alpha = 5\%$. Regarding its power, in general this strongly depends on the disease prevalence, and it is necessary to have very large samples ($n_i > 500$) when the prevalence is small and relatively small sample sizes ($n_i \geq 50$) when the prevalence is high, so that the power will be high.

The simulation experiments also showed that the methods based on individual hypothesis tests along with multiple comparison methods have type I errors and very similar power to those of the global test based on the chi-square distribution. Consequently, both methods can be used to compare the PVs of the two BDTs. Furthermore, the experiments also showed that the comparison of the predictive values of two diagnostic tests cannot be made independently i.e. comparing the two positive predictive values and comparing the two negative predictive values independently to an error $\alpha = 5\%$, as it is possible to obtain a type I error that clearly overwhelms the nominal error set.

Simulation experiments were carried out to study the effect of a misspecification of the prevalence in the asymptotic behaviour of the global hypothesis test based on the chi-square distribution and on the methods based on multiple comparisons. From the results obtained, we can conclude that light or moderate overestimations or underestimations of the prevalence do not have an important effect on the behaviour of these hypothesis tests.

Conclusion

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 (equation (4)) to an error α ; 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 error α , then the causes of the significance are studied solving the individual hypothesis tests (equation (5)) and applying Bonferroni's method or Holm's method to an error α . 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.

Authors' contributions

RN and SR performed all statistical models mentioned in the manuscript. The authors read and approved the final manuscript.

Funding

This research was supported by the Spanish Ministry of Economy, Grant Number MTM2016-76938-P.

Competing interests

The authors declare that they have no competing interests.

References

- 1 Pepe, MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2003.
- 2 Bennett, BM. On comparison of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics*. 1972;28:793-800.
- 3 Bennett, BM. On tests for equality of predictive values for t diagnostic procedures. *Statistics in Medicine*, 1985;4:535-539.
- 4 Leisenring, W, Alonzo, T, Pepe, MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics*. 2000;56:345-351.
- 5 Wang, W, Davis, CS, Soong SJ. Comparison of predictive values of two diagnostic tests from the same sample of subjects using weighted least squares. *Statistics in Medicine*. 2006;25:2215-2229.
- 6 Kosinski, AS. (2013). A weighted generalized score statistic for comparison of predictive values of diagnostic tests. *Statistics in Medicine*. 2013;32:964-977.
- 7 Moskowitz, CS, Pepe, MS. Comparing the predictive values of diagnostic tests: sample size and analysis for paired study designs. *Clinical Trials*. 2006;3:272-279.
- 8 Roldán-Nofuentes, JA, Luna del Castillo, JD, Montero-Alonso, MA. Global hypothesis test to simultaneously compare the predictive values of two binary diagnostic tests. *Computational Statistics and Data Analysis*, Special issue “Computational Statistics for Clinical Research”. 2012;56:1161-1173.
- 9 Mercaldo, ND, Kit, FL, Zhou, XH. Confidence intervals for predictive values with an emphasis to case-control studies. *Statistics in Medicine*. 2007;26:2170-2183.
- 10 Vacek, PM. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*. 1985;41:959-968.

- 544 11 Kocherlakota, S, Kocherlakota, K. Bivariate discrete distributions. New York: Marcel
545 Dekker INC; 1992.
- 546 12 Bonferroni, CE. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del
547 R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 1936;8:3-62.
- 548 13 Holm, S. A simple sequential rejective multiple testing procedure. Scandinavian Journal of
549 Statistics. 1979;6:65-70.
- 550 14 World Health Organization. [www.who.int/news-room/fact-sheets/detail/trypanosomiasis-](http://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness))
551 [human-african-\(sleeping-sickness\)](http://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness)). Accessed 4 Jul 2019.
- 552 15 Matuvo, E, Mugasa, CM, Ekangu, RA, Deborggrave, S, Lubega, GW, Laurent, T,
553 Schoone, GJ, Schallig, HD, Büscher, P. Phase II evaluation of sensitivity and specificity of
554 PCR and NASBA followed by oligochromatography for diagnosis of human African
555 trypanosomiasis in clinical samples from D.R. Congo and Uganda. PLOS Neglected
556 Tropical Diseases. 2010;4:e737.