Comparison of the predictive values of diagnostic tests subject to a case-control sampling with application to the diagnosis of Human African Trypanosomiasis

José Antonio Roldán-Nofuentes[1*], Saad Bouh Sidaty-Regad[2]

[1]Department of Statistics, School of Medicine, University of Granada, Spain

[2]Department of Public Health and Epidemiology, School of Medicine, University of Nouakchott, Mauritania

[*]Corresponding author

Email: jaroldan@ugr.es

Abstract. Case-control sampling to compare the accuracy of two binary diagnostic tests is frequent in clinical practice. This type of sampling 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 sampling, the sensitivities are compared from the case sample applying the McNemar's test, and the specificities from the control sample. Other parameters of binary 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. 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 the disease prevalence. 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

1

25   prevalence on the asymptotic behavior of the hypothesis tests and on the estimators of the

26   predictive values. The results obtained were applied to a real example on the diagnosis of the

27   Human African Trypanosomiasis. The model proposed was extended to the situation in which

28   there are more than two diagnostic tests.

29   **Keywords:** Case-control sampling, Diagnostic test, Human African Trypanosomiasis,

30   Predictive values.

31

1. Introduction

33   The main parameters to assess and compare the accuracy of binary diagnostic tests (BDTs)

34   are sensitivity and specificity. The sensitivity (Se) is the probability of the result of the BDT

35   being positive when the individual has the disease and the specificity (Sp) is the probability of

36   the result of the BDT being negative when the individual does not have the disease. Other

37   parameters that are used to assess and compare two BDTs are the predictive values (PVs).

38   The positive predictive value (PPV) is the probability of an individual having the disease

39   when the result of the BDT is positive, and the negative predictive value (NPV) is the

40   probability of an individual not having the disease when the result of the BDT is negative.

41   The PVs represent the accuracy of the diagnostic test when it is applied to a cohort of

42   individuals, and they are measures of the clinical accuracy of the BDT. The PVs depend on

43   the Se and the Sp of the BDT and on the disease prevalence (p), and are easily calculated

44   applying Bayes' Theorem i.e.

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

46   Whereas the Se and the Sp quantify how well the BDT reflects the true disease status (present

47   or absent), the PVs quantify the clinical value of the BDT, since both the individual and the

2

48   clinician are more interested in knowing how probable it is to have the disease given a BDT

49   result.

50   The comparison of the performance of two binary diagnostic tests is a topic of special

51   importance in the study of statistical methods for the diagnosis of diseases. This comparison

52   can be made through a cross-sectional sampling or a case-control sampling. Cross-sectional

53   sampling consists of applying the two BDTs and the gold standard to all of the individuals in

54   a single sample. Case-control sampling consists of applying the two BDTs to all of the

55   individuals in two samples, one made up of individuals who have the disease (case sample)

56   and another made up of individuals who do not have the disease (control sample). The

57   advantages and disadvantages of case-control sampling over the cross-sectional can be seen in

58   the book by Pepe (2003). Summarizing, case-control sampling has some advantages over

59   cross-sectional: a) case-control design is more efficient in terms of sample size requirements,

60   b) case-control studies allow for the exploration of subject-related characteristics of the test.

61   Nevertheless, the case-control design has the disadvantage is that by using it we cannot

62   estimate the prevalence of the disease.

63   The comparison of the sensitivities and the specificities of two BDTs subject to cross-

64   sectional sampling or subject to case-control sampling is made applying the exact comparison

65   test of two paired binomial proportions or McNemar's test (the asymptotic version of the

66   exact test).

67   In cross-sectional sampling, the comparison of PVs has been the subject of several studies.

68   Bennett (1972, 1985), Leisenring et al (2000), Wang et al (2006) and Kosinski (2013) studied

69   hypothesis tests to independently compare the PPVs and the NPVs of two BDTs. Moskowitz

70   and Pepe (2006) studied the estimation of the PVs through a confidence region. Roldán-

71   Nofuentes et al (2012) studied the joint comparison of the PPVs and NPVs of two BDTs, and

72   proposed a global hypothesis test based on the chi-square distribution to simultaneously

73   compare the PVs of two BDTs.

74   In a case-control sampling, Mercaldo et al (2007) have studied the estimation of the PVs of

75   a BDT, assuming that the disease prevalence is known. In this article, we extended the study

76   of Mercaldo et al to the case of two BDTs, studying the comparison of the PVs of the two

77   BDTs subject to a case-control sampling. Subject to a case-control sampling, the two BDTs

78   are applied to all of the individual in two samples, one of n_1 individuals who have the disease

79   (case sample) and another with n_2 individuals who do not have the disease (control sample).

80   In this sampling, the sample sizes n_1 and n_0 are set by the researcher. The sample of

81   individuals that have the disease is extracted from a population of individuals that have the

82   disease (e.g. registers of diseases), and the control sample is extracted from a population of

83   individuals who are known not to have the disease. As the PVs depend on the disease

84   prevalence and subject to a case-control sampling the quotient $n_1/(n_1 + n_2)$ is not an estimator

85   of the prevalence, in order to estimate and compare the PVs subject to this sampling it is

86   necessary to know the prevalence or an estimate of the prevalence. This estimation can be

87   obtained from health surveys or from previous studies. Consequently, the methods of

88   comparison of the PVs subject to a cross-sectional sampling cannot be applied when there is a

89   case-control sampling. In Section 2, we study hypothesis tests to jointly and individually

90   compare the PVs of two BDTs subject to case-control sampling. In Section 3, simulation

91   experiments are carried out to study the type I errors and the powers of the hypothesis tests

92   proposed in Section 2. In Section 4, we study the effect of the misspecification of the

93   prevalence on the asymptotic behavior of the hypothesis tests proposed in Section 2 and on

94   the estimators of the PVs. In Section 5, the results are applied to a real example of the

95   diagnosis of Human African Trypanosomiasis. In Section 6, the model proposed in Section 2

96   was extended to the situation in which we compare the PVs of more than two BDTs, and in

97   Section 7 the results are discussed.

98

2. The model

100   Let us consider two BDTs, Test 1 and Test 2, which are applied to all of the individuals in

101   two samples, one of n_1 individuals who have the disease (case sample) and another of n_2

102   individuals who do not have it (control sample). Let T_1 and T_2 be two binary variables that

103   model the results of each BDT, in such a way that $T_i = 1$ when the result of the corresponding

104   BDT is positive and $T_i = 0$ when it is negative. In Table 1, we can see the probabilities

105   associated to the application of both BDTs to both types of individuals (cases and controls), as

106   well as the frequencies observed.

107

108   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}	$n_{11\square}$	$T_1 = 1$	n_{211}	n_{210}	$n_{21\square}$
$T_1 = 0$	n_{101}	n_{100}	$n_{10\square}$	$T_1 = 0$	n_{201}	n_{200}	$n_{20\square}$
Total	$n_{1\square1}$	$n_{1\square0}$	n_1	Total	$n_{2\square1}$	$n_{2\square0}$	n_2

109

110   Using the conditional dependence model of Vacek (1985), the probabilities given in the table

111   are written as

$$\xi_{1jk} = Se_1^j \left(1 - Se_1\right)^{1-j} Se_2^k \left(1 - Se_2\right)^{1-k} + \delta_{jk}\varepsilon^+ \qquad (2)$$

5

113 and

$$\xi_{2jk} = Sp_1^{1-j}\left(1-Sp_1\right)^j Sp_2^{1-k}\left(1-Sp_2\right)^k + \delta_{jk}\varepsilon^-, \tag{3}$$

115 with $j,k = 0,1$. The parameter ε^+ $\left(\varepsilon^-\right)$ is the covariance between the two BDTs in cases

116 (controls), where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = -1$ if $j \neq k$, and it is verified that

117 $0 \leq \varepsilon^+ \leq \mathrm{Min}\left\{Se_1\left(1-Se_2\right), Se_2\left(1-Se_1\right)\right\}$ and $0 \leq \varepsilon^- \leq \mathrm{Min}\left\{Sp_1\left(1-Sp_2\right), Sp_2\left(1-Sp_1\right)\right\}$. If

118 $\varepsilon^+ = \varepsilon^- = 0$ then the two BDTs are conditionally independent on the disease status. In

119 practice, the assumption of the conditional independence is not realistic, and therefore $\varepsilon^+ > 0$

120 and/or $\varepsilon^- > 0$. In terms of the probabilities ξ_{ijk}, the sensitivities are written as

121 $$Se_1 = \xi_{111} + \xi_{110} \text{ and } Se_2 = \xi_{111} + \xi_{101},$$

122 and the specificities as

123 $$Sp_1 = \xi_{201} + \xi_{200} \text{ and } Sp_2 = \xi_{210} + \xi_{200}.$$

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

125 $$\hat{Se}_1 = \frac{n_{11\square}}{n_1} \text{ and } \hat{Se}_2 = \frac{n_{1\square1}}{n_1},$$

126 and

127 $$\hat{Sp}_1 = \frac{n_{20\square}}{n_2} \text{ and } \hat{Sp}_2 = \frac{n_{2\square0}}{n_1},$$

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

129 $\hat{Var}\left(\hat{Se}_2\right) = \hat{Se}_2\left(1-\hat{Se}_2\right)\big/n_1$, $\hat{Var}\left(\hat{Sp}_1\right) = \hat{Sp}_1\left(1-\hat{Sp}_1\right)\big/n_2$ and $\hat{Var}\left(\hat{Sp}_2\right) = \hat{Sp}_2\left(1-\hat{Sp}_2\right)\big/n_2$.

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

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

132 $n_{1\square1}$, and therefore these frequencies are the product of a type I bivariate binomial distribution

133 (Kocherlakota and Kocherlakota, 1992). In an analogous way, from the control sample, the

6

134 marginal frequencies $n_{20\square}$ and $n_{2\square0}$ are the product of a type I bivariate binomial distribution.

135 In the individuals with the disease, the type I bivariate binomial distribution is characterized

136 (Kocherlakota and Kocherlakota, 1992) by the two probabilities Se_1 and Se_2 and by the

137 correlation coefficient (ρ^+) between T_1 and T_2. In an analogous way, in the individuals who

138 do not have the disease, the type I bivariate binomial distribution is characterized by Sp_1, Sp_2

139 and the correlation coefficient (ρ^-) between T_1 and T_2. In the individuals with the disease

140 (cases), the correlation coefficient between the two BDTs is

$$\rho^+ = \frac{\xi_{111} - Se_1 Se_2}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}} = \frac{\varepsilon^+}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}},$$ (4)

142 and in the individuals who do not have the disease (controls), the correlation coefficient

143 between the two BDTs is

$$\rho^- = \frac{\xi_{200} - Sp_1 Sp_2}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}} = \frac{\varepsilon^-}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}}.$$ (5)

145 It is easy to check that

$$\hat{\varepsilon}^+ = \frac{n_1 n_{111} - n_{11\square} n_{1\square}}{n_1^2} \text{ and } \hat{\varepsilon}^- = \frac{n_2 n_{200} - n_{20\square} n_{2\square0}}{n_2^2},$$

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

148 and

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

150 All of the other covariances are zero, since the two samples are independent. The estimators

151 of ρ^+ and ρ^- are

$$\hat{\rho}^+ = \frac{n_1 n_{111} - n_{11\square} n_{1\square}}{\sqrt{n_{11\square}(n_1 - n_{11\square})n_{1\square}(n_1 - n_{1\square})}} \text{ and } \hat{\rho}^- = \frac{n_2 n_{200} - n_{20\square} n_{2\square0}}{\sqrt{n_{20\square}(n_2 - n_{20\square})n_{2\square0}(n_2 - n_{2\square0})}}.$$

7

153   Assuming that prevalence p (or an estimation) is known, the estimators of the predictive

154   values are

$$\hat{P}PV_1 = \frac{pn_2 n_{11\cdot}}{pn_2 n_{11\cdot} + qn_1 (n_2 - n_{20\cdot})} \quad \text{and} \quad \hat{N}PV_1 = \frac{qn_1 n_{20\cdot}}{pn_2 (n_1 - n_{11\cdot}) + qn_1 n_{20\cdot}} \qquad (6)$$

155

156   for Test 1, and

$$\hat{P}PV_2 = \frac{pn_2 n_{1\cdot 1}}{pn_2 n_{1\cdot 1} + qn_1 (n_2 - n_{2\cdot 0})} \quad \text{and} \quad \hat{N}PV_2 = \frac{qn_1 n_{2\cdot 0}}{pn_2 (n_1 - n_{1\cdot 1}) + qn_1 n_{2\cdot 0}} \qquad (7)$$

157

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

$$\Sigma_{\hat{S}e} = \begin{pmatrix} Var(\hat{S}e_1) & Cov(\hat{S}e_1, \hat{S}e_2) \\ Cov(\hat{S}e_1, \hat{S}e_2) & Var(\hat{S}e_2) \end{pmatrix} \qquad (8)$$

159

160   and

$$\Sigma_{\hat{S}p} = \begin{pmatrix} Var(\hat{S}p_1) & Cov(\hat{S}p_1, \hat{S}p_2) \\ Cov(\hat{S}p_1, \hat{S}p_2) & Var(\hat{S}p_2) \end{pmatrix}. \qquad (9)$$

161

162   Let $\boldsymbol{\theta} = (Se_1, Se_2, Sp_1, Sp_2)^T$ be a vector whose components are the sensitivities and the

163   specificities, and let $\boldsymbol{\omega} = (PPV_1, PPV_2, NPV_1, NPV_2)^T$ be a vector whose components are the

164   predictive values. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \Sigma_{\hat{S}e} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \Sigma_{\hat{S}p}, \qquad (10)$$

165

166   where $\otimes$ is the product of Kronecker. Applying the delta method, the matrix of variances-

167   covariances of $\hat{\boldsymbol{\omega}}$ is

$$\Sigma_{\hat{\boldsymbol{\omega}}} = \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\theta}} \right) \Sigma_{\hat{\boldsymbol{\theta}}} \left(\frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\theta}} \right)^T. \qquad (11)$$

168

169   In Appendix A, we can see the expressions of the variances-covariances of the PVs. Then, we

170   study the joint comparison and the individual comparison of the PVs of the two BDTs. In

8

171 both cases, and as has been explained in Section 1, it is assumed that there is an estimation of

172 the disease prevalence based on a health survey or other studies.

173

174 *2.1. Global hypothesis test*

175 The PVs of each BDT depend on the same parameters, the sensitivity and the specificity of

176 the test and disease prevalence, and therefore they are parameters that depend on each other.

177 Consequently, the PVs of the two BDTs can be compared simultaneously. The global

178 hypothesis test to simultaneously compare the PVs of the two BDTs is

179
$$H_0 : PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2$$
$$H_1 : \text{at least one equality is not true,}$$

180 which is equivalent to the hypothesis test

181
$$H_0 : \mathbf{A}\boldsymbol{\omega} = \mathbf{0} \text{ vs } H_1 : \mathbf{A}\boldsymbol{\omega} \neq \mathbf{0}, \tag{12}$$

182 where $\mathbf{A}$ is a complete range design matrix and a dimension 2×4, i.e.

183
$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \end{pmatrix}.$$

184 As the vector $\hat{\boldsymbol{\omega}}$ is distributed asymptotically according to a multivariate normal distribution,

185 i.e. $\sqrt{n_1 + n_2} (\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow[n_1 + n_2 \to \infty]{} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$, then the statistic for the global hypothesis test (12)

186 is

187
$$Q^2 = \hat{\boldsymbol{\omega}}^T \mathbf{A}^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}}, \tag{13}$$

188 which is distributed asymptotically according to Hotelling's *T*-squared distribution with a

189 dimension 2 and $n_1 + n_2$ degrees of freedom, where 2 is the dimension of the vector $\mathbf{A}\hat{\boldsymbol{\omega}}$.

190 When $n_1 + n_2$ is large, the statistic Q^2 is distributed according to a central chi-square

191 distribution with 2 degrees of freedom when the null hypothesis is true.

192

9

193   *2.2. Individual hypothesis tests*

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

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

196   where PV is PPV or NPV. Based on the asymptotic normality of the estimators, the statistic

197   for this hypothesis test is

198    $$z = \frac{\left| \hat{P}V_1 - \hat{P}V_2 \right|}{\sqrt{\hat{V}ar\left(\hat{P}V_1\right) + \hat{V}ar\left(\hat{P}V_2\right) - 2\hat{C}ov\left(\hat{P}V_1, \hat{P}V_2\right)}}, \tag{14}$$

199   which is distributed asymptotically according to a normal standard distribution, and where the

200   variances-covariances is obtained from the equation (11).

201

202   *2.3. Alternative methods to the global test*

203   The global hypothesis test (12) simultaneously compares the PPVs and the NPVs of the two

204   BDTs. Some alternative methods to this global hypothesis test, based on the individual

205   hypothesis tests, are: 1) Testing the hypotheses $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$,

206   each one to an error α; 2) Testing the hypotheses $H_0 : PPV_1 = PPV_2$ and $H_0 : NPV_1 = NPV_2$,

207   and applying a multiple comparison method such as Bonferroni's method (1936) or Holm's

208   method (1979), which are methods that are very easy to apply based on the p-values.

209   Bonferroni's method consists of solving each individual hypothesis test to an error $\alpha/2$; and

210   Holm's method is a step-down method which is based on Bonferroni's method but is more

211   conservative. In Appendix B, Holm's method is summarized.

212

213   **3. Simulation experiments**

214   Simulation experiments were carried out to study the type I errors and the powers of the four

215   methods proposed to solve the global hypothesis test: the hypothesis test based on the chi-

10

216 square (equation (13)), the individual hypothesis tests each one to an error α, and the

217 individual hypothesis tests applying Bonferroni's method and Holm's method. We have also

218 studied the effect of a misspecification of the prevalence on the asymptotic behaviour of the

219 global hypothesis test and on the estimators of the PVs.

220 The experiments were designed setting the values of the PVs. For each BDT, we took as

221 PVs the values $\{0.70, 0.75, ..., 0.90, 0.95\}$, and as disease prevalence we took the values 10%,

222 25% and 50%. Based on the PVs and the prevalence, the Se and the Sp of each BDT were

223 calculated from the equations (1) and (2), only considering those cases in which the solutions

224 are between 0 and 1. As values of the correlation coefficients ρ^+ and ρ^- we took low values

225 (25% of the maximum value), intermediate ones (50% of the maximum value) and high ones

226 (75% of the maximum value), and the maximum value of each correlation coefficient is:

227
$$\max\left(\rho^+\right) = \frac{\min\left\{Se_1\left(1-Se_2\right), \left(1-Se_1\right)Se_2\right\}}{\sqrt{Se_1\left(1-Se_1\right)Se_2\left(1-Se_2\right)}} \quad \text{and} \quad \max\left(\rho^-\right) = \frac{\min\left\{Sp_1\left(1-Sp_2\right), \left(1-Sp_1\right)Sp_2\right\}}{\sqrt{Sp_1\left(1-Sp_1\right)Sp_2\left(1-Sp_2\right)}}.$$

228 As sample sizes, we took the values $n_i = \{50, 75, 100, 200, 500\}$. The simulation experiments

229 were carried out with R, using the bindata package to generate the samples of each type I

230 bivariate binomial distribution.

231 Regarding the random samples, these were generated in the following way. Firstly, once

232 the values of the PVs and of the prevalence were set, we calculated the sensitivities and the

233 specificities and maximum values of the coefficients ρ^+ and ρ^-. We then generated 10,000

234 samples with a type I bivariate binomial distribution with a sample size n_1, probabilities Se_1

235 and Se_2 and correlation coefficient ρ^+, and another 10,000 samples with a type I bivariate

236 binomial distribution with a sample size n_0, probabilities Sp_1 and Sp_2 and correlation

237 coefficient ρ^-. In this way, we obtained the marginal frequencies $n_{11\square}$ and $n_{1\square}$ ($n_{20\square}$ and $n_{2\square0}$)

238 of each one of the 10, 000 case (control) samples. The rest of the marginal frequencies were

11

239   easily calculated: $n_{10\square} = n_1 - n_{11\square}$, $n_{1\square 0} = n_1 - n_{1\square 1}$, $n_{21\square} = n_2 - n_{20\square}$ and $n_{2\square 1} = n_2 - n_{2\square 0}$. Then and in

240   order to construct the 2×2 table of each case simple, we generated a random valor n_{111} of a

241   doubly truncated binomial distribution of parameters n_1 and $\xi_{111} = Se_1 Se_2 + \varepsilon^+$ with

242   $n_{11\square} + n_{1\square 1} - n_1 \leq n_{111} \leq \text{Min}\{n_{11\square}, n_{11\square}\}$. This is necessary so that the sum of the frequencies leads

243   to the marginal totals randomly generated through the type I bivariate binomial distribution.

244   In the same way, in order to construct the 2×2 table of each control sample, we generated a

245   random value n_{200} of a doubly truncated binomial distribution of parameters n_2 and

246   $\xi_{200} = Sp_1 Sp_2 + \varepsilon^-$ with $n_{20\square} + n_{2\square 0} - n_2 \leq n_{200} \leq \text{Min}\{n_{20\square}, n_{2\square 0}\}$. For each one of the 10,000 case

247   (control) samples, once we have generated the values $n_{11\square}$, $n_{1\square 1}$ and n_{111} ($n_{20\square}$, $n_{2\square 0}$ and n_{200})

248   it is easy to construct the complete 2×2 table. Thus, $n_{110} = n_1 - n_{11\square}$, $n_{101} = n_{1\square 1} - n_{111}$ and

249   $n_{100} = n_{10\square} - n_{101}$ for the case samples, and $n_{201} = n_{20\square} - n_{200}$, $n_{210} = n_{2\square 0} - n_{200}$ and

250   $n_{211} = n_{21\square} - n_{210}$. For the experiments, the error $\alpha = 5\%$ was set. Moreover, all of the samples

251   were generated in such a way that in all of them the parameters and the variances-covariances

252   can be estimated.

253

254   *3.1. Type I errors and powers*

255   In Tables 2 and 3, we can see some results obtained for the type I errors of the global test and

256   of the alternative methods proposed in Section 2.3. In these tables, we can only see the results

257   for the global test, the individual comparisons with $\alpha = 5\%$ and with Bonferroni's method.

258   The results obtained with Holm's method are not shown as they are practically the same as

259   those obtained with Bonferroni's method. From the results obtained we can draw the

260   following conclusions. In general terms, the type I error of the global hypothesis test

261   fluctuates around the error $\alpha = 5\%$, especially in the case of samples sized $n_i \geq 100$,

12

262 depending on the prevalence and the correlations between the two BDTs. For samples with

263 smaller sizes $\left(n_i \leq 75\right)$, the type I error of the global test is lower than the error $\alpha = 5\%$. The

264 correlations between the two BDTs have an important effect on the type I error of the global

265 test, with a decrease in the type I error when there is an increase in the correlation

266 coefficients. Regarding the method based on the individual hypothesis tests $H_0 : PPV_1 = PPV_2$

267 and $H_0 : NPV_1 = NPV_2$ to an error $\alpha = 5\%$ each one of them, the type I error may clearly

268 overwhelm the nominal error (a situation that we have considered when the type I error is

269 greater than 6.5%), especially when the correlations are not high. Consequently, this method

270 may lead to erroneous results (false significances) and, therefore, should not be used. As for

271 solving the global test from the individual tests applying Bonferroni's (Holm's) method, the

272 type I error has a very similar behaviour to that of the global hypothesis test.

273

274

275

276

277

278

279

280

281

282

283

284

285

286

13

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

$Se_1 = 0.5385$, $Sp_1 = 0.9744$, $Se_2 = 0.5385$, $Sp_2 = 0.9744$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 10\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
n_1	n_2	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

$Se_1 = 0.8615$, $Sp_1 = 0.8769$, $Se_2 = 0.8615$, $Sp_2 = 0.8769$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 25\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
n_1	n_2	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

$Se_1 = 0.9692$, $Sp_1 = 0.5846$, $Se_2 = 0.9692$, $Sp_2 = 0.5846$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 50\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
n_1	n_2	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

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

289    $\alpha = 5\%$: Individual hypothesis tests each one to an error $\alpha = 5\%$.

290   Bonf.: Bonferroni's method.

291

292

293

294

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

$Se_1 = 0.5312$, $Sp_1 = 0.9896$, $Se_2 = 0.5312$, $Sp_2 = 0.9896$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 10\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
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

$Se_1 = 0.85$, $Sp_1 = 0.95$, $Se_2 = 0.85$, $Sp_2 = 0.95$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 25\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
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

$Se_1 = 0.9562$, $Sp_1 = 0.8312$, $Se_2 = 0.9562$, $Sp_2 = 0.8312$

$0 \leq \rho^+ \leq 1$, $0 \leq \rho^- \leq 1$

$p = 50\%$

		$\rho^+ = 0.25$ $\rho^- = 0.25$			$\rho^+ = 0.50$ $\rho^- = 0.50$			$\rho^+ = 0.75$ $\rho^- = 0.75$		
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.

15

302   Regarding the power of the hypothesis tests, in Tables 4 and 5 we can see some of the

303   results obtained for the global test and other alternative methods (Section 2.3). Neither can we

304   see in these Tables the results obtained applying Holm's method as they are practically the

305   same as those obtained with Bonferroni's method. From the results, the following conclusions

306   are obtained. The disease prevalence has an important effect on the power of each one of the

307   methods to solve the global test, and the power increases with an increase in the prevalence.

308   Regarding the correlations ρ^+ and ρ^-, these do not have a clear effect on the power, and the

309   power increases sometimes and decreases other times when the correlations increase. In very

310   general terms, when the prevalence is relatively small $\left(p = 10\% \right)$ we need large samples

311     $\left(n_i > 500 \right)$ so that the power of the global hypothesis test (equation (13)) is greater than 80%;

312   for a prevalence of 25% with sample sizes $n_i \geq 200$ we obtain a power greater than 80%; and

313   for a very large prevalence $\left(p = 50\% \right)$ with sample sizes $n_i \geq 50$ we obtain a very higher

314   power, greater than 80%-90%, depending on the difference between the PVs. The power of

315   the method based on the individual hypothesis tests to an error $\alpha = 5\%$ is greater than that of

316   the global test based on the chi-square distribution due to the fact that its type I error is also

317   greater. Regarding the hypothesis tests based on the individual tests with Bonferroni's method

318   and Holm's method, their corresponding power is practically the same, and is also very

319   similar to the power of the global test based on the chi-square distribution.

320

321

322

323

324

325

16

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

$Se_1 = 0.5357$, $Sp_1 = 0.9802$, $Se_2 = 0.5455$, $Sp_2 = 0.9596$

$0 \le \rho^+ \le 0.9805$, $0 \le \rho^- \le 0.6933$

$p = 10\%$

		$\rho^+ = 0.25$ $\rho^- = 0.17$			$\rho^+ = 0.49$ $\rho^- = 0.35$			$\rho^+ = 0.74$ $\rho^- = 0.52$		
n_1	n_2	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

$Se_1 = 0.8571$, $Sp_1 = 0.9048$, $Se_2 = 0.8727$, $Sp_2 = 0.8061$

$0 \le \rho^+ \le 0.9354$, $0 \le \rho^- \le 0.6614$

$p = 25\%$

		$\rho^+ = 0.23$ $\rho^- = 0.17$			$\rho^+ = 0.47$ $\rho^- = 0.33$			$\rho^+ = 0.70$ $\rho^- = 0.50$		
n_1	n_2	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

$Se_1 = 0.9643$, $Sp_1 = 0.6786$, $Se_2 = 0.9818$, $Sp_2 = 0.3455$

$0 \le \rho^+ \le 0.7071$, $0 \le \rho^- \le 0.5$

$p = 50\%$

		$\rho^+ = 0.18$ $\rho^- = 0.13$			$\rho^+ = 0.35$ $\rho^- = 0.25$			$\rho^+ = 0.53$ $\rho^- = 0.38$		
n_1	n_2	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

327 Global: Global hypothesis test based on the chi-square distribution.
328 $\alpha = 5\%$: Individual hypothesis tests each one to an error $\alpha = 5\%$.
329 Bonf.: Bonferroni's method.
330

331

332

333

17

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

$Se_1 = 0.5278$, $Sp_1 = 0.9969$, $Se_2 = 0.5357$, $Sp_2 = 0.9802$

$0 \le \rho^+ \le 0.9841$, $0 \le \rho^- \le 0.3910$

$p = 10\%$

		$\rho^+ = 0.25$ $\rho^- = 0.10$			$\rho^+ = 0.49$ $\rho^- = 0.19$			$\rho^+ = 0.74$ $\rho^- = 0.29$		
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

$Se_1 = 0.8444$, $Sp_1 = 0.9852$, $Se_2 = 0.8571$, $Sp_2 = 0.9048$

$0 \le \rho^+ \le 0.9511$, $0 \le \rho^- \le 0.3779$

$p = 25\%$

		$\rho^+ = 0.24$ $\rho^- = 0.09$			$\rho^+ = 0.48$ $\rho^- = 0.19$			$\rho^+ = 0.71$ $\rho^- = 0.28$		
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

$Se_1 = 0.95$, $Sp_1 = 0.95$, $Se_2 = 0.9643$, $Sp_2 = 0.6786$

$0 \le \rho^+ \le 0.8388$, $0 \le \rho^- \le 0.3333$

$p = 50\%$

		$\rho^+ = 0.21$ $\rho^- = 0.08$			$\rho^+ = 0.42$ $\rho^- = 0.17$			$\rho^+ = 0.63$ $\rho^- = 0.25$		
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

335   Global: Global hypothesis test based on the chi-square distribution.
336   $\alpha = 5\%$: Individual hypothesis tests each one to an error $\alpha = 5\%$.
337   Bonf.: Bonferroni's method.
338

339   As conclusions of the results obtained in the simulation experiments, the global hypothesis

340   test based on the chi-square distribution behaves well in terms of the type I error (it does not

341   overwhelm the nominal error of 5%), the same as the individual tests along with Bonferroni's

18

342 method or Holm's method. The method based on the individual tests to a global error $\alpha = 5\%$

343 should not be used as it may clearly overwhelm the nominal error.

344 From the results obtained, we propose the following method to compare the PVs of two

345 BDTs subject to a case-control sampling: 1) Applying the hypothesis test based on the chi-

346 square distribution (equation (13)) to an error α, 2) If the global hypothesis test is not

347 significant, the equality hypothesis of the PVs is not rejected; if the global hypothesis test is

348 significant to an error α, the investigation of the causes of the significance is made by testing

349 the individual tests (equation (14)) and applying Bonferroni's method or Holm's method to an

350 error α.

351

352 *3.2. Effect of the prevalence*

353 The estimation and comparison of the PVs of two BDTs subject to a case-control sampling

354 requires knowledge of the disease prevalence, of an estimation of the disease prevalence

355 obtained from another study, e.g. a health survey. To study the effect of a misspecification of

356 the prevalence in the comparison of the PVs of two BDTs and in the estimators of the PVs,

357 we carried out simulation experiments similar to those made to study the type I errors and the

358 powers. For this purpose, we took as the prevalence for the inference an overestimation (and

359 an underestimation) equal to 5% and to 10% of the value of the prevalence set, and we have

360 studied the type I errors and the powers of the global test and of the Bonferroni and Holm

361 methods and the relative root mean square error (*RRMSE*) of the estimator of each PVs. Thus,

362 for each estimator we calculated the relative root mean square error (RRMSE) defined as

363
$$RRMSE\left(\hat{P}V_i\right) = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\hat{P}V_{ik} - PV_i\right)^2}}{PV_i},$$

19

364   where PV_i is the PPV or the NPV of the ith BDT $(i=1,2)$ and $\hat{P}V_{ik}$ is its estimator

365   calculated from the kth sample $(k=1,...,N)$, and $N=10,000$. For the values of the

366   parameters we took as prevalences $p=\{10\%,25\%,50\%\}$ respectively, and to estimate the

367   PVs we took as prevalences $p'=p\pm d\times p$ with $d=\{5\%,10\%\}$.

368   In Table 6, we show some of the results obtained for the type I errors and the powers of the

369   global test and the Bonferroni method (the results of the Holms method are not shown as they

370   are practically identically to those obtained with the Bonferroni method). In the Table we

371   show the results when there is no misspecification of the prevalence $(p'=p)$ and when the

372   prevalence is underestimated $(p'<p)$ and overestimated $(p'>p)$. From the results of these

373   experiments, it is verified that the type I errors of the methods studied do not overwhelm the

374   nominal error $\alpha=5\%$, and in general terms there are no important differences between the

375   type I errors when there is a misspecification of the prevalence and when there is not.

376   Regarding the powers, the conclusions are also very similar: there are no important

377   differences between the powers when there is a misspecification of the prevalence and when

378   there is not. Regarding the estimators, in Table 6 we show some of the results obtained for the

379   RRMSEs (in %) of the estimators of the PVs of the two BDTs. There is no important

380   difference between the RRMSEs when there is a misspecification of the prevalence ($p'<p$

381   or $p'>p$) and the RRMSEs when there is no misspecification of the prevalence $(p'=p)$. In

382   general terms, this difference is not usually over 5% when the samples are small, and this is

383   even lower when the samples are large. Consequently, misspecifications (5% or 10%) of the

384   disease prevalence do not have any important effect on the type I errors and on the powers of

385   the global hypothesis test and on the alternative methods (Bonferroni and Holm), and nor do

386   they have an important effect on the estimators of the PVs.

387

388

Table 6. Effect of a misspecification of the prevalence.

Type I errors

$PPV_1 = PPV_2 = 0.90$, $NPV_1 = NPV_2 = 0.80$

$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.2571$, $Sp_2 = 0.9905$, $\rho^+ = 0.75$, $\rho^- = 0.75$, $p = 25\%$

		$p'=p=25\%$		$p'=22.50\%$		$p'=23.75\%$		$p'=26.25\%$		$p'=27.50\%$	
n_1	n_2	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

$PPV_1 = 0.90$, $PPV_2 = 0.70$, $NPV_1 = 0.80$, $NPV_2 = 0.90$

$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.70$, $Sp_2 = 0.90$, $\rho^+ = 0.29$, $\rho^- = 0.22$, $p = 25\%$

		$p'=p=25\%$		$p'=22.50\%$		$p'=23.75\%$		$p'=26.25\%$		$p'=27.50\%$	
n_1	n_2	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

$PPV_1 = 0.90$, $PPV_2 = 0.70$, $NPV_1 = 0.80$, $NPV_2 = 0.90$

$Se_1 = 0.2571$, $Sp_1 = 0.9905$, $Se_2 = 0.70$, $Sp_2 = 0.90$, $\rho^+ = 0.29$, $\rho^- = 0.22$, $p = 25\%$

		$p'=p=25\%$		$p'=22.50\%$		$p'=23.75\%$		$p'=26.25\%$		$p'=27.50\%$	
n_1	n_2	$\hat{P}PV_1$	$\hat{N}PV_1$	$\hat{P}PV_1$	$\hat{N}PV_1$	$\hat{P}PV_1$	$\hat{N}PV_1$	$\hat{P}PV_1$	$\hat{N}PV_1$	$\hat{P}PV_1$	$\hat{N}PV_1$
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

		$p'=p=25\%$		$p'=22.50\%$		$p'=23.75\%$		$p'=26.25\%$		$p'=27.50\%$	
n_1	n_2	$\hat{P}PV_2$	$\hat{N}PV_2$	$\hat{P}PV_2$	$\hat{N}PV_2$	$\hat{P}PV_2$	$\hat{N}PV_2$	$\hat{P}PV_2$	$\hat{N}PV_2$	$\hat{P}PV_2$	$\hat{N}PV_2$
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

389 Global: Global hypothesis test based on the chi-square distribution.
390 Bonf.: Bonferroni's method.
391

392

4. Example

The results obtained were applied to the study by Matovu et al (2010) on the diagnosis of Human African Trypanosomiasis (HAT) in Uganda. 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. In some rural areas of Africa, the disease prevalence may reach 50% in periods of epidemics, and is a significant cause of death. Matovu et al (2010) applied two diagnostic tests to a sample of 75 cases and another sample of 65 controls. 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: 10% and 50%. The first case $(p = 10\%)$ corresponds to a situation of low disease prevalence, and the second one $(p = 50\%)$ corresponds to a situation of a HAT epidemic. In Table 7, we can also see the estimations of the sensitivities and the specificities (and their standard errors, SE) of the BDTs and the correlations.

Table 7. Study by Matuvo et al.

Observed frequencies							
Case				**Control**			
	$T_2 = 1$	$T_2 = 0$	Total		$T_2 = 1$	$T_2 = 0$	Total
$T_1 = 1$	57	4	61	$T_1 = 1$	1	4	5
$T_1 = 0$	6	8	14	$T_1 = 0$	0	60	60
Total	63	12	75	Total	1	64	65

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	0.840 ± 0.042	0.538	0.923 ± 0.033	0.985 ± 0.015	0.433

22

412   For a prevalence value equal to 10%, the estimations of the PVs are $\hat{PPV}_1 = 0.540$,

413    $\hat{PPV}_2 = 0.858$, $\hat{NPV}_1 = 0.978$ and $\hat{NPV}_2 = 0.982$, and the estimated variance and covariance

414   matrix of the estimators of the PVs is

415
$$
\hat{\Sigma}_{\hat{\omega}} = \begin{pmatrix} 0.01158 & 0.00562 & 0.00015 & 0.00005 \\ 0.00562 & 0.01456 & 0.00006 & 0.00006 \\ 0.00015 & 0.00006 & 0.00003 & 0.00001 \\ 0.00005 & 0.00006 & 0.00001 & 0.00002 \end{pmatrix}.
$$

416   The value of the test statistic for the test

417
$$
H_0 : PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2
$$
$$
H_1 : \text{at least one equality is not true,}
$$

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

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

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

421   statistic for the test $H_0 : NPV_1 = NPV_2$ is equal to 0.886 $(\text{two sided p-value} = 0.375)$.

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

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

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

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

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

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

428   matrix of the estimators of the PVs is

429
$$
\hat{\Sigma}_{\hat{\omega}} = \begin{pmatrix} 0.00117 & 0.00026 & 0.00032 & 0.00010 \\ 0.00026 & 0.00031 & 0.00005 & 0.00006 \\ 0.00032 & 0.00005 & 0.00116 & 0.00058 \\ 0.00010 & 0.00006 & 0.00058 & 0.00102 \end{pmatrix}.
$$

430   The value of the test statistic for the test

23

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

431

432 is $Q^2 = 5.048$ $\left(P - value = 0.080\right)$, and therefore with an error $\alpha = 5\%$ we do not reject the

433 equality of the positive predictive values and the negative predictive values of both diagnostic

434 tests, although there are signs of significance and, therefore, an increase in the two sample

435 sizes may be necessary. Solving the individual hypothesis tests it is found that the value of the

436 test statistic for the test $H_0 : PPV_1 = PPV_2$ is equal to 2.21 $\left(\text{two sided p-value} = 0.027\right)$, and

437 that the value of the test statistic for the $H_0 : NPV_1 = NPV_2$ is equal to 0.892

438 $\left(\text{two sided p-value} = 0.373\right)$. If each individual hypothesis test is solved to an error $\alpha = 5\%$,

439 then $H_0 : PPV_1 = PPV_2$ is rejected and $H_0 : NPV_1 = NPV_2$ is not rejected, and the result is

440 contradictory to that obtained with the global test to an error $\alpha = 5\%$.

441

6. More than two BDTs

443 Let us consider that J BDTs $\left(J \geq 3\right)$ are applied to all of the individuals in the case sample

444 and the control sample. For each BDT we define the random variable T_j in a similar way to

445 how this was done in Section 2. Let Se_j and Sp_j be the sensitivity and the specificity of the

446 jth BDT, with con $j = 1, ..., J$. Let $n_{1 i_1 ... i_J}$ be the number of individuals with the disease for

447 whom $T_1 = i_1, ..., T_J = i_J$, with $i_j = 1$ when the result of the jth BDT is positive and $i_j = 0$

448 when it is negative. In a similar way, $n_{2 i_1 ... i_J}$ is the number of without the disease for whom

449 $T_1 = i_1, ..., T_J = i_J$. Let us consider the probabilities $\xi_{h i_1, ..., i_J} = P\left(T_1 = i_1, T_2 = i_2, ..., T_J = i_J\right)$, with

450 $h = 1, 2$. Thus, for example for three BDTs, using the dependence model of Torrance-Rynard

451 and Walter (1997), these probabilities are

24

452
$$\xi_{1i_1i_2i_3} = \prod_{j=1}^{3} Se_j^{i_j} \left(1 - Se_j\right)^{1-i_j} + \sum_{j,k,j<k}^{3} (-1)^{\left|i_j - i_k\right|} \varepsilon_{jk}^{+}$$

453   and

454
$$\xi_{2i_1i_2i_3} = \prod_{j=1}^{3} Sp_j^{1-i_j} \left(1 - Sp_j\right)^{i_j} + \sum_{j,k,j<k}^{3} (-1)^{\left|i_j - i_k\right|} \varepsilon_{jk}^{-},$$

455   with $i_j = 0,1$, $i_k = 0,1$ and $j,k = 1,2,3$, and where ε_{jk}^{+} $\left(\varepsilon_{jk}^{-}\right)$ is the covariance between the jth

456   BDT and the kth BDT for individuals with the disease (without the disease). The estimators of

457   these probabilities are $\hat{\xi}_{hi_1...i_J} = n_{hi_1...i_J}/n_h$, with $h = 1,2$. The sensitivity and the specificity of

458   the jth BDT are

459
$$Se_j = \sum_{\substack{i_1,...,i_J=0 \\ i_j=1}}^{1} \xi_{1i_1,...,i_J} \quad \text{and} \quad Sp_j = \sum_{\substack{i_1,...,i_J=0 \\ i_j=0}}^{1} \xi_{2i_1,...,i_J},$$

460   and its estimators are

461
$$\hat{Se}_j = \frac{\displaystyle\sum_{\substack{i_1,...,i_J=0 \\ i_j=1}}^{1} n_{1i_1,...,i_J}}{n_1} \quad \text{and} \quad \hat{Sp}_j = \frac{\displaystyle\sum_{\substack{i_1,...,i_J=0 \\ i_j=0}}^{1} n_{2i_1,...,i_J}}{n_2}.$$

462   The estimators of the variances-covariances of these estimators are

463    $\hat{Var}\left(\hat{Se}_j\right) = \hat{Se}_j\left(1 - \hat{Se}_j\right)/n_1$, $\hat{Var}\left(\hat{Sp}_j\right) = \hat{Sp}_j\left(1 - \hat{Sp}_j\right)/n_2$, $\hat{Cov}\left(\hat{Se}_j,\hat{Se}_k\right) == \hat{\varepsilon}_{jk}^{+}/n_1$ and

464    $\hat{Cov}\left(\hat{Sp}_j,\hat{Sp}_k\right) = \hat{\varepsilon}_{jk}^{-}/n_2$, and the rest of the covariances are equal to zero. If we know the

465   prevalence p (or and an estimation), the estimators of the PVs of the jth BDT are

466
$$\hat{PPV}_j = \frac{pn_2 \displaystyle\sum_{\substack{i_1,...,i_J=0 \\ i_j=1}}^{1} n_{1i_1,...,i_J}}{pn_2 \displaystyle\sum_{\substack{i_1,...,i_J=0 \\ i_j=1}}^{1} n_{1i_1,...,i_J} + qn_1 \displaystyle\sum_{\substack{i_1,...,i_J=0 \\ i_j=1}}^{1} n_{2i_1,...,i_J}}$$

467   and

25

468
$$\hat{NPV}_j = \frac{qn_1 \sum\limits_{\substack{i_1,\dots,i_J=0 \\ i_j=0}}^{1} n_{2i_1,\dots,i_J}}{qn_1 \sum\limits_{\substack{i_1,\dots,i_J=0 \\ i_j=0}}^{1} n_{2i_1,\dots,i_J} + pn_2 \sum\limits_{\substack{i_1,\dots,i_J=0 \\ i_j=0}}^{1} n_{1i_1,\dots,i_J}} .$$

469 Let $\boldsymbol{\theta} = \left(Se_1,\dots,Se_J,Sp_1,\dots,Sp_J \right)^T$ be the vector whose components are the sensitivities and the

470 specificities, and let $\boldsymbol{\omega} = \left(PPV_1,\dots,PPV_J,NPV_1,\dots,NPV_J \right)^T$ be the vector whose components

471 are the PVs. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, with a dimension $2J \times 2J$, is similar to

472 that given in expression (10), where $\sum_{\hat{Se}}$ and $\sum_{\hat{Sp}}$ are matrixes with a dimension $J \times J$.

473 Applying the delta method, the variance-covariance matrix of $\hat{\boldsymbol{\omega}}$, with a dimension $2J \times 2J$,

474 has an expression similar to that given in equation (11).

475 The PVs of each one of the J BDTs depend on the same parameters (the sensitivity and the

476 specificity of the jth diagnostic test) and, therefore, these parameters can be compared

477 simultaneously. The global hypothesis test to simultaneously compare the PVs of the J BDTs

478 is

479
$$H_0 : PPV_1 = PPV_2 = \dots = PPV_J \text{ and } NPV_1 = NPV_2 = \dots = NPV_J$$
$$H_1 : \text{at least one equality is not true,}$$

480 which is equivalent to the hypothesis test

481
$$H_0 : \mathbf{A}\boldsymbol{\omega} = 0 \text{ vs } H_1 : \mathbf{A}\boldsymbol{\omega} \neq 0$$

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

483
$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_0 \\ \mathbf{A}_0 & \mathbf{A}_1 \end{pmatrix},$$

484 where $\mathbf{A}_0$ is a matrix with a dimension $(J-1) \times J$ whose elements are all equal to 0, and $\mathbf{A}_1$

485 is a matrix with a dimension $(J-1) \times J$ where each component (i,i) is equal to 1, each

486 element $(i,i+1)$ is equal to -1 for $i = 1,\dots,J-1$, and the rest of the elements in this matrix

26

487   are equal to 0. Applying the multivariate central limit theorem it is verified that

488    $\sqrt{n_1 + n_2} \left(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega} \right) \xrightarrow[n_1+n_2 \to \infty]{} N_{2J} \left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \right)$. Then, the statistic $Q^2 = \left(\hat{\boldsymbol{\omega}} \mathbf{A} \right)^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}}$ is

489   distributed according to Hotelling's T-squared distribution with a dimension $2(J-1)$ and

490    $n_1 + n_2$ degrees of freedom, where $2(J-1)$ is the dimension of the vector $\mathbf{A} \hat{\boldsymbol{\omega}}$. When $n_1 + n_2$

491   is large, the statistic Q^2 is distributed according to a central chi-squared distribution with

492    $2(J-1)$ degrees of freedom when the null hypothesis is true, i.e.

493
$$Q^2 = \left(\hat{\boldsymbol{\omega}} \mathbf{A} \right)^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}} \xrightarrow[n \to \infty]{} \chi^2_{2(J-1)}.$$

494   Finally, the method to compare the PVs of the J BDTs would consist of the following steps:

495   1) Solve the global hypothesis test to an error α calculating the statistic

496    $Q^2 = \left(\hat{\boldsymbol{\omega}} \mathbf{A} \right)^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}}$ based on the chi-squared distribution; 2) if the global test is not

497   significant to an error α then we do not reject the homogeneity of the J PVs, but if the

498   hypothesis test is significant then the causes of significance are investigated comparing the

499   PPVs (NPVs) in pairs (equation (14)) and applying an adjustment method of the p-value

500   based on multiple comparisons (e.g. Bonferroni or Holm).

501

7. Discussion

502

503   The comparison of the positive and negative predictive values of two binary diagnostic tests is

504   an important topic in the study of Statistical Methods in Diagnostic Medicine. Subject to a

505   cross-sectional sampling, this topic has been subject to different studies. In this article we

506   studied the simultaneous comparison of the predictive values of two diagnostic tests subject to

507   a case-control sampling, analysing and comparing several methods. These methods consisted

508   of a global test based on the chi-square distribution, a method based on the individual

509   comparisons each one to a nominal error α, and another three methods based on individual

510   comparisons along with a multiple comparison method. The multiple comparison methods

511   that were used were Bonferroni's method and Holm's method, which are methods based on

512   the p-values of the individual hypothesis tests and are very easy to apply.

513   Simulation experiments were carried out to study the type I errors and the power of the

514   four methods proposed. These experiments were based on the generation samples with type I

515   bivariate binomial distributions, which are the distributions that are inherent to case-control

516   design, since from these samples proportions of marginal totals are estimated. The results

517   have shown that the global hypothesis test based on the chi-square distribution behaves well

518   in terms of type I error, and does not overwhelm the nominal error $\alpha = 5\%$. Regarding its

519   power, in general this strongly depends on the disease prevalence, and it is necessary to have

520   very large samples $(n_i > 500)$ when the prevalence is small and relatively small sample sizes

521    $(n_i \geq 50)$ when the prevalence is high, so that the power will be high. The simulation

522   experiments also showed that the methods based on individual hypothesis tests along with

523   multiple comparison methods have type I errors and very similar power to those of the global

524   test based on the chi-square distribution. Consequently, both methods can be used to compare

525   the PVs of the two BDTs. Furthermore, the experiments also showed that the comparison of

526   the predictive values of two diagnostic tests cannot be made independently i.e. comparing the

527   two positive predictive values and comparing the two negative predictive values

528   independently to an error $\alpha = 5\%$, as it is possible to obtain a type I error that clearly

529   overwhelms the nominal error set. Based on the results of the simulation experiments, a

530   method has been proposed to compare the predictive values of two diagnostic tests subject to

531   a case-control sampling. This method, which is similar to that proposed by Roldán-Nofuentes

532   et al (2012), consists of the following steps: 1) Simultaneously comparing the predictive

533   values applying the global hypothesis test based on the chi-square distribution (equation (13))

534   to an error α; 2) If the global hypothesis test is not significant, then the equality hypothesis of

28

535   the PVs is not rejected. If the global hypothesis test is significant to an error α, then the

536   causes of the significance are studied solving the individual hypothesis tests (equation (14))

537   and applying Bonferroni's method or Holm's method to an error α. This procedure that we

538   propose is similar to the Analysis of Variance: firstly, the global test is solved and, if this is

539   significant, then the causes of the significance are studied starting with paired comparisons

540   along with some multiple comparison method.

541   Simulation experiments were carried out to study the effect of a misspecification of the

542   prevalence in the asymptotic behaviour of the global hypothesis test based on the chi-square

543   distribution and on the methods based on multiple comparisons. From the results obtained, we

544   can conclude that light or moderate overestimations or underestimations of the prevalence do

545   not have an important effect on the behaviour of these hypothesis tests.

546   The proposed model has been applied to a real example on the diagnosis of the Human

547   African Trypanosomiasis (HAT) in Uganda, disease that is a major public health problem in

548   some African countries, and whose correct diagnosis is essential for a proper treatment. The

549   results obtained have shown that, when the prevalence is small, the positive predictive value

550   of the NASBA-OC test is significantly greater than that of the PCR-OC test, and there are no

551   significant differences between the negative predictive values of both diagnostic tests.

552   Therefore, when the prevalence of HAT is small, the NASBA-OC test is a better test than the

553   PCR-OC test to confirm the presence of the HAT. When the prevalence of HAT is very high,

554   the equality of the predictive values has not been rejected (although an increase of the two

555   samples may be convenient), and therefore it is not rejected that both diagnostic tests are

556   equally valid to confirm and to exclude the presence of the HAT.

557   Finally, the global hypothesis test was extended to the situation in which we

558   simultaneously compare the PVs of more than two BDTs, and for this we propose a method

559   which is similar to that proposed for two BDTs. To be able to calculate the global test

29

560   statistic, $Q^2 = \hat{\boldsymbol{\omega}}^T \mathbf{A}^T \left(\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T \right)^{-1} \mathbf{A} \hat{\boldsymbol{\omega}}$, it is necessary for the matrix $\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T$ to be non-

561   singular. For two BDTs, the matrix $\mathbf{A} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\omega}}} \mathbf{A}^T$ is non-singular when it is verified that

562    $n_{110} + n_{101} > 0$ and that $n_{210} + n_{201} > 0$; therefore, if $n_{110} = n_{101} = 0$ and $n_{210} = n_{201} = 0$ then the

563   method proposed to compare the PVs cannot be applied.

564

Acknowledgements

568

References

570   Bennett BM. On comparison of sensitivity, specificity and predictive value of a number of

571   diagnostic procedures. Biometrics 1972;28:793-800.

572   Bennett BM. On tests for equality of predictive values for t diagnostic procedures. Statistics in

573   Medicine 1985;4:535-9.

574   Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R

575   Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936;8:3-62.

576   Holm S. A simple sequential rejective multiple testing procedure. Scandinavian Journal of

577   Statistics 1979;6:65-70.

578   Kocherlakota S, Kocherlakota, K. Bivariate discrete distributions. New York: Marcel Dekker

579   INC; 1992.

580   Kosinski AS. A weighted generalized score statistic for comparison of predictive values of

581   diagnostic tests. Statistics in Medicine 2013;32:964-977.

582   Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical

583   diagnostic tests for paired designs. Biometrics 2000;56:345-351.

584   Matuvo E, Mugasa CM, Ekangu RA, Deborggrave S, Lubega GW, Laurent T, Schoone GJ,

585   Schallig HD, Büscher P. Phase II evaluation of sensitivity and specificity of PCR and NASBA

586   followed by oligochromatography for diagnosis of human African trypanosomiasis in clinical

587   samples from D.R. Congo and Uganda. PLOS Neglected Tropical Diseases2010;4: e737.

588   Mercaldo ND, Kit FL, Zhou XH. Confidence intervals for predictive values with an emphasis to

589   case-control studies. Statistics in Medicine 2007;26:2170-2183.

590   Moskowitz CS, Pepe MS. Comparing the predictive values of diagnostic tests: sample size

591   and analysis for paired study designs. Clinical Trials 2006;3:272-9.

592   Pepe MS (2003). The statistical evaluation of medical tests for classification and prediction. New

593   York: Oxford University Press; 2003.

594   Roldán-Nofuentes JA, Luna del Castillo JD, Montero-Alonso MA. Global hypothesis test to

595   simultaneously compare the predictive values of two binary diagnostic tests. Computational

596   Statistics and Data Analysis, Special issue "Computational Statistics for Clinical Research"

597   2012;56:1161-1173.

598   Torrance-Rynard VL, Walter SD. Effects of dependent errors in the assessment of diagnostic

599   test performance. Statistics in Medicine 1997;16:2157-2175.

600   Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests.

601   Biometrics 1985;41:959-968.

602   Wang W, Davis CS, Soong SJ. Comparison of predictive values of two diagnostic tests from

603   the same sample of subjects using weighted least squares. Statistics in Medicine

604   2006;25:2215-2229.

605

606

607

608

Appendix A

Performing algebraic operations in equation (11) it is found that:

$$Var\left(PPV_1\right)=\left(\frac{p^2Se_1+pQ_1}{Q_1^2}\right)^2 Var\left(Se_1\right)+\left(\frac{pqSe_1}{Q_1^2}\right)^2 Var\left(Sp_1\right),$$

$$Var\left(PPV_2\right)=\left(\frac{p^2Se_2+pQ_2}{Q_2^2}\right)^2 Var\left(Se_2\right)+\left(\frac{pqSe_2}{Q_2^2}\right)^2 Var\left(Sp_2\right),$$

$$Var\left(NPV_1\right)=\left(\frac{pqSp_1}{\left(1-Q_1\right)^2}\right)^2 Var\left(Se_1\right)+\left(\frac{q\left(1-Q_1\right)-q^2Sp_1}{\left(1-Q_1\right)^2}\right)^2 Var\left(Sp_1\right),$$

$$Var\left(NPV_2\right)=\left(\frac{pqSp_2}{\left(1-Q_2\right)^2}\right)^2 Var\left(Se_2\right)+\left(\frac{q\left(1-Q_2\right)-q^2Sp_2}{\left(1-Q_2\right)^2}\right)^2 Var\left(Sp_2\right),$$

$$Cov\left(PPV_1,PPV_2\right)=\left(\frac{pQ_1-p^2Se_1}{Q_1^2}\right)\left(\frac{pQ_2-p^2Se_2}{Q_2^2}\right)Cov\left(Se_1,Se_2\right)+\frac{p^2q^2Se_1Se_2}{Q_1^2Q_2^2}Cov\left(Sp_1,Sp_2\right),$$

$$Cov\left(PPV_1,NPV_1\right)=$$
$$\frac{pq}{Q_1^2\left(1-Q_1\right)^2}\left[\left(pQ_1-p^2Se_1\right)Sp_1Var\left(Se_1\right)+\left\{q\left(1-Q_1\right)-q^2Sp_1\right\}Se_1Var\left(Sp_1\right)\right],$$

$$Cov\left(PPV_1,NPV_2\right)=$$
$$\frac{pq\left(pQ_1-p^2Se_1\right)}{Q_1^2\left(1-Q_2\right)^2}Sp_2Cov\left(Se_1,Se_2\right)+\frac{pq\left\{q\left(1-Q_2\right)-q^2Sp_2\right\}}{Q_1^2\left(1-Q_2\right)^2}Se_1Cov\left(Sp_1,Sp_2\right),$$

$$Cov\left(PPV_2,NPV_1\right)=$$
$$\frac{pq}{Q_2^2\left(1-Q_1\right)^2}\left[\left(pQ_2-p^2Se_2\right)Sp_1Cov\left(Se_1,Se_2\right)+\left\{q\left(1-Q_1\right)-q^2Sp_1\right\}Se_2Cov\left(Sp_1,Sp_2\right)\right],$$

$$Cov\left(PPV_2,NPV_2\right)=$$
$$\frac{pq}{Q_2^2\left(1-Q_2\right)^2}\left[\left(pQ_2-p^2Se_2\right)Sp_2Var\left(Se_2\right)+\left\{q\left(1-Q_2\right)-q^2Sp_2\right\}Se_2Var\left(Sp_2\right)\right]$$

and

$$Cov\left(NPV_1, NPV_2\right) =$$

621
$$\frac{p^2 q^2 Sp_1 Sp_2}{\left(1-Q_1\right)^2 \left(1-Q_2\right)^2} Cov\left(Se_1, Se_2\right) + \left(\frac{q\left(1-Q_1\right)-q^2 Sp_1}{\left(1-Q_1\right)^2}\right)\left(\frac{q\left(1-Q_2\right)-q^2 Sp_2}{\left(1-Q_2\right)^2}\right) Cov\left(Sp_1, Sp_2\right),$$

622   where $q = 1 - p$ and $Q_i = P\left(T_i = 1\right) = p \times Se_i + q \times \left(1 - Sp_i\right)$.

623

624   **Appendix B**

625   Let us assume that we are going to solve K hypothesis test H_{0k} vs H_{1k} with $k = 1,...,K$. Let

626    $p_{[1]} \leq p_{[2]} \leq ... \leq p_{[K]}$ be the *p-values* obtained ordered from the lowest to the highest, and

627   therefore $p_{[k]}$ is the *p-value* that corresponds to the hypothesis test $H_{0[k]}$ vs $H_{1[k]}$. Holm's

628   method [12] consists of the following steps:

629   Step 1. If $p_{[1]} \leq \alpha/K$ hypothesis $H_{0[1]}$ is rejected and we go to the next step; if $p_{[1]} > \alpha/K$

630   no null hypothesis is rejected and the process finishes.

631   Step 2. If $p_{[2]} \leq \alpha/\left(K-1\right)$ hypothesis $H_{0[2]}$ is rejected and we go to the next step; if

632    $p_{[2]} > \alpha/\left(K-1\right)$ we do not reject the null hypotheses $H_{0[k]}$ with $k = 2,...,K$ and the process

633   finishes....

634   Step K. If $p_{[K]} \leq \alpha$ hypothesis $H_{0[K]}$ is rejected and the process finishes; and if $p_{[K]} > \alpha$

635    $H_{0[K]}$ is not rejected and the process finishes.