Revista Colombiana de Estadística - Applied Statistics
July 2025, volume 48, issue 2, pp. 227–246
http://doi.org/10.15446/rce.v48n2.115786

Steps for Diagnostic Precision Meta-Analysis in Binary Response Studies: A Case of the Application by Hierarchical Modeling

Pasos para el metaanálisis de precisión diagnóstica en estudios de respuesta binaria: un caso de aplicación mediante modelización jerárquica

Johny J. Pambabay-Calero[1,2,a], Sergio A. Bauz-Olvera[1,2,b], Omar H. Ruiz-Barzola[2,3,c], M. Purificacion Galindo-Villardon[2,4,5,6,d]

[1] Faculty of Natural Sciences and Mathematics, ESPOL-Polytechnic University, Guayaquil, Ecuador

[2] Centro de Estudios e Investigaciones Estadisticas, ESPOL-Polytechnic University, Guayaquil, Ecuador

[3] Facultad de Ciencias de la Vida, ESPOL-Polytechnic University, Guayaquil, Ecuador

[4] Department of Statistics, University of Salamanca, Salamanca, Spain

[5] Instituto de Investigación Biomédica de Salamanca (IBSAL), University of Salamanca, Salamanca, Spain

[6] Centro de Estudios Estadísticos, Universidad Estatal de Milagro (UNEMI), Milagro, Ecuador

Abstract

The discriminatory capacity of a test is commonly expressed in terms of sensitivity and specificity, and there is generally a compromise relationship between these two measures, as an increasing threshold for defining the positivity of the test results in a decrease in sensitivity and an increase in specificity. Recommended methods for the meta-analysis of diagnostic tests, such as the bivariate model, focus on estimating a summary sensitivity and specificity at a common threshold, while the Hierarchical Summary Receiver Operating Characteristic (HSROC) model focuses on estimating a summary curve from studies that have used different thresholds. Therefore, we will explain the hierarchical modeling for meta-analysis of the study of precision in diagnostic tests, and we will design a decision scheme that helps

[a] Ph.D. E-mail: jpambaba@espol.edu.ec
[b] Ph.D. E-mail: serabauz@espol.edu.ec
[c] Ph.D. E-mail: oruiz@espol.edu.ec
[d] Ph.D. E-mail: pgalindo@usal.es

to understand the models to choose the most appropriate in situations of heterogeneity in the studies, and illustrate its application in a systematic review, studying the properties and assumptions of meta-analytic procedures to synthesize the quantitative evidence of the parameters. For which, we used a systematic review that obtained summary estimates for the diagnosis of invasive aspergillosis, being our modeling framework the NLMIXED procedure of SAS, obtaining summary estimates for sensitivity and specificity for the bivariate model of 0.7708 and 0.8521, from a total of 27 studies involving 3,943 patients, and for the HSROC case the values were 0.7304 and 0.8867 respectively. Finally, we hope that this article will provide clinicians with a sufficient understanding of the terminology and statistical methods, obtaining plausible interpretations of the results in systematic reviews.

Key words**:** Bivariate model; Diagnostic accuracy; Heterogeneity; HSROC; Meta-analysis.

Resumen

La capacidad discriminatoria de una prueba se expresa comúnmente en términos de sensibilidad y especificidad, y generalmente existe una relación de compromiso entre estas dos medidas, ya que un umbral creciente para definir la positividad de la prueba resulta en una disminución de la sensibilidad y un aumento de la especificidad. Los métodos recomendados para el metaanálisis de pruebas diagnósticas, como el modelo bivariado, se centran en estimar una sensibilidad y especificidad resumidas en un umbral común, mientras que el modelo HSROC se enfoca en estimar una curva resumen a partir de estudios que han utilizado diferentes umbrales. Por lo tanto, explicaremos el modelado jerárquico para el metaanálisis del estudio de precisión en pruebas diagnósticas, y diseñaremos un esquema de decisión que ayude a comprender los modelos para elegir el más adecuado en situaciones de heterogeneidad en los estudios, e ilustraremos su aplicación en una revisión sistemática, estudiando las propiedades y supuestos de los procedimientos metaanalíticos para sintetizar la evidencia cuantitativa de los parámetros. Para ello, utilizamos una revisión sistemática que obtuvo estimaciones resumidas para el diagnóstico de aspergilosis invasiva, siendo nuestro marco de modelado el procedimiento NLMIXED de SAS, obteniendo estimaciones resumidas de sensibilidad y especificidad para el modelo bivariado de 0.7708 y 0.8521, a partir de un total de 27 estudios que involucraron a 3,943 pacientes, y para el caso del HSROC los valores fueron 0.7304 y 0.8867 respectivamente. Finalmente, esperamos que este artículo brinde a los clínicos una comprensión suficiente de la terminología y los métodos estadísticos, obteniendo interpretaciones plausibles de los resultados en revisiones sistemáticas.

Palabras clave**:** Modelo bivariante; Precisión diagnóstica; Heterogeneidad; HSROC; Meta-análisis.

Abbreviations

The following abbreviations are used throughout this manuscript:

D+	Sick
D-	No sick
T+	Positive Test
T-	Negative Test
TP	True positive
FP	False positive
TN	True negative
FN	False negative
Se	Sensitivity
Sp	Specificity
LR+	Positive maximum likelihood ratio
LR-	Negative maximum likelihood ratio
DOR	Diagnostic Odd ratio
AUC	Area Under the Curve
FPR	False Positive Rate
TPR	True Positive Rate

1. Introduction

In medical research, the most common known type of study is the clinical trial or interventional study. It is conducted, for example, to evaluate new drugs or therapies. The outcomes of interest could be a cure for a disease. For a good introduction to interventional studies, see, for example, (Schumacher & Schulgen-Kristiansen, 2008). Contrary to that, in other cases, researchers are interested in a binary outcome according to disease status (1: sick (D+), 0: not sick (D-)), Table 1.

Table 1: Summary table of data.

Test Results	$D+$	$D-$	Total
Positive $T+$	$TP(a)$	$FP(b)$	Positive tests $(a+b)$
Negative $T-$	$FN(c)$	$TN(d)$	Negative tests $(c+d)$
Total	Patients with the disease (a+c)	Patients without the disease $(b+d)$	Total patients $(a+b+c+d)$

Health care professionals (primarily physicians) use diagnostic tests to determine whether or not a person (usually a patient) has a particular disease or condition. Diagnostic test accuracy studies provide information on how well the tests distinguish patients with the disease from those without the disease. Most tests are imperfect and errors will occur. Therefore, statistical methods focus on two statistical measures of diagnostic accuracy, test sensitivity (the probability of a diseased individual having a positive test result) and test specificity (the probability of a healthy individual having a negative test result). A diagnostic test accuracy (DTA) study aims to quantify and compare these measures for one or more diagnostic tests to describe how well each test classifies individuals, and to estimate and compare the likely error rates (false positive and false negative diagnoses) that may be encountered. Pooling the results of multiple studies addressing the same question through meta-analysis will provide a more accurate estimate of test performance than is possible in a single study. The degree of variability in test performance between studies (heterogeneity) can be quantified in a meta-analysis and formal investigations of potential sources of heterogeneity can also be performed to explain why results differ between studies.

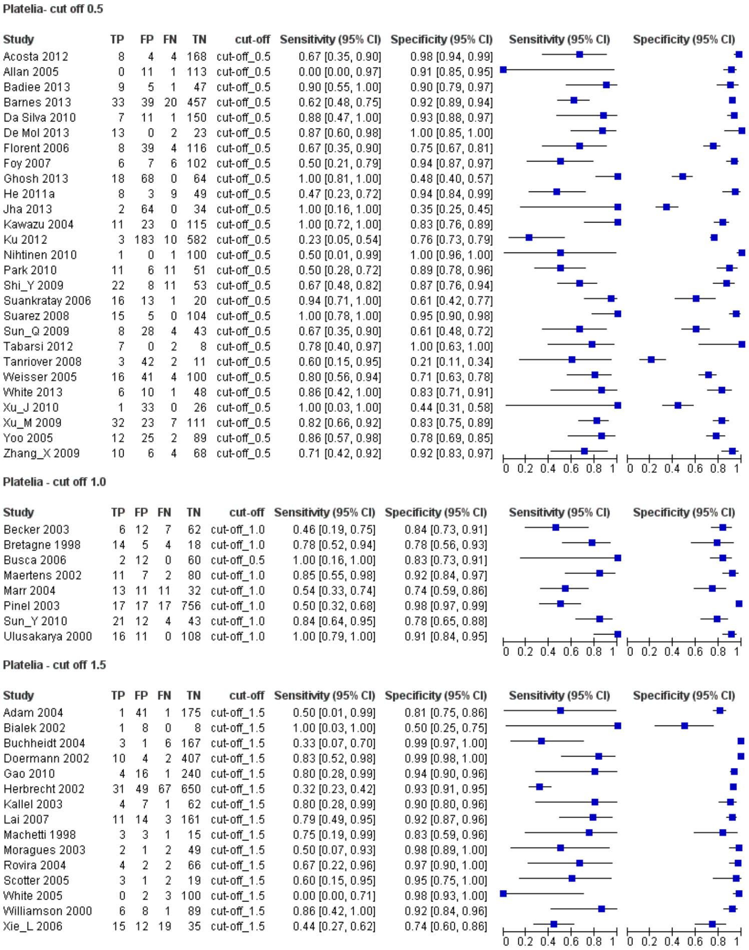
It is critical that the recommended methods for pooling study results are well understood to ensure appropriate application. In this paper, we summarize the basic concepts in diagnostic accuracy research as a prelude to explaining the rationale for the recommended methods for DTA meta-analysis, describe the properties of the methods, and use a published example to illustrate their application in different types of analyses, making use of a modification to the scheme proposed by Pambabay-Calero et al. (2020).

2. Materials and Methods

In this section, we present hierarchical modeling and the most common statistical measures to summarize test accuracy. We describe how the parameters of hierarchical models are estimated and discuss analytically the estimation of the correlation between studies. To ensure a real-world context, this section includes an illustrative example from the medical literature where the HSROC (Rutter & Gatsonis, 2001) and bivariate (Reitsma et al., 2005) approach are potentially important.

2.1. Illustrative Example

Data from a previous systematic review (Leeflang et al., 1996) were used, the aim of which was to obtain synthesized estimates of the diagnostic accuracy of serum galactomannan detection for the diagnosis of invasive aspergillosis by ELISA (Platelia© sandwich test). Fifty-four studies were included in the review (50 in the meta-analysis), containing 7955 patients, of whom 748 had proven or probable invasive aspergillosis. Three different cut-off points (thresholds) were used to measure the optical density index (ODI), whose values were 0.5, 1 and 1.5, see Figure 1.



Platelia- cut off 0.5

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Acosta 2012	8	4	4	168	cut-off_0.5	0.67 [0.35, 0.90]	0.98 [0.94, 0.99]
Allan 2005	0	11	1	113	cut-off_0.5	0.00 [0.00, 0.97]	0.91 [0.85, 0.95]
Badiee 2013	9	5	1	47	cut-off_0.5	0.90 [0.55, 1.00]	0.90 [0.79, 0.97]
Barnes 2013	33	39	20	457	cut-off_0.5	0.62 [0.48, 0.75]	0.92 [0.89, 0.94]
Da Silva 2010	7	11	1	150	cut-off_0.5	0.88 [0.47, 1.00]	0.93 [0.88, 0.97]
De Mol 2013	13	0	2	23	cut-off_0.5	0.87 [0.60, 0.98]	1.00 [0.85, 1.00]
Florent 2006	8	39	4	116	cut-off_0.5	0.67 [0.35, 0.90]	0.75 [0.67, 0.81]
Foy 2007	6	7	6	102	cut-off_0.5	0.50 [0.21, 0.79]	0.94 [0.87, 0.97]
Ghosh 2013	18	68	0	64	cut-off_0.5	1.00 [0.81, 1.00]	0.48 [0.40, 0.57]
He 2011a	8	3	9	49	cut-off_0.5	0.47 [0.23, 0.72]	0.94 [0.84, 0.99]
Jha 2013	2	64	0	34	cut-off_0.5	1.00 [0.16, 1.00]	0.35 [0.25, 0.45]
Kawazu 2004	11	23	0	115	cut-off_0.5	1.00 [0.72, 1.00]	0.83 [0.76, 0.89]
Ku 2012	3	183	10	582	cut-off_0.5	0.23 [0.05, 0.54]	0.76 [0.73, 0.79]
Nihtinen 2010	1	0	1	100	cut-off_0.5	0.50 [0.01, 0.99]	1.00 [0.96, 1.00]
Park 2010	11	6	11	51	cut-off_0.5	0.50 [0.28, 0.72]	0.89 [0.78, 0.96]
Shi_Y 2009	22	8	11	53	cut-off_0.5	0.67 [0.48, 0.82]	0.87 [0.76, 0.94]
Suankratay 2006	16	13	1	20	cut-off_0.5	0.94 [0.71, 1.00]	0.61 [0.42, 0.77]
Suarez 2008	15	5	0	104	cut-off_0.5	1.00 [0.78, 1.00]	0.95 [0.90, 0.98]
Sun_Q 2009	8	28	4	43	cut-off_0.5	0.67 [0.35, 0.90]	0.61 [0.48, 0.72]
Tabarsi 2012	7	0	2	8	cut-off_0.5	0.78 [0.40, 0.97]	1.00 [0.63, 1.00]
Tanriover 2008	3	42	2	11	cut-off_0.5	0.60 [0.15, 0.95]	0.21 [0.11, 0.34]
Weisser 2005	16	41	4	100	cut-off_0.5	0.80 [0.56, 0.94]	0.71 [0.63, 0.78]
White 2013	6	10	1	48	cut-off_0.5	0.86 [0.42, 1.00]	0.83 [0.71, 0.91]
Xu_J 2010	1	33	0	26	cut-off_0.5	1.00 [0.03, 1.00]	0.44 [0.31, 0.58]
Xu_M 2009	32	23	7	111	cut-off_0.5	0.82 [0.66, 0.92]	0.83 [0.75, 0.89]
Yoo 2005	12	25	2	89	cut-off_0.5	0.86 [0.57, 0.98]	0.78 [0.69, 0.85]
Zhang_X 2009	10	6	4	68	cut-off_0.5	0.71 [0.42, 0.92]	0.92 [0.83, 0.97]

Platelia - cut off 1.0

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Becker 2003	6	12	7	62	cut-off_1.0	0.46 [0.19, 0.75]	0.84 [0.73, 0.91]
Bretagne 1998	14	5	4	18	cut-off_1.0	0.78 [0.52, 0.94]	0.78 [0.56, 0.93]
Busca 2006	2	12	0	60	cut-off_0.5	1.00 [0.16, 1.00]	0.83 [0.73, 0.91]
Maertens 2002	11	7	2	80	cut-off_1.0	0.85 [0.55, 0.98]	0.92 [0.84, 0.97]
Marr 2004	13	11	11	32	cut-off_1.0	0.54 [0.33, 0.74]	0.74 [0.59, 0.86]
Pinel 2003	17	17	17	756	cut-off_1.0	0.50 [0.32, 0.68]	0.98 [0.97, 0.99]
Sun_Y 2010	21	12	4	43	cut-off_1.0	0.84 [0.64, 0.95]	0.78 [0.65, 0.88]
Ulusakarya 2000	16	11	0	108	cut-off_1.0	1.00 [0.79, 1.00]	0.91 [0.84, 0.95]

Platelia - cut off 1.5

Study	TP	FP	FN	TN	cut-off	Sensitivity (95% CI)	Specificity (95% CI)
Adam 2004	1	41	1	175	cut-off_1.5	0.50 [0.01, 0.99]	0.81 [0.75, 0.86]
Bialek 2002	1	8	0	8	cut-off_1.5	1.00 [0.03, 1.00]	0.50 [0.25, 0.75]
Buchheidt 2004	3	1	6	167	cut-off_1.5	0.33 [0.07, 0.70]	0.99 [0.97, 1.00]
Doermann 2002	10	4	2	407	cut-off_1.5	0.83 [0.52, 0.98]	0.99 [0.98, 1.00]
Gao 2010	4	16	1	240	cut-off_1.5	0.80 [0.28, 0.99]	0.94 [0.90, 0.96]
Herbrecht 2002	31	49	67	650	cut-off_1.5	0.32 [0.23, 0.42]	0.93 [0.91, 0.95]
Kallel 2003	4	7	1	62	cut-off_1.5	0.80 [0.28, 0.99]	0.90 [0.80, 0.96]
Lai 2007	11	14	3	161	cut-off_1.5	0.79 [0.49, 0.95]	0.92 [0.87, 0.96]
Machetti 1998	3	3	1	15	cut-off_1.5	0.75 [0.19, 0.99]	0.83 [0.59, 0.96]
Moragues 2003	2	1	2	49	cut-off_1.5	0.50 [0.07, 0.93]	0.98 [0.89, 1.00]
Rovira 2004	4	2	2	66	cut-off_1.5	0.67 [0.22, 0.96]	0.97 [0.90, 1.00]
Scotter 2005	3	1	2	19	cut-off_1.5	0.60 [0.15, 0.95]	0.95 [0.75, 1.00]
White 2005	0	2	3	100	cut-off_1.5	0.00 [0.00, 0.71]	0.98 [0.93, 1.00]
Williamson 2000	6	8	1	89	cut-off_1.5	0.86 [0.42, 1.00]	0.92 [0.84, 0.96]
Xie_L 2006	15	12	19	35	cut-off_1.5	0.44 [0.27, 0.62]	0.74 [0.60, 0.86]

Figure 1: Forest Plot of sensitivity and specificity. The squares represent the sensitivity and specificity of a study, the black line is its confidence interval. Studies are grouped according to the reported cut-off point.

2.2. Main Measures of Test Accuracy

The most commonly used measures are summarized in Table 2. These measures are global indicators of test performance, including sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV, NPV), and positive and negative likelihood ratios (LR+, LR-). Sensitivity and specificity are the most frequently reported measures in primary studies and are also used for meta-analysis (Dahabreh et al., 2012). Primary studies evaluating a test at various thresholds sometimes present results as an Receiver Operating Characteristic (ROC) curve. The ROC curve of a test is the graph of the sensitivity and specificity values obtained by varying the positivity threshold by all possible values, indicating an inverse relationship between these measures (Macaskill et al., 2010). In other words, an ROC curve is a graph used to represent the performance of a test by showing the impact of sensitivity and specificity as the threshold is varied, Figure 2.

The position of the ROC curve depends on the discriminatory ability of the test; the more accurate the test, the closer the curve is to the upper left corner of the ROC space. The diagonal line joins the lower left corner with the upper right corner of the ROC space (see Figure 2), and represents a test that provides essentially no information, since the ability to detect genuine cases is no better than chance. The upper left corner represents a test with 100% sensitivity and specificity, in other words, a perfect dichotomous test. Clearly, a desirable test is as close to the upper left corner as possible and as far from the diagonal as possible.

The most common global measures are the diagnostic odd ratio (DOR) and the area under the curve (AUC). These measures summarize the accuracy of the test at all possible thresholds, but are not useful in clinical practice because they do not provide information on error rates in diseased (false negative) and non-diseased (false positive) groups. Error rates are important for judging the extent and possible impact of subsequent consequences of testing. In meta-analysis, DOR can be a useful measure when comparing tests or subgroups, particularly if there is no preference for sensitivity or specificity.

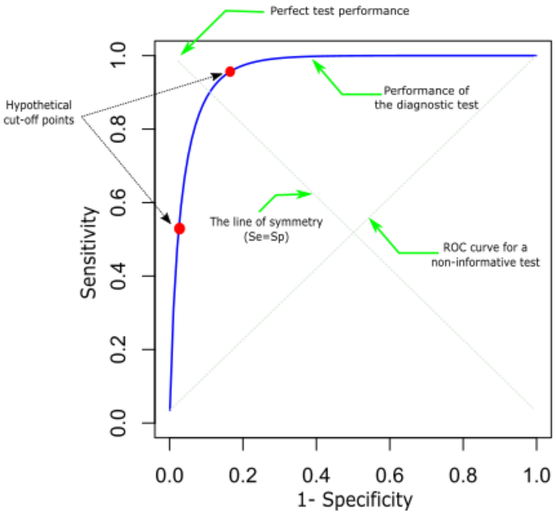


Figure 2: Example of a receiver operating characteristic (ROC) curve. Note: The ROC curve is based on the hypothetical distributions for a hypothetical test and cut-off points in Table 1.

Table 2: Definition of summary measures for diagnostic test precision, formulas based on the notation used in Table 1.

Parameter	Formula	Description
Se	$a/(a+c)$	The likelihood of a sick individual having a positive test result
Sp	$d/(b+d)$	The proportion of healthy people who have a negative test.
PPV	$a/(a+b)$	The probability that an individual with a positive test result has the disease.
NPV	$d/(c+d)$	The probability that an individual with a negative test result does not have the disease.
LR+	$\frac{a/(a+c)}{b/(b+d)}$	Describes how much more likely a positive test result is in the diseased group compared to the non-diseased group.
LR-	$\frac{c/(a+c)}{d/(b+d)}$	Describes how many times a negative test result is less likely in the unaffected group compared to the unaffected group.
DOR	$\frac{ad}{bc} = \frac{LR+}{LR-}$	Describes how many times more likely it is to get a positive test result in a sick person than in an unaffected person.

2.3. Meta-analytical methods for diagnostic accuracy

Since different studies included in a meta-analysis may explicitly use different thresholds or variations in the way the test is interpreted and applied, the recommended meta-analytic methods explicitly or implicitly allow for negative correlation between sensitivity and specificity across studies, induced by threshold variation.

Simple univariate meta-analytical methods pool sensitivity and specificity separately, ignoring the potential threshold effect. Such analyses can give misleading results as illustrated (Irwig et al., 1995). A perfect test would have 100% sensitivity and specificity. However, in reality the two measures are almost always negatively correlated, so that higher sensitivity is associated with decreased specificity (see Figure 3). Negative correlation is usually a function of the threshold beyond which a test result is considered positive. For example, a very large threshold will result in fewer false positives (increased specificity) but more false negatives (reduced sensitivity).

An Summary Receiver Operating Characteristic (SROC) curve approach was developed by Moses et al. (1993) to account for possible threshold heterogeneity. It fits a straight line to the logit transformations of the false positive rate (FPR)(1-Sp) and true positive rate (TPR) (Se) of each study, and their slope and intersection give the curve parameters. The SROC curve summarizes the sensitivity and specificity pairs from multiple studies. The model proposed by Irwig et al. (1995) may be appropriate if all the observed heterogeneity is due to a threshold effect. That is, where all the observed heterogeneity is due to the use of different thresholds in the included studies. This method allows correlation between sensitivity and specificity, but is not statistically rigorous, as the assumptions of linear regression are not met (Harbord et al., 2008). Moreover, since it is based on a DOR analysis, summary measures of sensitivity and specificity are not directly available.

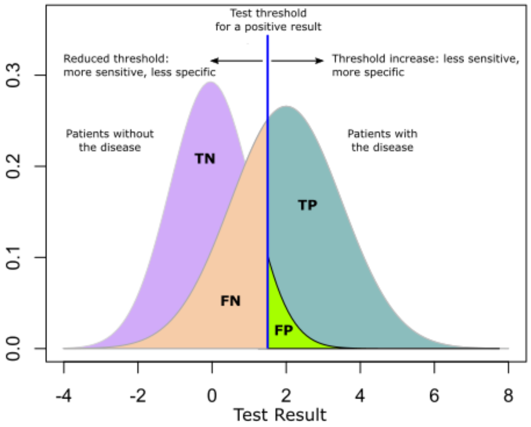


FIGURE 3: Test threshold and impact on diagnostic accuracy.

Two statistically rigorous approaches based on hierarchical models have been proposed that overcome the limitations of the model. In this section, the bivariate model and the hierarchical model (HSROC) are described and analyzed.

2.4. Hierarchical ROC Summary Model (HSROC)

The HSROC model proposed by Rutter & Gatsonis (2001) is based on a latent-scale logistic regression model. The HSROC model assumes that there is an underlying ROC curve in each study with parameters α and β characterizing the precision and skewness of the curve, in a manner similar (although technically different) to the parameters in the linear regression method of Irwig et al. (1995). The HSROC model more appropriately incorporates both within- and between-study variability, and allows greater flexibility in the statistical estimation of summary measures. The HSROC model describes within-study variability using a binomial distribution for the number of positive tests in diseased and nondiseased patients. The model is specified at two levels: within-study modeling and between-study modeling. The within-study modeling takes the following form:

$$logit\left(\pi_{ij}\right) = \left(\theta_i + \alpha_i X_{ij}\right) exp\left(-\beta X_{ij}\right) \tag{1}$$

The variable π_{ij} is the probability that a patient in study i with disease condition j will get a positive test result. Defining $j = 0$ for a patient without the disease (Rutter & Gatsonis, 2001) $X_{ij} = -1/2$) and $j = 1$ for a patient with the condition of interest (Rutter & Gatsonis, 2001) $X_{ij} = 1/2$, it follows that for study i, π_{i0} is the false positive rate and π_{i1} is the true positive rate. The parameter X_{ij} is a dummy variable for the true disease state of a study patient i with disease state j. The parameters θ_i and α_i are the cutoff point and precision parameter, respectively, and are allowed to vary between studies. Finally β, is a scaling parameter that models the possible asymmetry in the ROC curve. The following definitions for parameters θ_i and α_i include a common covariate Z that affects both parameters, although they can be modeled without covariates or with multiple covariates.

$$\theta_i \sim N\left(\Theta + \gamma Z_i, \sigma_\theta^2\right) \tag{2}$$

$$\alpha_i \sim N\left(\Lambda + \lambda Z_i, \sigma_\alpha^2\right) \tag{3}$$

The model was originally formulated in a Bayesian framework and, therefore, a priori distributions must be specified (Rutter & Gatsonis, 2001). Thus, the specification of the hierarchical model is completed with the choice of a priori distributions for the parameters. In particular, uniform and gamma (Γ) distributions are chosen, as shown below:

$$\begin{aligned}
\Theta &\sim Uniform\left(\mu_{\theta 1}, \mu_{\theta 2}\right)\\
\Lambda &\sim Uniform\left(\mu_{\alpha 1}, \mu_{\alpha 2}\right)\\
\beta &\sim Uniform\left(\mu_{\beta 1}, \mu_{\beta 2}\right)\\
\gamma &\sim Uniform\left(\mu_{\gamma 1}, \mu_{\gamma 2}\right)\\
\lambda &\sim Uniform\left(\mu_{\lambda 1}, \mu_{\lambda 2}\right)\\
\sigma_\theta^2 &\sim \Gamma^{-1}\left(\xi_{\theta 1}, \xi_{\theta 2}\right)\\
\sigma_\alpha^2 &\sim \Gamma^{-1}\left(\xi_{\alpha 1}, \xi_{\alpha 2}\right)
\end{aligned}$$

The parameters $\Theta, \Lambda, \beta, \gamma, \lambda, \sigma_\theta^2$ and σ_α^2 are assumed to be mutually independent. Likewise the parameters $\mu_{\theta 1}, \mu_{\theta 2}, \mu_{\alpha 1}, \mu_{\alpha 2}, \mu_{\beta 1}, \mu_{\beta 2}, \mu_{\gamma 1}, \mu_{\gamma 2}, \mu_{\lambda 1}, \mu_{\lambda 2}, \xi_{\theta 1}, \xi_{\theta 2}$ are assumed to be fixed by choosing values for them that reflect plausible ranges.

The model produces an SROC curve by allowing the cutoff point parameter to vary while keeping the precision parameter at its mean value, i.e., the summary ROC (SROC) curve can be plotted using expected values of $\Lambda + \lambda Z$ and β. If the true disease state is coded by $1/2$ for positive cases and $-1/2$ for cases without disease, then for a given value of covariate Z_i, the true positive rate (TPR) can be expressed as:

$$TPR = logit^{-1}\left(\left(logit\left(FP\right) exp\left[E\left(\beta\right)/2\right] + E\left[\Lambda + \lambda Z\right]\right) exp\left[E\left[\beta\right]/2\right]\right)$$

Then the SROC curve is plotted by the ordered pair (FP, TPR) for $FP \in [0, 1]$.

2.5. Bivariate Random-Effects Model for Sensitivity and Specificity

As with the HSROC method, the bivariate approach preserves the two-dimensional nature of the original data by jointly modeling sensitivity and specificity (Reitsma et al., 2005). This method also incorporates modeling for the correlation that may exist between these two measures using a random effects approach. The evaluation of the bivariate model requires the specification of an appropriate transformation (e.g., a logit transformation that is applied in the generalized linear mixed model) (Menke, 2010). Explanatory variables can be added to the bivariate model and lead to separate effects on sensitivity and specificity, rather than a net effect on the odds ratio scale as in the SROC approach. The bivariate model is specified as follows, (Harbord et al., 2008):

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma_{AB}\right) \tag{4}$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \tag{5}$$

The variables μ_{Ai} and μ_{Bi} are the logit transformations of the sensitivity and specificity, with variances σ_A^2 and σ_B^2, respectively, for study i. Where σ_A^2 and σ_B^2 describe the between-study variability of the true value of the logit transforms of sensitivity and specificity and σ_{AB} is the covariance between the logit of sensitivity and specificity. The model can also be parameterized by employing the correlation $\rho_{AB} = \sigma_{AB}/\left(\sigma_A \sigma_B\right)$ which may be more interpretable than the covariance. Therefore, the bivariate model has five parameters: $\mu_A, \mu_B, \sigma_A^2, \sigma_B^2$ and ρ_{AB}.

The Model 4 can be extended by incorporating the precision of the sensitivity (Se_i) and specificity (Sp_i) that have been measured in each study, the corresponding estimate of the variance for the logit transformations of sensitivity and specificity of each study are given by:

$$S_{Ai}^2 = \frac{1}{(a+c)\, Se_i \left(1 - Se_i\right)}$$

$$S_{Bi}^2 = \frac{1}{(b+d)\, Sp_i \left(1 - Sp_i\right)}$$

The standard outputs of the bivariate model include, means for the logit transformations of sensitivity (μ_A) and specificity (μ_B) with their standard errors and respective 95% confidence intervals; and estimates of the between-study variability of logit sensitivity σ_A^2 and specificity σ_B^2 and the covariance between them σ_{AB}. In relation to the above parameters, the following measures of interest can be estimated, Table 3.

Table 3: Standard outputs of the bivariate model.

Parameter	**Formula**
$LR+$	$\dfrac{\exp\left\{\mu_A/\left(1+\exp\left(\mu_A\right)\right)\right\}}{1-\left[\exp\left\{\mu_B/\left(1+\exp\left(\mu_B\right)\right)\right\}\right]}$
LR-	$\dfrac{1-\left[\exp\left\{\mu_A/\left(1+\exp\left(\mu_A\right)\right)\right\}\right]}{\exp\left\{\mu_A/\left(1+\exp\left(\mu_B\right)\right)\right\}}$
DOR	$\exp\left\{\mu_A+\mu_B\right\}$
ρ_{AB}	$\dfrac{\sigma_{AB}}{\sqrt{\sigma_A^2}\sqrt{\sigma_B^2}}$

The bivariate model also allows covariates to affect sensitivity and/or specificity. Assuming we have a single covariate level Z study that can affect both sensitivity and specificity, then the model can be extended as follows:

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_A + \nu_A Z_i \\ \mu_B + \nu_B Z_i \end{pmatrix}, \Sigma_{AB}\right) \tag{6}$$

where ν_A and ν_B (which are treated as fixed effects) are coefficients representing the effects of the covariate Z on the logit transformation of Se and Sp respectively.

2.6. Comparison of hierarchical methods

The HSROC and bivariate methods are equivalent under certain parameterizations, such as in the absence of covariates or when the same covariates affect both sensitivity and specificity (bivariate case) and the accuracy and cutoff parameters (HSROC case) (Harbord et al., 2007). Therefore, in situations where there are no covariates, the two models will return equivalent summary estimates for sensitivity and specificity (and also any other measures derived from these two measures stated above). The relationship of the HSROC and bivariate model is clarified below, (Harbord et al., 2007).

From Equation 1, it follows that, $logit\,(\pi_{i1}) = \left(\theta_i + \frac{\alpha_i}{2}\right)\exp\left(-\beta/2\right)$ and $logit\,(\pi_{i0}) = \left(\theta_i - \frac{\alpha_i}{2}\right)\exp\left(\beta/2\right)$. Similarly, from Equation 4, we have that the logit transformations for sensitivity (π_{i1}) and specificity $(1-\pi_{i0})$ are, $\mu_{Ai} = logit\,(\pi_{i1})$ and $\mu_{Bi} = logit\,(1-\pi_{i0}) = -logit\,(\pi_{i0})$ respectively. Therefore, we can relate the random variables that form the basis of the two models, by the following expressions:

$$\mu_{Ai} = \left(\theta_i + \frac{\alpha_i}{2}\right)\exp\left\{-\beta/2\right\} \tag{7}$$

$$\mu_{Bi} = \left(\theta_i - \frac{\alpha_i}{2}\right)\exp\left\{\beta/2\right\} \tag{8}$$

Equations 7 and 8, imply that μ_{Ai} and μ_{Bi} are linear combinations of two random variables, θ_i and α_i, which the HSROC model assumes have independent normal distributions. Higgins & Thompson (2002) states that any pair of independent normal random variables has a bivariate normal distribution. It is known that the HSROC model assumes a bivariate normal distribution for μ_{Ai} and μ_{Bi}, (Rutter & Gatsonis, 2001); therefore, the HSROC model is equivalent to the bivariate model.

Making use of a matrix notation and making $b = \exp\{\beta/2\}$, one can more formally express Equations 7 and 8 as follows:

$$\begin{gathered} \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} = S^{-1}\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} \\ S^{-1} = \begin{pmatrix} b^{-1} & 1/2b^{-1} \\ -b & 1/2b \end{pmatrix} \end{gathered} \tag{9}$$

Inverting Equation 9, it follows that:

$$\begin{aligned} \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} &= S \begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \\ S &= \begin{pmatrix} b^{-1} & 1/2b^{-1} \\ -b & 1/2b \end{pmatrix} \end{aligned} \tag{10}$$

Where S is the transformation matrix associated with the change from the bivariate model coordinates (logit transformations of sensitivity and specificity) to the HSROC model coordinates (cut-off point and precision parameters).

3. Steps to Follow in Performing a Meta-Analysis of Diagnostic Tests

There is a growing consensus that the HSROC and bivariate approaches offer the best methodologies for pooling diagnostic test accuracy studies, but there are differences between the two approaches and the nature of the underlying data may determine which approach is more appropriate.

A first step is to examine the sensitivity and specificity distributions of the included studies separately (Pambabay-Calero et al., 2020). If either measure shows a lack of heterogeneity, then it is more appropriate to analyze the data using univariate meta-analysis to derive point estimates for sensitivity and specificity with their respective confidence intervals. However, an integration of all studies may provide contextual information that may warrant a full analysis in these situations. If only one study is available, then there is clearly no basis for meta-analysis. If only two studies are available, then there is insufficient information available to reliably estimate all parameters in the HSROC and bivariate models. Therefore, in the case of two studies, meta-analysis is not recommended and a narrative description of the studies should be presented.

The correlation between sensitivity and specificity is important and is estimated by HSROC and bivariate methods. A positive correlation between TPR and FPR is generally expected. However, study data often exhibit significant volatility and a positive correlation cannot be estimated. If there is a significant negative correlation, this implies that sensitivity improves with increasing specificity, which is unlikely to occur in practice because of the relationship between disease status and test cutoff point (see Figure 3). In the case of a negative correlation, such a finding should be discussed in relation to the nature of the test and the amount of evidence.

A key consideration is whether or not a threshold effect is present, which is generally evidenced by a positive correlation between false positive rate and sensitivity. When a threshold effect is present, then an SROC approach is appropriate, which can be achieved using HSROC or bivariate models. Test accuracy estimates can be plotted in ROC space. In the absence of a threshold effect, the SROC approach is not appropriate. There will be situations where a threshold effect may or may not be plausible, depending on the nature of the test and the indication. Because of differences in the way test results are interpreted, a threshold effect may arise even when there is a universally employed cutoff point.

If a threshold effect is plausible and heterogeneity is observed, it should be evaluated whether the heterogeneity can be attributed to a threshold effect. The determination of whether the observed heterogeneity is due to a threshold effect is generally based on a visual inspection of the distribution of the study points relative to the SROC curve. If the study points are very close to the SROC curve, there will be reasonable confidence that the threshold effect is responsible for the heterogeneity. An inspection of the shape of the confidence region is also useful, particularly to verify whether the region spans and largely follows the shape of the SROC curve. If, on the other hand, the prediction region bears little relation to the SROC curve, or the study points are not close to the SROC curve, then it is reasonable to conclude that factors other than just a threshold effect are responsible for the observed heterogeneity. The choice of method used should be justified by the context (e.g., studies, inclusion of covariates, correlation between sensitivity and specificity), and the potential impact of assumptions on the interpretation of results should be clarified, for a clearer idea of how to choose the hierarchical model, see Figure 4.

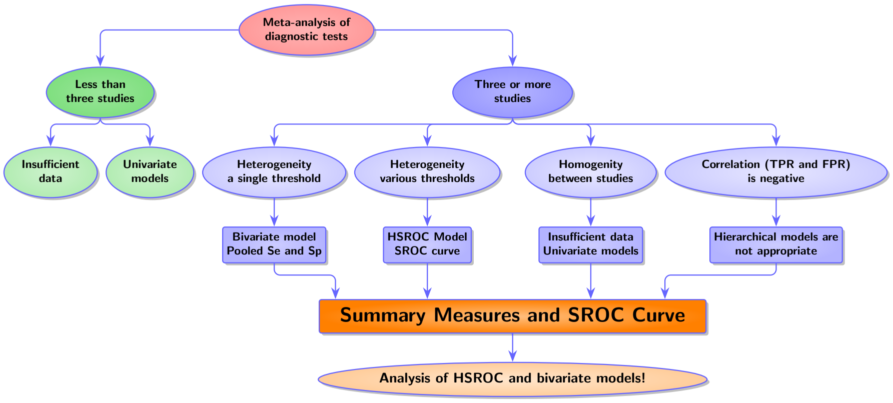

Meta-analysis of diagnostic tests
Less than three studies
Three or more studies
Insufficient data
Univariate models
Heterogeneity a single threshold
Heterogeneity various thresholds
Homogenity between studies
Correlation (TPR and FPR) is negative
Bivariate model Pooled Se and Sp
HSROC Model SROC curve
Insufficient data Univariate models
Hierarchical models are not appropriate
Summary Measures and SROC Curve
Analysis of HSROC and bivariate models!

Figure 4: Steps to follow for fitting a bivariate model or HSROC in a meta-analysis of diagnostic tests.

Three kinds of research can be conducted in a DTA review that reflect the types of questions that can be addressed. Such questions are: what is the diagnostic accuracy of a test, how does the accuracy of two or more tests differ, and how does test accuracy vary with clinical and methodological characteristics? Each of these questions will be considered below, using the illustrative example.

4. Results

4.1. Estimation of a Global Measure for Sensitivity and Specificity at a Fixed Threshold

The bivariate model jointly synthesizes sensitivity and specificity to provide summary estimates that are plotted as a single point in SROC space. Confidence and prediction regions plotted around the summary point allow joint inferences to be made about sensitivity and specificity. These regions take into account the correlation between sensitivity and specificity, and are useful to illustrate, uncertainty and the extent of heterogeneity. Since summary points should only be calculated when studies share a common threshold, the available data are reduced. The choice of a common threshold is often based on the available data and may not be the threshold used in clinical practice. In addition, a common threshold is difficult to define for non-numerical tests.

The summary point for the Platelia© test can be estimated by selecting the studies that used the recommended threshold of 0.5, see Figure 1. This restriction reduces the data for the meta-analysis from 50 to 27 studies. When observing the Forest Plot in Figure 1, we note a clear heterogeneity among the studies for sensitivity and specificity values, which indicates that it is feasible to fit a hierarchical model: in this case the bivariate model. Note also that the correlation coefficient between sensitivity and false positive rate is 0.172, which implies that the result should be presented as a summary point with the corresponding confidence region for sensitivity and specificity (or false positive rate).

The overall sensitivity and specificity were 0.7708 (with 95% CI: 0.6806-0.8610) and 0.8521 (with 95% CI: 0.7791-0.9250) respectively. Figure 5 shows this summary point with a 95% confidence region (green color) and a 95% prediction region (blue color). The confidence region is based on a CI around the summary point and indicates that, based on the available data, we would expect the "true value" to be within that region 95% of the time. The prediction region around the summary point indicates the region in which we would expect to find the results of a new study in the future, and is therefore broader than the confidence region, as it goes beyond the uncertainty of the available data. Despite the use of a common threshold, heterogeneity is considerable and evident in Figure 5.

4.2. SROC CURVE SUMMARY

The HSROC model is based on the estimation of an SROC curve. The strength of this approach lies in the inclusion of data from each study, regardless of the type of threshold used in them, thus maximizing the use of the available data. Note that a meta-analysis of diagnostic tests usually includes information from each study in a 2×2 table, considering that care should be taken with studies that use different types of thresholds (Macaskill et al., 2010).

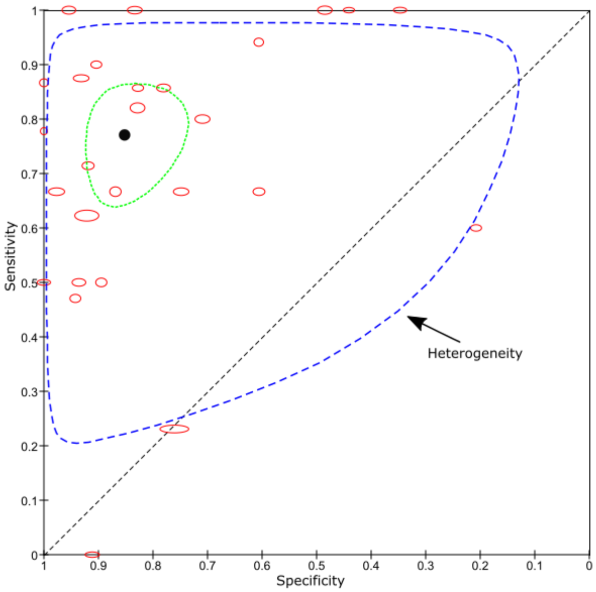


FIGURE 5: Summary point for sensitivity and specificity in the SROC space for detection of invasive aspergiosis cut-off point 0.5. The size of each study (red color) is plotted according to the sample size (number of patients included in the study). The black solid circle (summary point) represents the overall estimate of sensitivity and specificity. The summary point is surrounded by a dashed dotted line (green color) representing the 95% confidence region and a dashed line (blue color) representing the 95% prediction region (the region within which we are 95% certain that the results of a new study will be found).

Although a global measure for an SROC curve (including different thresholds) is clinically impossible to interpret. Estimates for sensitivity with their respective confidence intervals can be calculated from the HSROC model by fixing the specificity or vice versa, allowing to observe the changes between these measures along the curve.

The SROC curve for the Platelia© test can be estimated by selecting the 50 studies of the meta-analysis, see Figure 1. Looking at the Forest Plot in Figure 1, we notice a clear heterogeneity and different cut-off points (values of 0.5, 1.00 and 1.50) among the studies, which indicates that it is feasible to fit a hierarchical model. Furthermore, the correlation coefficient between sensitivity and false positive rate is 0.209. As mentioned above, due to the variation in the threshold between studies, an SROC curve (HSROC model) is appropriate to summarize these data with the corresponding confidence region for TPR and FPR.

The dispersion in the threshold used to define test positivity across studies is reflected in the variability of the sensitivity and specificity estimates shown in Figure 6. To avoid extrapolation beyond the data, the curve was drawn within the range of observed specificities (0.21 to 1.00) of the 50 included studies. Given the relationship between the bivariate and HSROC (Harbord et al., 2007) models, it is possible to calculate the overall sensitivity and specificity by estimating an average point on the SROC curve. These values are 0.7304 (0.6597 to 0.8011) and 0.8867 (0.8478 to 0.9257) respectively.

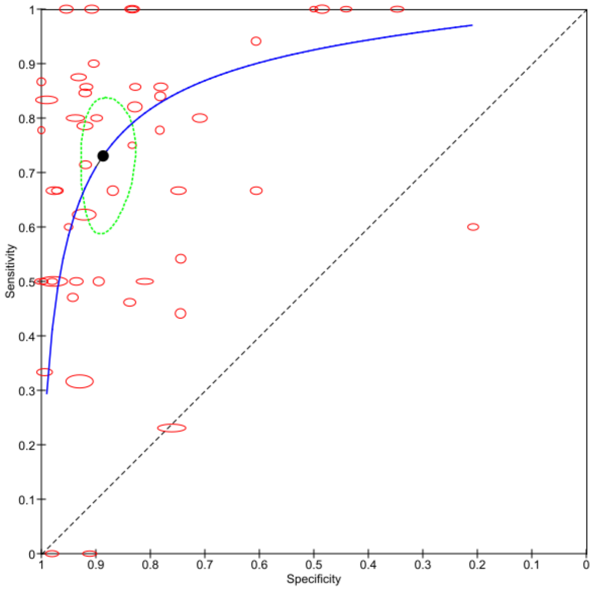


Figure 6: SROC curve for invasive aspergillosis by ELISA (Platelia© test type with three cut-off points). Each study point is plotted according to the number of patients. The SROC curve was plotted restricted to the specificity range (0.21 to 1.00) of the 50 studies included in the evaluation of aspergillosis. The solid black circle (summary point) represents the overall estimate of sensitivity and specificity. The summary point is surrounded by a dashed dotted line (green color) representing the 95% confidence region.

4.3. Explaining Heterogeneity

Although measures of heterogeneity exist for univariate meta-analyses (e.g., index I) (Higgins & Thompson, 2002), there is no analogous measure for bivariate meta-analyses. The amount of observed heterogeneity is quantified in terms of the random effects captured by bivariate and HSROC models, but these are not easily interpreted (Macaskill et al., 2010; Pambabay-Calero et al., 2018, 2020). The distribution of studies, in SROC space using the ordered pair (TPR; FPR) and the prediction ellipse, can give clues to the existence of heterogeneity due to variation in the test threshold (Macaskill et al., 2010; Pambabay-Calero et al., 2020, 2021).

A common approach to explore heterogeneity is meta-regression, where study-level covariates are included to estimate overall statistical measures. Both HSROC and the Bivariate model facilitate the use of study-level covariates such as, categorical (e.g., test type) or continuous (e.g., mean patient age) (Macaskill et al., 2010). In the bivariate model, covariates can be incorporated to affect sensitivity and/or specificity.

The HSROC model, on the other hand, allows the addition of covariates to affect test positivity, position and shape of the curve. A covariate can be associated with some of these three parameters.

The effect of the cut-off point on the accuracy of aspergiosis was evaluated by comparing SROC curves for the three subgroups in an HSROC model, see Figure 7.

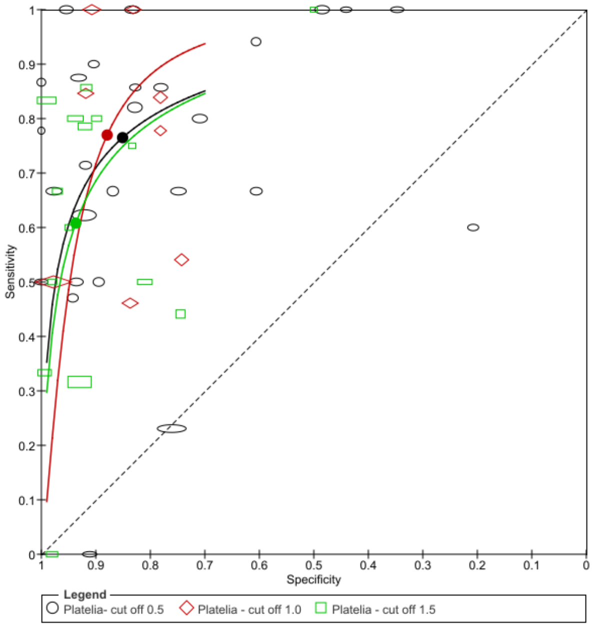


Figure 7: SROC curves for invasive aspergillosis by ELISA (Platelia© test type with three cut-off points). Each study point was scaled according to study size and cut-off point (0.21 to 1.00 for the 0.5 cut-off point, 0.74 to 0.98 for the 1.00 cut-off point and 0.50 to 0.99 for the 1.5 cut-off point).

The model was adjusted by including a covariate (cutoff points) for precision, threshold and curve shape. Based on the relationship between HSROC and the bivariate model, summary points for sensitivity and specificity were calculated by applying the HSROC model only to studies that used the recommended threshold for each instrument (test). The overall estimates are shown in Table 4.

Looking at Figure 7, it is noticeable that the SROC curves for the three types of thresholds cross. This indicates that no test is more accurate than the others and that the relative accuracy depends on the threshold, i.e., the estimates of the sensitivities and/or specificities of the tests are not statistically significant with a p-value = 0.195.

Table 4: Diagnostic accuracy for the Platelia© test according to cut-off points for the detection of invasive aspergiosis. Summary sensitivity and specificity are shown for each instrument at the recommended cut-off point.

Prueba	Cut-off point	N	Cases	Patients	Sensitivity (95% CI)	Specificity (95% CI)
Platelia©	0.5	27	394	3943	0.7650 (0.6781-0.8341)	0.8509 (0.7716-0.9061)
Platelia©	1.00	8	145	1391	0.7698 (0.5538-0.9001)	0.8792 (0.7878-0.9346)
Platelia©	1.5	15	209	2621	0.6080 (0.4526-0.7443)	0.9368 (0.8811-0.9674)

5. Discussion

Estimation of diagnostic test accuracy is often intended to compare two or more tests for the same indication. In this situation, the diagnostic accuracy of all tests should be compared (Bossuyt et al., 2006; Pambabay-Calero et al., 2020, 2021). However, the evidence derived from comparative and non-comparative studies often differs.

The option to show a SROC curve depends on whether or not the studies were subjected to a common positivity threshold. In cases where the threshold varies, it is appropriate to represent the results by an SROC curve, since a global measure for sensitivity and specificity is not advisable, because it represents an average between the different thresholds. The SROC curve represented by the HSROC model is always increasing. Although it is feasible to generate SROC curves with the bivariate approach, such a model can generate curves with negative slopes (Dahabreh et al., 2012).

It is suggested that the SROC curve should be restricted to the observed range of specificities in the included studies and should not be extrapolated beyond the observed data. Estimation of diagnostic test accuracy is often intended to compare two or more tests for the same condition of interest. In such a situation, the diagnostic accuracy of all tests should be compared (Bossuyt et al., 2006; Ristow et al., 2023; Watjer et al., 2023). However, evidence derived from comparative and noncomparative studies often differs. Ideally, for the purposes of comparing two diagnostic tests, studies in which all patients are evaluated by all tests or are randomly assigned to receive one or the other of the tests to guide test selection are preferred.

When multiple tests are used to diagnose the condition of interest, it is very likely that the tests will not be performed independently of one another (Novielli et al., 2013). When the assumption of dependence between the tests is ignored, this may lead to erroneous estimation of the probability of disease.

Meta-analysis of diagnostic tests can be used to generate more robust estimates by grouping studies according to their characteristics (type of test, cut-off points, etc). The accuracy of the diagnostic test is not a measure of clinical effectiveness and improvement of the test does not necessarily imply a satisfactory outcome for the patient.

There are a variety of global measures (abstracts) that describe the accuracy of the diagnostic test, although the most common measures in a meta-analysis are sensitivity and specificity for a meta-analysis. Since there is an inverse (negative) relationship between sensitivity and specificity, a DTA meta-analysis should consider this particularity.

Acknowledgements

Thanks. We are very grateful to the anonymous reviewers for their helpful comments.

[Received: July 2024 — Accepted: April 2025]

References

Bossuyt, P. M., Irwig, L., Craig, J. & Glasziou, P. (2006), 'Comparative accuracy: assessing new tests against existing diagnostic pathways', *Bmj* **332**(7549), 1089–1092.

Dahabreh, I. J., Trikalinos, T. A., Lau, J. & Schmid, C. (2012), 'An empirical assessment of bivariate methods for meta-analysis of test accuracy [internet]'.

Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. & Sterne, J. A. (2007), 'A unification of models for meta-analysis of diagnostic accuracy studies', *Biostatistics* **8**(2), 239–251.

Harbord, R. M., Whiting, P., Sterne, J. A., Egger, M., Deeks, J. J., Shang, A. & Bachmann, L. M. (2008), 'An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary', *Journal of Clinical Epidemiology* **61**(11), 1095–1103.

Higgins, J. P. & Thompson, S. G. (2002), 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine* **21**(11), 1539–1558.

Irwig, L., Macaskill, P., Glasziou, P. & Fahey, M. (1995), 'Meta-analytic methods for diagnostic test accuracy', *Journal of Clinical Epidemiology* **48**(1), 119–130.

Leeflang, M. M., Debets-Ossenkopp, Y. J., Wang, J., Visser, C. E., Scholten, R. J., Hooft, L., Bijlmer, H. A., Reitsma, J. B., Zhang, M., Bossuyt, P. M. et al. (1996), 'Galactomannan detection for invasive aspergillosis in immunocompromised patients', *Cochrane Database of Systematic Reviews* **2017**(9).

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R. & Takwoingi, Y. (2010), 'Cochrane handbook for systematic reviews of diagnostic test accuracy'.

Menke, J. (2010), 'Bivariate random-effects meta-analysis of sensitivity and specificity with sas proc glimmix', *Methods of Information in Medicine* **49**(01), 54–64.

Moses, L. E., Shapiro, D. & Littenberg, B. (1993), 'Combining independent studies of a diagnostic test into a summary roc curve: data-analytic approaches and some additional considerations', *Statistics in Medicine* **12**(14), 1293–1316.

Novielli, N., Cooper, N. J. & Sutton, A. J. (2013), 'Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency?', *Value in health* **16**(4), 536–541.

Pambabay-Calero, J., Bauz-Olvera, S., Nieto-Librero, A., Sánchez-García, A. & Galindo-Villardón, P. (2021), 'Hierarchical modeling for diagnostic test accuracy using multivariate probability distribution functions', *Mathematics* **9**(11), 1310.

Pambabay-Calero, J. J., Bauz-Olvera, S. A., Nieto-Librero, A. B., Galindo-Villardon, M. P. & Hernandez-Gonzalez, S. (2018), 'An alternative to the cochran-(q) statistic for analysis of heterogeneity in meta-analysis of diagnostic tests based on hj biplot', *Investigación Operacional* **39**(4), 536–545.

Pambabay-Calero, J. J., Bauz-Olvera, S. A., Nieto-Librero, A. B., Galindo-Villardón, M. P. & Sánchez-García, A. B. (2020), 'A tutorial for meta-analysis of diagnostic tests for low-prevalence diseases: Bayesian models and software', *Methodology* **16**(3), 258–277.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. & Zwinderman, A. H. (2005), 'Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews', *Journal of Clinical Epidemiology* **58**(10), 982–990.

Ristow, O., Schnug, G., Smielowksi, M., Moratin, J., Pilz, M., Engel, M., Freudlsperger, C., Hoffmann, J. & Rückschloß, T. (2023), 'Diagnostic accuracy comparing opt and cbct in the detection of non-vital bone changes before tooth extractions in patients with antiresorptive intake', *Oral Diseases* **29**(3), 1039–1049.

Rutter, C. M. & Gatsonis, C. A. (2001), 'A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations', *Statistics in Medicine* **20**(19), 2865–2884.

Schumacher, M. & Schulgen-Kristiansen, G. (2008), *Methodik klinischer Studien: Methodische Grundlagen der Planung, Durchführung und Auswertung*, Springer-Verlag.

Watjer, R. M., Bonten, T. N., Arkesteijn, M. A., Quint, K. D., van der Beek, M. T., van der Raaij-Helmer, L. M., Numans, M. E. & Eekhof, J. A. (2023), 'The accuracy of clinical diagnosis of onychomycosis in dutch general practice: a diagnostic accuracy study', *BJGP open* **7**(3).