



Bivariate hierarchical model for the Meta-analysis of diagnostic tests in studies with binary responses: its application from SAS and R

Sergio A. Bauz-Olvera¹ orcid.org/0000-0003-2366-5358

Johnny J. Pambabay-Calero² orcid.org/0000-0002-9474-9586

Ana B. Nieto-Librero³ orcid.org/0000-0001-6643-247X

Ma. Purificación Galindo-Villardón⁴ orcid.org/0000-0001-6977-7545

¹Escuela Superior Politécnica del Litoral, Facultad de Cs. de la Vida, Guayaquil, Ecuador.

✉ serabauz@espol.edu.ec

²Escuela Superior Politécnica del Litoral, Facultad de Cs. Naturales y Matemáticas, Guayaquil, Ecuador.

✉ jpambaba@espol.edu.ec

Universidad de Salamanca, Dept. de Estadística, Facultad de Medicina, Salamanca, España.

³✉ ananieto@usal.es ; ⁴ ✉ pgalindo@usal.es

Received: April 2020 | Accepted: August 2020

Abstract:

Studies on the precision of diagnostic tests usually report the number of true positives, false positives, true negatives, and false negatives. There is generally a negative association between the number of true positives and true negatives, as studies that adopt less strict criteria to declare a test as positive need higher sensitivities and lower specificities. Given this particularity, the bivariate nature of the data must be preserved, by modeling sensitivity and specificity together. In this paper, we will use the bivariate hierarchical model applied to a meta-analysis data set which was an update to a previous systematic review of diagnostic tests for chronic Chagas disease. Our modeling framework was implemented with SAS NLMIXED procedure, making it possible to obtain summary measures for sensitivity and specificity, with values of 0.725 and 0.995, respectively, out of a total of 35 studies with 6057 patients.

Keywords: Sensitivity; Specificity; Diagnostic accuracy; Random effects; Bivariate approach.

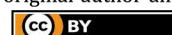
MSC (2010): 62-07, 62H05, 62H10, 65C60.

Cite this article as (IEEE citation style):

S. A. Bauz-Olvera, J. J. Pambabay-Calero, A. B. Nieto-Librero, and Ma. P. Galindo-Villardón, "Bivariate hierarchical model for the Meta-analysis of diagnostic tests in studies with binary responses: its application from SAS and R", *Proyecciones (Antofagasta, On line)*, vol. 39, no. 5, pp. 1365-1380, Oct. 2020, doi: 10.22199/issn.0717-6279-2020-05-0083.



Article copyright: © 2020 Sergio A. Bauz-Olvera, Johnny J. Pambabay-Calero, Ana B. Nieto-Librero, and Ma. Purificación Galindo-Villardón. This is an open access article distributed under the terms of the Creative Commons Licence, which permits unrestricted use and distribution provided the original author and source are credited.



1. Introduction

Although statistical methods for the meta-analysis of clinical trials are currently well developed and understood, there are still challenges when performing meta-analysis of data from studies on the accuracy of the diagnostic test. This is mainly due to the bivariate nature of the response, where information about sensitivity and specificity should be summarized [1, 2]. This correlation can be induced by populations or heterogeneous environments in the different trials, but it primarily stems from the balance between sensitivity and specificity when the threshold for the definition of the test positivity varies. The bivariate model [1] is a random-effects model that focuses on the estimation of meta-analytic sensitivities and specificities and has the advantages of accessing individual data and allowing unexplained heterogeneity, as well as allowing for the correlation between sensitivity and specificity [3]. In addition, it can be generalized for modeling covariates and deal with extreme values of 100% for sensitivity and specificity without applying artificial continuity corrections, using standard software, such as SAS [4] and R [5], which can be used for the analysis. Considering the conclusion of the study conducted by Brasil et al. [6], which reported that tests are probably biased and overestimated, we propose the use of the bivariate model to improve assessment of the diagnosis.

2. Methods

In this section, hierarchical model for the bivariate approach (bivariate hierarchical model) is presented. We describe how the model is estimated and analyze the estimation of the correlation between studies in an analytical manner. To ensure a real-world context, this section includes a motivating example of medical literature where the bivariate approach is potentially important.

3. Illustrative example

We used data from a previous systematic review [7]. The previous systematic review included information from 1980 to 2009, and the update period included data from 2009 to May 2014. For our cases, we selected 35 studies involving 6,057 patients. Two tests were used to measure the feature of interest, the first of which was enzyme-linked immunosorbent assay, while the second one was the index test, which was based on the

molecular-based technology called polymerase chain reaction (PCR). For our purposes, we selected the studies subjected to the PCR test, because it makes it possible to constitute groups of interest according to the test methodology, as suggested by Schijman et al. [8]. These methodologies include: deoxyribonucleic acid (DNA) (M1) extraction [extraction from blood in ethylenediamine tetraacetic acid (EDTA)-guanidine with phenol-chloroform and amplification of sat-DNA using quantitative PCR], DNA (M4) extraction (extraction from blood in EDTA guanidine with phenol-chloroform and amplification of k-DNA [121-122 primers] with hot-start PCR), and the Coris Oligo-C test, without any of the improvements proposed by Schijman.

4. Evaluation of the accuracy of diagnostic test studies

There are two groups of patients in a typical study of diagnostic test analysis: those who actually have a certain condition or disease, and those who are truly free of the disease. We assume that the presence or absence of the disease has been determined through a completely accurate standard procedure, and therefore we assume that the true state of each patient is known.

Our interest is focused on a so-called index test that is expected to be a reliable guide as to whether the patient has the disease or not [24] (Figure 1).

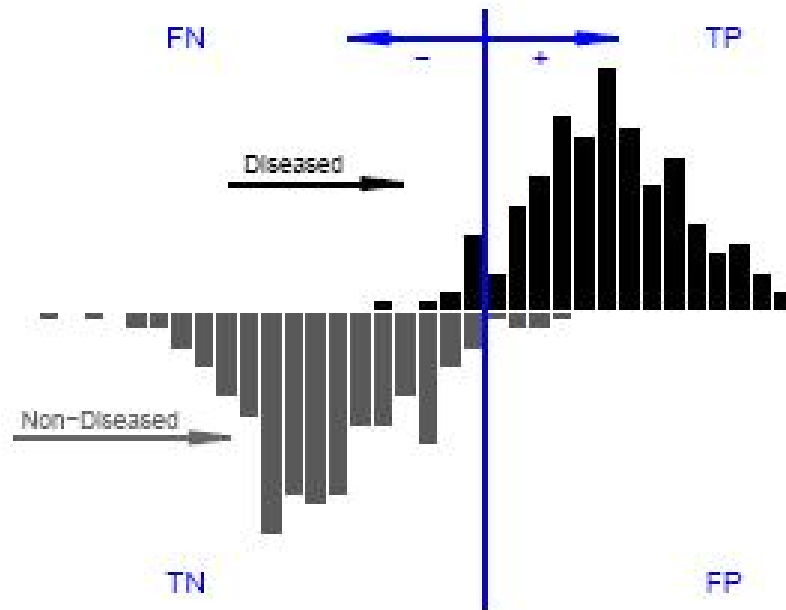


Figure 1: Test threshold and impact on diagnostic accuracy

The test can be based on a biomarker, a questionnaire, an imaging modality, or a more complex diagnostic procedure. However, the index test is not completely reliable. Therefore, we will express a proportion of patients who have a positive index test result as $T+$. Similarly, a proportion of patients who show negative index test results shall be expressed as $T-$. A generic representation of the resulting data is shown in Table 1, where n_1 denotes patients with the condition of interest and n_2 refers to healthy individuals. True positives, false positives, false negatives, and true negatives and denoted by the values of TP, FP, FN, and TN respectively (Table 1).

Test result	Patient status		Total
	D^+	D^-	
T^+	TP	FP	TP + FP
T^-	FN	TN	FN + TN
Total	n_1	n_2	n

Table 1: Generic representation of the precision analysis of a diagnostic test

Sensitivity and specificity are two common measures to evaluate the performance of a diagnostic test, and we will therefore define these two measures. Sensitivity is defined as the probability that a patient will have a positive test result if he/she actually has the disease. It is estimated by

$$(4.1) \quad \widehat{Se} = \frac{TP}{n_1}$$

The estimated sensitivity is also called the TP rate. The corresponding FN rate is expressed as $1 - \widehat{Se} = \frac{FN}{n_1}$. Specificity is defined as the probability that a patient will have a negative test result if he/she does not actually have the disease. This measure is estimated by the TN rate, expressed as

$$(4.2) \quad \widehat{Sp} = \frac{TN}{n_2}$$

The corresponding FP rate (FPR) is expressed as $1 - \widehat{Sp} = \frac{FP}{n_2}$. Ideally, both the sensitivity and the specificity of a diagnostic test should be close to one. The results of a diagnostic test study are usually reported as a pair (sensitivity, specificity). However, some attempts have been made to condense the result into a single number. The most common approach is the use of the diagnostic odds ratio (DOR) [9]. DOR can be defined as

$$(4.3) \quad DOR = \frac{TP/FN}{FP/TN} = \frac{TP \times TN}{FP \times FN} = \frac{\widehat{Se} \times \widehat{Sp}}{(1 - \widehat{Se}) \times (1 - \widehat{Sp})}$$

DOR can also be defined in terms of the likelihood ratios. The positive and negative likelihood ratio is defined as

$$LR+ = \frac{\widehat{Se}}{1 - \widehat{Sp}}; LR- = \frac{1 - \widehat{Se}}{\widehat{Sp}}, \text{ respectively.}$$

Therefore, an alternative definition of DOR using likelihood ratios is expressed as

$$(4.4) \quad DOR = \frac{LR+}{LR-}$$

5. Bivariate general normal model for the meta-analysis of random effects

The bivariate approach [1] directly models sensitivity and specificity. The model can be considered to have two levels corresponding to the variation within and between the studies. At the first level, it is assumed that the variability within the study for both measures follows a binomial distribution. We will define the modeling as

$$(5.1) \quad y_{A_i} \sim \text{Binomial}(n_{A_i}, \pi_{A_i}), y_{B_i} \sim \text{Binomial}(n_{B_i}, \pi_{B_i})$$

In this equation, if A is the sensitivity; the number of positive tests of each study, y_{A_i} , is distributed according to a binomial, that is, $y_{A_i} \sim B(n_{A_i}, \pi_{A_i})$, where n_{A_i} refers to the number of individuals with the condition of interest and π_{A_i} is the probability of a positive test result in the study group i. Similarly, if specificity is denoted as B, the number of negative tests in each study is distributed according to a binomial $y_{B_i} \sim B(n_{B_i}, \pi_{B_i})$, where n_{B_i} denotes the total number of healthy individuals evaluated and π_{B_i} indicates the probability of a negative result of the test in the study group i. The sensitivity and specificity pair for each study should be modeled jointly at level one (within the study) of the analysis, because they are linked by shared study characteristics, including the positivity threshold. In the upper level, it is assumed that the logit transformations of the sensitivities have a normal distribution, with mean μ_A and variance σ_A^2 , while the logit transformations of the specificities are distributed according to a normal with mean μ_B and variance σ_B^2 . Its correlation is included by

modeling both normal distributions using a normal bivariate distribution as

$$(5.2) \quad \begin{pmatrix} \mu_{A_i} \\ \mu_{B_i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma \right), \Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

In these equations, σ_A^2 and σ_B^2 describe the variability between studies of the true value of the logit transformations of sensitivity and specificity, and σ_{AB} is the covariance between the sensitivity and specificity logit. The model can also be parameterized using the correlation $\rho_{AB} = \sigma_{AB} / (\sigma_A \sigma_B)$ which may be more interpretable than covariance. Therefore, the bivariate model has five parameters: μ_A , μ_B , σ_A^2 , σ_B^2 and σ_{AB} (or ρ_{AB}). The model can extend by incorporating the precision used to measure the sensitivity and specificity in each study. Therefore, the corresponding variances of the logit transformations for sensitivity and specificity in each study are expressed as

$$S_{A_i}^2 = \frac{1}{n_{A_i} A_i (1 - A_i)} \text{ and } S_{B_i}^2 = \frac{1}{n_{B_i} B_i (1 - B_i)}, \text{ respectively.}$$

If we treat the observed variance of logit sensitivity and specificity as fixed quantities, a standard approach can be written as follows:

$$\begin{pmatrix} \hat{\mu}_{A_i} \\ \hat{\mu}_{B_i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{A_i} \\ \mu_{B_i} \end{pmatrix}, C_i \right), C_i = \begin{pmatrix} S_{A_i}^2 & 0 \\ 0 & S_{B_i}^2 \end{pmatrix} \quad (5.3)$$

The final model can be expressed using the following equation, where C_i is a diagonal matrix:

$$(5.4) \quad \begin{pmatrix} \hat{\mu}_{A_i} \\ \hat{\mu}_{B_i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{A_i} \\ \mu_{B_i} \end{pmatrix}, \Sigma + C_i \right)$$

This model can be adjusted by means of the methods based on the likelihood, in particular with SAS NLMIXED procedure, because it allows the user to adjust the variance within the assay to specific values according to the study, [10].

6. Data synthesis and analysis

We performed a descriptive or exploratory analysis using the Meta-Analysis of Diagnostic Accuracy package from R [5], and obtained graphs such as crosshair and receiver operating characteristic (ROC) ellipses to summarize the confidence intervals of the studies according to sensitivity (Se) and FP. We estimated the summary measures of Se, Sp, LR+, LR-, and DOR for the included studies. It should be noted that these measures are estimated using a univariate approach, because there is no joint modeling to capture the relationship between sensitivity and specificity. The summary measures can be calculated using a fixed-effect (Mantel-Haenszel) [11] or random (DerSimonian-Laird) [12] model, depending on the value of I^2 [13].

To establish whether there was inconsistency and heterogeneity in the meta-analysis, we summarized the performance characteristics using the Higgins I^2 index [15] and we evaluated the heterogeneity by visually inspecting a forest plot, biplot [17] and the Cochran Q test ($p > 0.1$). The DOR summary measures were estimated using the random effect model (DerSimonian-Laird) [12] following the recommendations [14]. Because I^2 was equal to zero and the Q test was > 0.5 , the variability of the studies between different studies was not confirmed even though the crosshair and ROC ellipses plots showed a true heterogeneity with greater intensity of sensitivity. Therefore, an investigation of the heterogeneity was necessary and bivariate adjustment was used to estimate a summary measure of the studies in terms of sensitivity and specificity.

The separate sensitivity and specificity study can lead to biased results because different thresholds were used in different studies. Therefore, we used the bivariate hierarchical [1] study to estimate diagnostic accuracy parameters and to generate a summary-receiver operating characteristic (SROC) curve [16].

7. Results

The Se of each study varied between 0.01 and 1, while the Sp ranged from 0.49 to 1 (Figure 2). A considerable degree of heterogeneity was seen in the sensitivities in regard to the specificities. Figure 3 shows the studies responsible for the high levels of heterogeneity, how the cutoff points vary and how correlations are shown between the sensitivity and the false-positive rate. On the other hand, the value of the Q statistic was 28,978 with 34

degrees of freedom and a p-value of 0.712, implying that there is no statistical heterogeneity in the analyzed data, and the value of index I^2 was zero. The aforementioned results express a contradiction, because the graphic evidence shows the opposite. From the above, it is clear that we need to adjust the data to a statistical model that captures true heterogeneity. It is important to note that different cutoff points were used in different studies, which generates heterogeneity and different levels of sensitivity [18]. Therefore, heterogeneity should be modeled with the bivariate approach.

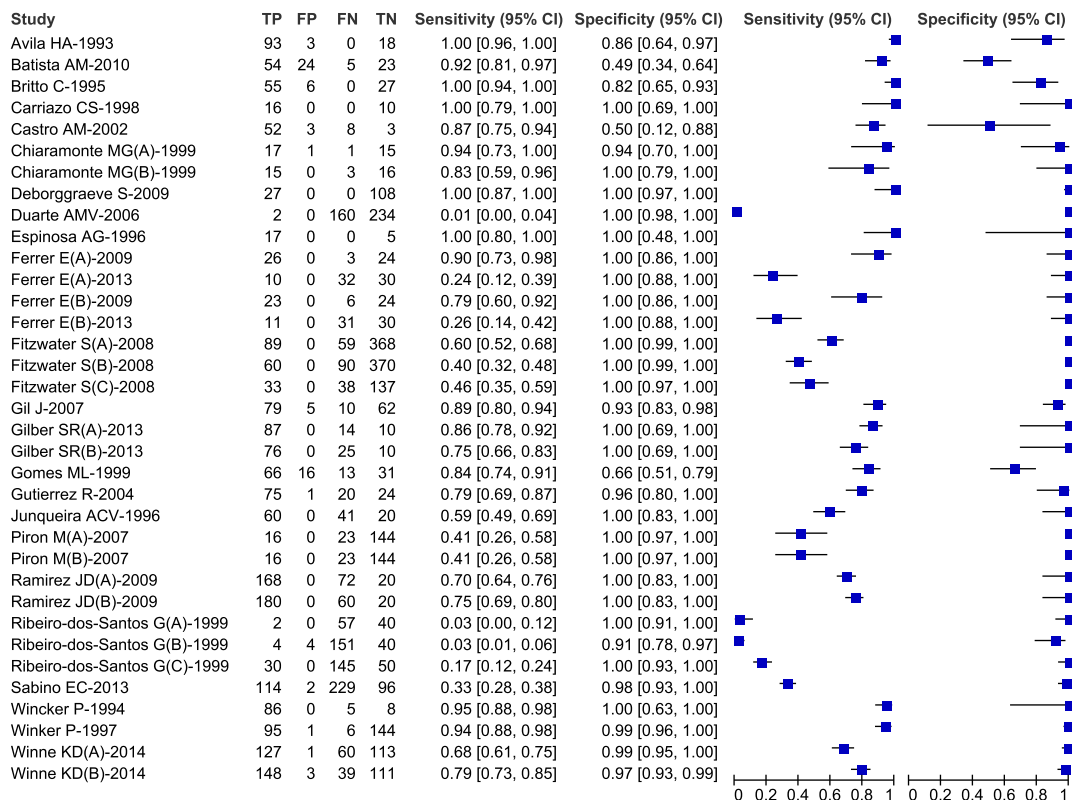


Figure 2: Forest Plot of Sensitivity and Specificity, Chagas data

In this bivariate model, the SROC curve is plotted as Se versus FPR to describe the uncertainty of the sensitivity pair and the false-positive rates, to observe how the different cutoff points vary, which will have an impact on the heterogeneity of the data and on the overall diagnostic accuracy. The SROC curve summarizes the relationship between Se and $(1-Sp)$ in all the studies, considering the heterogeneity between them. We created an SROC curve using all the selected studies. The SROC curves for the data are shown in (Figure 3). It is worth noting that it is an important graphical tool to understand how the diagnostic accuracy of the different tests depends on the different cutoff points [13].

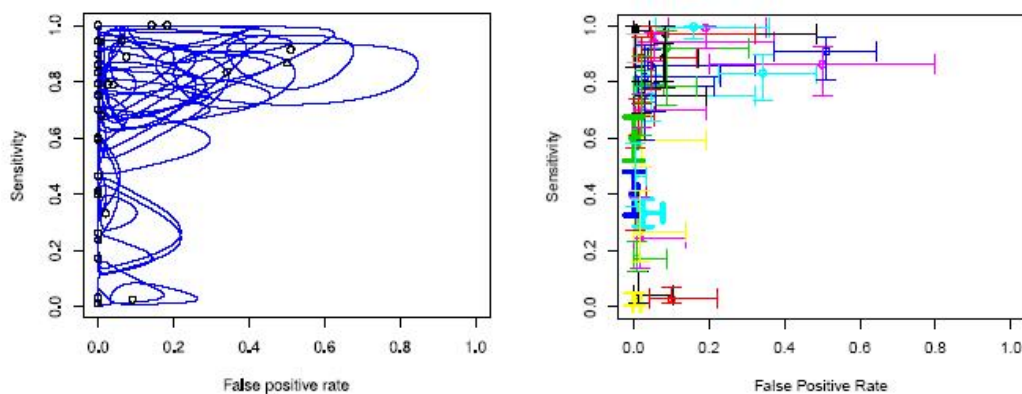


Figure 3: ROC ellipses and crosshair that describe regions of confidence and the uncertainty of the sensitivity and specificity pair

Figure 4 shows that the prediction region covers a greater range of Se than that of Sp. This may be because of the fact that most of the studies had more participants with negative test results than those with positive results of the test, which led to greater sampling variability when we estimated Se vs. Sp. Figure 4 also shows an asymmetry of the test performance measures toward a higher Sp with a greater Se variability, providing indirect evidence of some variability in the threshold. It also shows that, when the threshold is increased, Se drops but Sp increases.

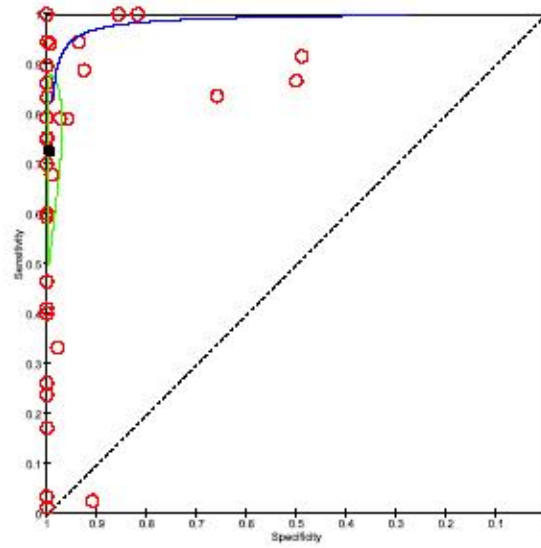


Figure 4: The SROC curve of the bivariate model showing the accuracy of the test (all the selected studies)

It is also possible to appreciate the summaries of sensitivity and specificity with values of 0.725 and 0.995 and standard errors of 0.076 and 0.003, respectively. On the other hand, we have the estimations for log LR+, log LR-, and log DOR whose values are 6.282, 4.993, and -1.289, respectively. Note that these results do not include the methodology as a covariate.

8. Analysis of Heterogeneity using the bivariate model

What is described here is the exploration of the role of moderating variables. In our case, we have the methodology of the test, which we will categorize as 1 (M1, M4, and Coris Oligo-C for 12 studies) for the studies that applied the PCR test with methodology and 2 (no methodology for 23 studies) for the studies in which no methodology was applied. The bivariate model allows covariates to affect sensitivity and/or specificity. Assuming that we have a single Z-level covariate study that can affect the sensitivity and specificity, the model can be extended as shown below:

$$(8.1) \quad \begin{pmatrix} \mu_{A_i} \\ \mu_{B_i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A + \nu_A Z_i \\ \mu_B + \nu_B Z_i \end{pmatrix}, \Sigma \right),$$

where ν_A y ν_B (which are treated as fixed effects) are coefficients that represent the effects of the covariate Z on logit of Se and Sp, respectively.

According to the confidence regions in Figure 4, it is clear that sensitivity varies according to the methodology, but not specificity. Summary estimates for the specificity were 0.9956 for the studies where the methodology test was not applied, and 0.9942 for the studies where the methodology test was applied. The estimation summaries for sensitivity were 0.7390 and 0.6970 for the tests where methodology was applied and not applied, respectively. Figure 5 compares the result of this model with that of the model without covariates; it is clear that the variances of random effects have been decreased, particularly for sensitivity.

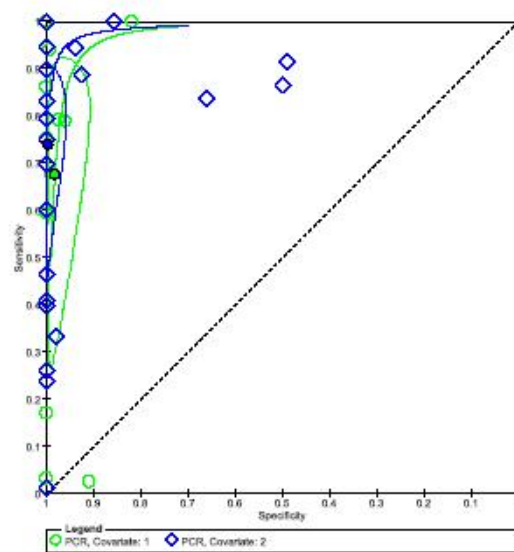


Figure 5: SROC curve, bivariate adjustment with covariate

9. Discussion

The bivariate method is an adequate method to perform a meta-analysis for diagnostic tests. Harbord et al. [19] have shown that the bivariate approach is adequate because it recognizes and differentiates between intrastudy and interstudy variation through a second level of random effects, making it possible to preserve the two-dimensional nature of the studies, that is, to model the relationship between the sensitivity and specificity.

We use a bivariate adjustment [1], which is robust when there is no standard reference. Another hierarchical model, i.e., hierarchical summary-receiver operating characteristic (HSROC) model, was proposed by [2], which is a multilevel model, but instead of directly modeling Se and Sp, it fits the classification data of the studies using a logistic regression of random effects. However, [20] demonstrated that the likelihood functions of both the HSROC and the bivariate models are algebraically equivalent and produce identical summary measures for sensitivity and specificity. It is important to keep in mind that the estimation of a single summary point using the bivariate model, or the estimation of a single summary curve using the HSROC model, requires five parameters to be calculated in the full specification of the model. There is little information about the estimate when the number of studies is small, which analysts should keep in mind when interpreting the results. In some situations, the models may not converge.

Although Brasil et al. [7] express that the usefulness of the PCR test is debatable, and that health service providers should not order it as a routine test for the diagnostic investigation of chronic Chagas disease, the bivariate model estimated a pooled sensitivity and specificity without covariates with values of 0.7258 and 0.9951, respectively, making it clear that the PCR test is very specific and significantly sensitive. The contribution of the previous summary measures (sensitivity and specificity) to the research conducted by [7] is evidenced by the application of mixed nonlinear algorithms, and their convergence designed by SAS.

We adapted the bivariate model using the NLMIXED procedure of the SAS statistical program [20], using a code similar to the one provided by [21]; this code is presented in the Annex. Note that our results differ slightly from the model proposed by [1], because we chose to model the binomial structure directly using the NLMIXED procedure. The diagnostic accuracy of the studies that used the PCR test were compared using a bivariate model, [22]. There was statistical evidence that the shape of the SROC curves varies, as shown in (Figure 4), which implies that the relative accuracy of the methodologies applied to the PCR tests is modified with the thresholds.

Conclusion

This document focused on the analysis of data from a systematic assessment for the accuracy of diagnostic tests. For a general description of the assessments review process see [3, 23]. Meta-analyses of diagnostic accuracy studies can provide answers to important clinical questions, but the recommended methods are demanding. We advise the review teams to seek the support of a statistician if there is no statistical experience within the team. Hierarchical models can be problematic to adjust, especially when there is little data. In summary, at present there are very powerful models to perform meta-analysis of the accuracy of binary response diagnoses, constructed as hierarchical models, which are proposed as random-effects models.

References

- [1] J. B. Reitsma, A. S. Glas, A. W. S. Rutjes, R. J. P. M. Scholten, P. M. Bossuyt, and A. H. Zwinderman, "Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews", *Journal of clinical epidemiology*, vol. 58, no. 10, pp. 982–990, 2005, doi: 10.1016/j.jclinepi.2005.02.022
- [2] C. M. Rutter and C. A. Gatsonis, "A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations", *Statistics in medicine*, vol. 20, no. 19, pp. 2865–2884, 2001, doi: 10.1002/sim.942
- [3] M. M. G. Leeflang, "Systematic reviews of diagnostic test accuracy", *Annals of internal medicine*, vol. 149, no. 12, p. 889, Dec. 2008, doi: 10.7326/0003-4819-149-12-200812160-00008
- [4] J. Menke, "Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX", *Methods of information in medicine*, vol. 49, no. 01, pp. 54–64, 2010, doi: 10.3414/me09-01-0001
- [5] P. Doebler and H. Holling, "Meta-analysis of diagnostic accuracy with mada", *R Packag*, vol. 1, p. 15, 2015. [On line]. Available: <https://bit.ly/36agmQy>
- [6] P. E. Brasil, L. D. Castro, A. M. Hasslocher-Moreno, L. H. Sangenis, and J. U. Braga, "ELISA versus PCR for diagnosis of chronic Chagas disease: systematic review and meta-analysis", *BMC infectious diseases*, vol. 10, no. 1, Nov. 2010, doi: 10.1186/1471-2334-10-337

- [7] P. E. Brasil, R. Castro, and L. D. Castro, "Commercial enzyme-linked immunosorbent assay versus polymerase chain reaction for the diagnosis of chronic Chagas disease: a systematic review and meta-analysis", *Memórias do Instituto Oswaldo Cruz*, vol. 111, no. 1, pp. 1–19, Jan. 2016, doi: 10.1590/0074-02760150296
- [8] C. Morel, "Faculty opinions recommendation of International study to evaluate PCR methods for detection of *Trypanosoma cruzi* DNA in blood samples from Chagas disease patients", *Faculty Opinions – post-publication peer review of the biomedical literature*, Jan. 2011, doi: 10.3410/f.7957957.8322055
- [9] L. E. Moses, D. Shapiro, and B. Littenberg, "Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations", *Statistics in medicine*, vol. 12, no. 14, pp. 1293–1316, Jul. 1993, doi: 10.1002/sim.4780121403
- [10] H. C. V. Houwelingen, L. R. Arends, and T. Stijnen, "Advanced methods in meta-analysis: multivariate approach and meta-regression", *Statistics in medicine*, vol. 21, no. 4, pp. 589–624, 2002, doi: 10.1002/sim.1040
- [11] J. J. Deeks, "Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests", *BMJ*, vol. 323, no. 7305, pp. 157–162, Jul. 2001, doi: 10.1136/bmj.323.7305.157
- [12] R. Dersimonian and N. Laird, "Meta-analysis in clinical trials", *Controlled clinical trials*, vol. 7, no. 3, pp. 177–188, Sep. 1986, doi: 10.1016/0197-2456(86)90046-2
- [13] J. P. T. Higgins and S. G. Thompson, "Quantifying heterogeneity in a meta-analysis", *Statistics in medicine*, vol. 21, no. 11, pp. 1539–1558, 2002, doi: 10.1002/sim.1186
- [14] M. Leeflang, J. J. Deeks, Y. Takwoingi, and P. Macaskill, "Cochrane diagnostic test accuracy reviews", *Systematic reviews*, vol. 2, no. 1, Oct. 2013, doi: 10.1186/2046-4053-2-82
- [15] J. P. T. Higgins, "Measuring inconsistency in meta-analyses", *BMJ*, vol. 327, no. 7414, pp. 557–560, Sep. 2003, doi: 10.1136/bmj.327.7414.557
- [16] T. A. Trikalinos, C. M. Balion, C. I. Coleman, L. Griffith, P. L. Santaguida, B. Vandermeer, and R. Fu, "Chapter 8: Meta-analysis of test performance when there is a 'Gold Standard'", *Journal of general internal medicine*, vol. 27, no. S1, pp. 56–66, May 2012, doi: 10.1007/s11606-012-2029-1

- [17] J. J. Pambabay-Calero, S. A. Bauz-Olvera, A. B. Nieto-Librero, M. P. Galindo-Villardón, and S. Hernández-González, “An alternative to the cochrane-(q) statistic for analysis of heterogeneity in meta-analysis of diagnostic tests based on hj biplot”, *Investigación operacional*, vol. 39, no. 4, pp. 536–544, 2018. [On line]. Available: <https://bit.ly/3j8U30t>
- [18] P. Doebler, H. Holling, and D. Böhning, “A mixed model approach to meta-analysis of diagnostic studies with binary test outcome”, *Psychological methods*, vol. 17, no. 3, pp. 418–436, Sep. 2012, doi: 10.1037/a0028091
- [19] R. M. Harbord, J. J. Deeks, M. Egger, P. Whiting, and J. A. C. Sterne, “A unification of models for meta-analysis of diagnostic accuracy studies”, *Biostatistics*, vol. 8, no. 2, pp. 239–251, May 2006, doi: 10.1093/biostatistics/kxl004
- [20] R. M. Harbord and P. Whiting, “Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression”, *The Stata journal: promoting communications on statistics and Stata*, vol. 9, no. 2, pp. 211–229, Aug. 2009., doi: 10.1177/1536867x0900900203
- [21] P. Macaskill, “Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis”, *Journal of clinical epidemiology*, vol. 57, no. 9, pp. 925–932, Sep. 2004, doi: 10.1016/j.jclinepi.2003.12.019
- [22] A. F. Carvalho, Y. Takwoingi, P. M. G. Sales, J. K. Soczynska, C. A. Köhler, T. H. Freitas, J. Quevedo, T. N. Hyphantis, R. S. McIntyre, and E. Vieta, “Screening for bipolar spectrum disorders: a comprehensive meta-analysis of accuracy studies”, *Journal of affective disorders*, vol. 172, pp. 337–346, Feb. 2015, doi: 10.1016/j.jad.2014.10.024
- [23] M. M. Leeflang, J. J. Deeks, Y. Takwoingi, and P. Macaskill, “Cochrane diagnostic test accuracy reviews”, *Systematic reviews*, vol. 2, no. 1, Oct. 2013, doi: 10.1186/2046-4053-2-82
- [24] S. A. Bauz-Olvera, J. J. Pambabay-Calero, A. B. Nieto-Librero, and M. P. Galindo-Villardón, “Meta-analysis in dta with hierarchical models Bivariate and HSROC: Simulation study”, *Springer proceedings in mathematics & statistics selected contributions on statistics and data science in Latin America*, pp. 33–42, 2019, doi: 10.1007/978-3-030-31551-1_3