AN ALTERNATIVE TO THE COCHRAN-(Q) STATISTIC FOR ANALYSIS OF HETEROGENEITY IN META-ANALYSIS OF DIAGNOSTIC TESTS BASED ON HJ BIPLOT

Johny J. Pambabay-Calero*,[a], Sergio A. Bauz-Olvera**,[b], Ana B. Nieto-Librero***, ****,[c], Ma. Purificación Galindo-Villardón***, ****,[d], Sergio Hernández-González*****,[e]
*Escuela Superior Politécnica del Litoral, Facultad de Ciencias Naturales y Matemáticas, Guayaquil, Ecuador
**Escuela Superior Politécnica del Litoral, Facultad de Ciencias de la Vida, Guayaquil, Ecuador
***Dpto. de Estadística, Facultad de Medicina, Universidad de Salamanca, Salamanca, España
****Instituto de Investigación Biomédica (IBSAL), Salamanca, España.
*****Universidad Veracruzana, Xalapa, México

ABSTRACT
The possible heterogeneity among individual studies constituting a meta-analysis is traditionally evaluated using Cochran Q and Higgins I^2 statistics. However, both indices have deficiencies: The Q statistic detects heterogeneity but does not allow its quantification, whereas the I^2 index allows for quantification of heterogeneity but does not indicate which studies are responsible for it. This problem is solved by additionally using the HJ biplot of the matrix containing the information about true positives, true negatives, false positives, and false negatives for each study. This means that the information contained in such a tetrachoric table contains the joint frequency distribution of the true classification of the disease and that provided by the diagnostic test.

MSC: 62H30

KEYWORDS: Meta-analysis, effect size, heterogeneity, Q-statistic, I^2 index, HJ biplot`

RESUMEN
La forma tradicional de evaluar la posible heterogeneidad de los estudios individuales que conforman un meta-análisis, es utilizando los estadísticos Q de Cochran e I^2 de Higgins. Sin embargo, ambos índices presentan deficiencias: el Q detecta heterogeneidad, pero no permite cuantificarla y el I^2 permite cuantificarla, pero no conocer cuáles son los estudios responsables de la heterogeneidad. Este problema se resuelve utilizando, además, el HJ-Biplot de la matriz que contiene la información sobre VP, VN, FP y FN, para todos y cada uno de los estudios; es decir la información contenida en la tabla tetracórica que contiene la distribución de frecuencias conjunta de la verdadera clasificación de la enfermedad y la proporcionada por el test diagnóstico.

PALABRAS CLAVES: Meta-análisis, tamaño del efecto, heterogeneidad, estadístico Q, índice I2, HJ biplot

1. INTRODUCTION

In a meta-analysis, the results of each study are quantified by means of an index called effect size (for example, standardized mean difference, correlation coefficient, and odds ratio, etc.) that can be applied to all studies. This allows the use of the same metric for the analysis of the results of the studies.
In general, a meta-analysis has three main objectives: **(a)** to test whether the results of the studies are homogeneous; **(b)** to obtain an overall index of the magnitude of the effect of the relationship studied, with their respective confidence intervals and their statistical significance; and **(c)** to analyze whether a significant

[a] E-mail: jpambaba@espol.edu.ec
[b] E-mail: serabauz@espol.edu.ec
[c] E-mail: ananieto@usal.es
[d] E-mail: pgalindo@usal.es
[e] E-mail: sehernadez@uv.mx

heterogeneity exists among studies and to identify possible variables or moderating characteristics of the results obtained.
The purpose of this study is to compare the performance of the **Q** test and the $\mathbf{I^2}$ index, using the HJ biplot for the analysis of heterogeneity, using data from the Audit questionnaire (Alcohol Use Disorder Identification Test). Because of the length of this questionnaire, the **Audit C** data (a short alcohol detection test, see [15]) will be used to answer only the first three questions of the AUDIT, as shown in Table 1. The table contains 14 rows and five columns: the first column identifies the respective studies with their authors and year, and the remaining columns show the values of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN).

Table 1.- Datos Audit C

Estudio - Año	TP	FN	FP	TN
Aalto et al., 2006	47	9	101	738
Aertgeerts et al., 2001	126	51	272	1543
Aertgeerts et al., 2002	19	10	12	192
Bradley et al., 2003	36	3	78	276
Bradley et al., 2007	130	19	211	959
Bush et al., 1998	84	2	68	89
Gómez et al., 2006	68	0	112	423
Gordon et al., 2001	752	0	3226	2977
Gual et al., 2002	59	5	55	136
Rumpf et al., 2002	142	50	571	2788
Seale et al., 2006	137	24	107	358
Selin, 2006	57	3	103	437
Tsai et al., 2005	34	1	21	56
Tuunanen et al., 2007	152	51	88	264

This process was conducted using a program in the R environment, whose code is attached in the Appendix.

2. MATERIALS AND METHODS

2.1. The Q-statistic and $\mathbf{I^2}$ index

The evaluation of heterogeneity in a meta-analysis is a crucial issue because the presence or absence of true heterogeneity can affect the statistical model that the individuals performing the meta-analysis (the meta-analyst) decides to apply. Therefore, when the results of the studies only vary according to one sampling error (homogeneous case), a fixed-effects model can be applied to obtain a mean effect size. Conversely, if the study results differ by more than one sampling error (heterogeneous case), the meta-analyst can assume a random-effects model, taking into account intrastudy and interstudy variability. The most common method to assess whether there is actual heterogeneity in a meta-analysis is using the Q test, a statistical test defined by [2].
Under the hypothesis of homogeneity among effect sizes, the Q-statistic follows a chi-square distribution with $k - 1$ degrees of freedom, where k is the number of studies. Not rejecting the homogeneity hypothesis usually guides the meta-analyst to adopt a fixed-effects model, because it is assumed that the sizes of the estimated effect are only differentiated by the sampling error. Conversely, the rejection of the homogeneity hypothesis may lead to the use of a random-effects model that includes both intrastudy and interstudy variability. The deficiencies of the Q-statistic are its deficient power to spot true heterogeneity between studies when a meta-analysis includes a small number of studies and its excessive power to detect insignificant variability with a large number of studies [10], [18], and [3]; i.e., its power depends on the sample size. Another strategy to quantify true heterogeneity in a meta-analysis is the estimation of the variance between studies $\boldsymbol{\tau^2}$, which represents the true value of heterogeneity among study effects produced by innumerable substantive factors (treatment type, patient characteristics), methodological (design type, sample size, wear, etc.), and study

characteristics. Because the value of $\boldsymbol{\tau^2}$ depends on the particular effect of the metric used in a meta-analysis, comparing the estimated $\boldsymbol{\tau^2}$ values from meta-analyses that have used different effect size indices (i.e., standardized mean difference, correlation coefficient and odds ratio, etc.) is not possible.

To overcome deficiencies of the **Q** test and $\boldsymbol{\tau^2}$, [12]; see also [13] have proposed the $\mathbf{I^2}$ index to evaluate heterogeneity in a meta-analysis. This index can be interpreted as a percentage of heterogeneity—i.e., the fraction of variation that is because of interstudy variance $\boldsymbol{\tau^2}$; this is obtained by dividing the difference of **Q** and $(\mathbf{k - 1})$ for Q. Thus, the $\mathbf{I^2}$ index is similar to an intraclass correlation in cluster sampling [12]. For example, a meta-analysis with $\mathbf{I^2 = 0}$ means that all variability in effect size estimates is a result of a sampling error within the studies. On the other hand, a meta-analysis with $\mathbf{I^2 = 50}$ means that half of the total variability among effect sizes is not caused by a sampling error but is caused by true heterogeneity between the studies. [12] proposed a tentative classification of $\mathbf{I^2}$ to help interpret its magnitude; therefore, the percentages **25%** $\mathbf{(I^2 = 25)}$**, 50%** $\mathbf{(I^2 = 50)}$**,** and **75%** $\mathbf{(I^2 = 75)}$ would mean a low, average, and high heterogeneity, respectively.

There is a direct correlation between the $\mathbf{I^2}$ index and interstudy variance $\boldsymbol{\tau^2}$—when $\boldsymbol{\tau^2}$ increases, $\mathbf{I^2}$ will also increase, and vice versa. One of the advantages of the $\mathbf{I^2}$ index with respect to $\boldsymbol{\tau^2}$ is stated by [12], who state that $\mathbf{I^2}$ indices obtained from meta-analysis with varying numbers of studies and varying effect metrics are directly comparable. Along with this descriptive interpretation of the $\mathbf{I^2}$ index, [12] proposed a confidence interval that could be used in the same way that the Q test is used for assessing heterogeneity in meta-analyses. Therefore, if the confidence interval regarding $\mathbf{I^2}$ includes the value of 0%, then the meta-analyst can support the hypothesis of homogeneity. If, on the other hand, the confidence interval does not include the value of 0%, there is no evidence of the existence of true heterogeneity. The use of the $\mathbf{I^2}$ index and its confidence interval is similar to the Q test application. Because the $\mathbf{I^2}$ index assesses not only the heterogeneity in the meta-analysis but also the degree of heterogeneity, it should be a more advisable procedure than the **Q test** to assess whether true heterogeneity exists between studies in a meta-analysis.

2.2. Biplot Method

The biplot method is a low-dimensional graphical representation of a data matrix X (with *I* individuals and *J* variables), i.e., a multivariate matrix; for more detail see [7]. The biplot is based on the singular value decomposition in vectors—i.e., it tries to reproduce the data incorporating a simultaneous representation of individuals and variables. The two most important representations proposed by [7] were the GH and JK biplots. The former offers a high quality in the representation of the columns (variables), whereas the latter offers a high representation of the rows (individuals). [8] named this type the HJ biplot and demonstrated that it is possible to represent rows and columns simultaneously on the same low-dimensional space, with a high-quality representation for individuals and variables.

The biplot method is useful for the visual inspection of data matrices because it allows the collection of atypical patterns and values, which can be graphically represented in a manner analogous to principal component analysis and discriminant analysis. Because of these advantages of the biplot method, the HJ biplot will be used to analyze the heterogeneity of the Audit C data.

Measures such as factor contributions to the element [8] will be concepts employed when analyzing the results. The proximity between individuals (rows) is interpreted as similarity; the angles formed by vector variables (columns) is interpreted as correlation, and the proximity between row and column markers is interpreted as preponderance.

2.3. Meta-analysis of diagnostic tests

When original studies that evaluate the quality of a test producing binary results are available, performing a diagnostic meta-analysis has become an important tool for investigating the information available on a test [6], [11], [19] and [20]. In a primary diagnostic study, the quality of a diagnostic test is often measured in terms of the sensitivity (TP rate) and the specificity (TN rate = 1 − FP rate) of the test; that is, parallel to a standard gold procedure, which defines the presence of a certain condition, the diagnostic test is performed, after which the sensitivity and specificity can be calculated.

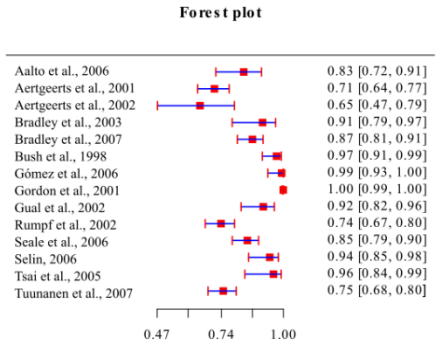
From the sensitivity and specificity, summary graphs can be made that show the variability between the studies. Thus, we have the following tools: **(1)** forest plot, which describes the sensitivity and specificity of each study with their respective confidence intervals; **(2)** crosshair, which demonstrates the bivariate relationship and degree of heterogeneity between sensitivity and FP rate; and **(3)** ROCellipse, which shows a region of confidence describing the uncertainty of the pair of sensitivity and specificity of each study.

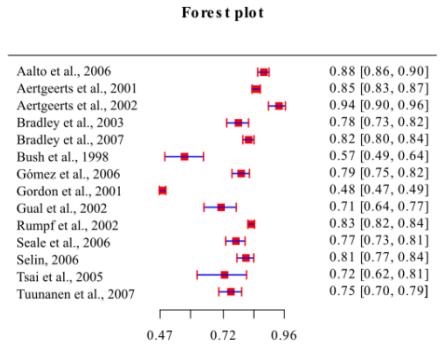
2.4. Heterogeneity in meta-analysis diagnosis

The observed sensitivities and specificities can be expected to vary in the original studies that constitute a meta-analysis. This is because of two main reasons:
1. Different authors will calibrate a test differently. Given a questionnaire score or a biomarker level, an investigator will have to decide which minimum value (or maximum value) should give a positive test result. This value is known as the cutoff point. Sometimes, especially in screening for rare diseases, a cutoff value will be set to reach a certain level of sensitivity, such as 95%, which often leads to a small specificity; however, a certain type of relationship exists between sensitivity and specificity. Both approaches result in a different cutoff points. In general, a specific calibration of the population directed to a certain level of sensitivity will give rise to different cutoff points for different populations.
2. When a diagnostic test is applied to several populations, different sensitivities and specificities can be expected even if the same cutoff value is used.
In a diagnostic meta-analysis, the quality of a diagnostic test is evaluated by integrating data from individual studies, which usually include sensitivities and specificities; some of these original studies may not report the cutoff point. This challenge—i.e., nonhomogeneity and typically unknown cutoff points—is known as the cutoff point issue. Therefore, the bivariate nature of the data must be preserved, modeling sensitivity and specificity together. During the last decade, two models have been established: a hierarchical model [17] and a bivariate model [16]. However, two groups of researchers independently demonstrated [1] and [9] that hierarchical and bivariate models are equivalent in special cases (and most common) in the absence of covariance.

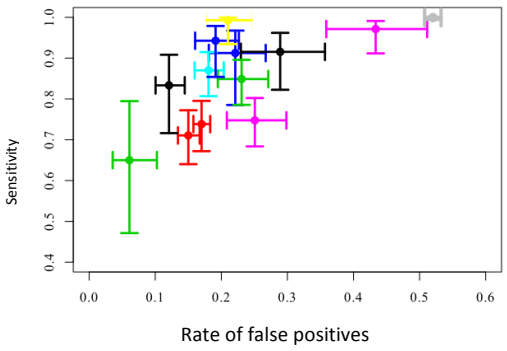
3. RESULTS

In this review, we included the Audit C data, which consists of data from14 studies that evaluated 18,332 individuals, out of which 2071 had alcohol problems. Figure 1 shows the variation of studies in the sensitivities and specificities, despite the fact that the specificity was perfect (100%) in two studies, as shown in Figure 1 (a) and Figure 1 (b). This variability can be observed in the graphs in Figure 1 (c) and Figure (d) as crosshair and ROCellipse, respectively.

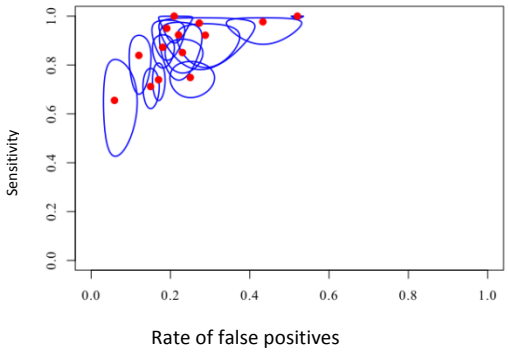




a. Forest plot for sensitivity **b. Forest plot for specificity**



c. Weighted crosshair graph



d. Graph, regions of confidence in the form of ellipse

Figure 1. Forest plot, crosshair and ROCellipse graphs, Audit C data

Table 2 shows the measures that corroborate the previous graphical analysis. Thus, there is a correlation between the sensitivity and the FP rate of 0.677, which denotes a significant relationship between both measures and a clear problem regarding the cutoff points in the studies. The true heterogeneity between studies, denoted by τ^2, has a value of 0.311. I^2 is 33.955%, which indicates the percentage of heterogeneity among the studies because of the true variability (τ^2) between studies. In addition, the Q-statistic is available with a p-value of 0, which means that the null hypothesis of homogeneity between the studies is rejected in favor of the alternative hypothesis that there is heterogeneity among the studies analyzed.

Table 2. Correlation values, τ^2, I^2, and Q of Audit C data

Measure-statistic	Value	Lower limit (95%)	Higher limit (95%)	p-value
Correlation between sensitivity and specificity	0.677	0.228	0.888	-
τ^2	0.311	0.00	3.787	-
Q	4177.192	-	-	0
I2	33.955	-	-	-

By performing the HJ biplot analysis on the data matrix described above, the following absorption of inertia (or variability explained) was obtained (see Table 3).

Table 3. Inertia absorbed by the two main axes in the HJ biplot representation, Audit C data.

Axis 1	Axis 2	Accumulated
66.46	28.69	95.15

As can be seen in Table 3, the accumulated absorption with the first two main axes is 95.15%, with a clear differentiation between axes 1 and 2. This leads us to interpret the first main plane (axes 1 and 2) together.

Figure 2 shows the simultaneous representations in the first main plane of the analysis performed. The number following the name of the author corresponds to the year that each study was conducted. Figure 2 shows the studies that form three clusters. Thus, the studies by Seale et al., 2006, Bradley et al., 2007, Aalto et al., 2006, Aertageerts et al., 2002, Bradley et al., 2003, Gual et al., 2002, Selin, 2006, Gómez et al., 2006, Tsai et al., 2005, and Bush et al., 1998 belong to the first cluster, i.e., cluster 1 (green); the studies by Aertgeerts et al., 2001, Rumpf et al., 2002 and Tuunanen et al., 2007 belong to cluster 2 (red); finally, the study by Gordon et al., 2001 belongs to cluster 3 (blue). Note that the studies within a cluster have common characteristics among themselves, but that there is heterogeneity among the clusters.

The following tables of Relative Contribution of the Factor to the element, which show the value first for the rows (studies; Table 4) and then the columns (variables; Table 5), indicate the elements that are characteristic of axes 1 and 2. Note that these values of contributions come in a scale of 0–1000. Therefore, when an element that receives a high contribution of the axis 1 and a low contribution from the rest is an exclusive characteristic of that axis, and its interpretation will be done with respect to this axis. Values higher than a scale of 500 will be considered for our analysis. Thus, studies by Gordon et al., 2001, Aertgeerts et al., 2002, Gual et al., 2002, Tsai et al., 2005, Bradley et al., 2003, Aalto et al., 2006, Aalto et al., 2006, Selin, 2006, and Bush et al., 1998 are representative of axis 1 as well as the variables FP, TP and TN. On the other hand, the

studies by Aertgeerts et al., 2001, Tuunanen et al., 2007, Rumpf et al., 2002, Bradley et al., 2007 are representative of axis 2 as well as the variable FN.

Table 4. Relative contributions of the factor to the element-studies

Contributions rows - studies			
Studies	**Axis 1**	**Studies**	**Axis 2**
Gordon et al., 2001	937	Aertgeerts et al., 2001	946
Aertgeerts et al., 2002	899	Tuunanen et al., 2007	636
Gual et al., 2002	677	Rumpf et al., 2002	621
Tsai et al., 2005	635	Bradley et al., 2007	595
Bradley et al., 2003	629	Gómez et al., 2006 et al., 2007	549
Aalto et al., 2006	589	Bush et al., 1998	470
Selin, 2006	523	Selin, 2006	423
Bush et al., 1998	517	Tsai et al., 2005	362
Gómez et al., 2006	398	Bradley et al., 2003	347
Seale et al., 2006	335	Gual et al., 2002	320
Rumpf et al., 2002	274	Seale et al., 2006	170
Aertgeerts et al., 2001	53	Aalto et al., 2006	109
Tuunanen et al., 2007	31	Aertgeerts et al., 2002	85
Bradley et al., 2007	3	Gordon et al., 2001	61

Table 5. Relative contributions of the factor to the element-variables

Contributions rows - variables			
Variables	**Axis 1**	**Variables**	**Axis 2**
FP	940	FN	965
TP	922	TN	103
TN	787	FP	53
FN	9	TP	27

In Figure 2, it can be seen that there is atypical data corresponding to the study by **Gordon et al., 2001**, which is characterized by very high FP and TP values.

The horizontal gradient (axis 1) exclusively marks the difference between the study by **Gordon et al., 2001** and all the others, because the study by Gordon et al. shows high PF and TP values. The vertical gradient (axis 2) differentiates cluster 2 (red) studies (Aertgeerts et al., 2001, Rumpf et al., 2002 and Tuunanen et al., 2007) from cluster 1 (green) studies (Bush et al., 1998, Aertgeerts et al., 2002, Gual et al., 2002, Bradley et al., 2003, Tsai et al., 2005, Selin, 2006, Seale et al., 2006, Gómez et al., 2006 and Aalto et al., 2006, Bradley et al., 2007).

All cluster 2 (red) studies have very high FN values, but among them they differ by the values in the other three variables. Cluster 1 (green)- studies have very low values in all variables.

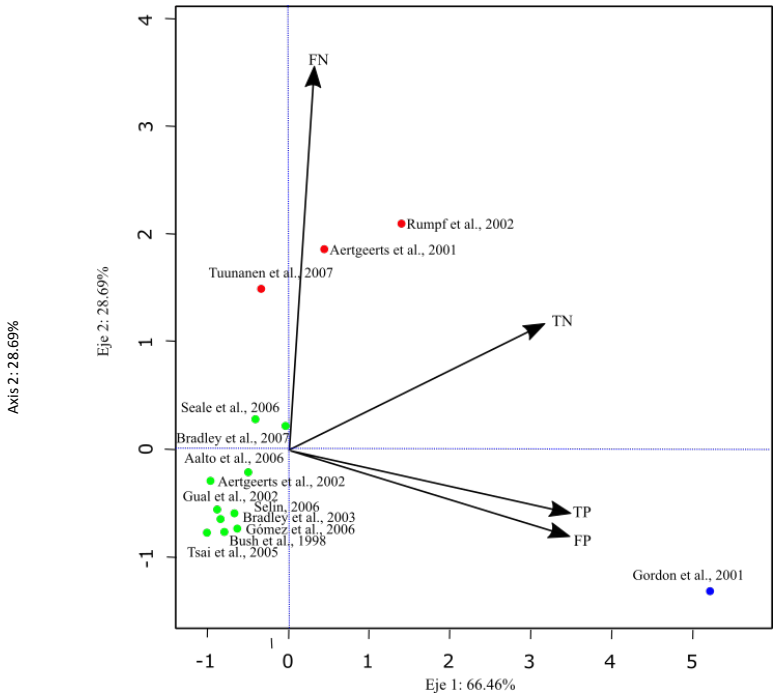


Figure 2. HJ biplot of Audit C data.

3. DISCUSSION AND CONCLUSIONS

• In this research, we analyzed the heterogeneity of the studies in meta-analysis of diagnostic tests using the Q-statistic, I^2 index, and HJ biplot. The evaluation of heterogeneity in a meta-analysis is a crucial question, because the meta-analyst's decision to select the statistical model to be applied (fixed-effects model versus random-effects model) depends on the homogeneity test employed [13]. Considering the importance of this question, the purpose of this study is to compare the behavior of the **Q**-statistic and the $\mathbf{I^2}$ index, and to complement the information they provide with the results obtained using the HJ biplot to evaluate the heterogeneity between cluster of individual studies in a meta-analysis.

• The $\mathbf{I^2}$ index behaves similarly to the **Q** test from an inferential point of view [14], but the $\mathbf{I^2}$ index has advantages over the classical **Q** test—it is easily interpretable because it is a percentage and does not depend on degrees of freedom [4]. Another advantage is that it provides a way of assessing the magnitude of heterogeneity, whereas the **Q** test reports on the statistical significance of the homogeneity hypothesis [4]. However, these two tests do not provide information about the possible causes of heterogeneity; that is, they do not indicate whether studies are similarly responsible for the heterogeneity or which variables used in the individual studies are responsible for the heterogeneity.

• The HJ biplot method [8] provides a graphical representation of the studies and the TP, TN, FP, and FN in the same system of Cartesian axes with minimal information loss. This representation provides information on the overall structure of the studies that is complementary to that provided by Higgins' Q and I^2 statistics. The length of the vectors representing the variables informs the variability of the variables; thus, analyzing this length can detect which variables are responsible for the heterogeneity of a meta-analysis. The cosines of the angles between the vectors show the correlation between the variables they represent, thus providing useful information when choosing the appropriate model for the available data; if there is a strong correlation between FP and TP, a clear cutoff point issue becomes evident [5]. Thus, it is necessary to model these measures using a hierarchical approach [17] or a bivariate approach [16]. The points represent the different studies to be integrated into the meta-analysis of a diagnostic test. The distances between them allow the visualization of clusters of studies with homogenous characteristics. Those clusters that are further away are responsible for the heterogeneity. Thus, by projecting such studies orthogonally on the direction of the TP, TN, FP and FN, the causes of heterogeneity can be obtained.

- The HJ biplot has proven to be a multivariate tool that is extremely useful in the analysis of meta-analysis data of diagnostic tests, both in the descriptive phase and in the search for the causes of variability. The points in Figure 2 display the different studies to be integrated into the meta-analysis of a diagnostic test.
- In conclusion, the use of the HJ biplot in the meta-analysis of diagnostic tests allowed us the following: (i) characterization of the heterogeneity of the studies regarding measures such as TP, FP, TN, and FN; (ii) identification of groups of homogeneous studies; and (iii) analysis of the relations of these measures to choose the most appropriate model that facilitates the subsequent integration of the individual characteristics of the studies analyzed.

RECEIVED: JULY, 2017
REVISED: JANUARY, 2018

REFERENCES

[1] ARENDS, L., HAMZA, T., VAN HOUWELINGEN, J., HEIJENBROK-KAL, M., HUNINK, M., & STIJNEN, T. (2008). Bivariate random effects meta-analysis of roc curves. **Medical Decision Making**, 28, 621-638.

[2] COCHRAN, W. G. (1954). The combination of estimates from different experiments. **Biometrics**, 10, 101-129.

[3] CORNWELL, J. M. (1993). Monte Carlo comparisons of three tests for homogeneity of independent correlations. **Educational and Psychological Measurement**, 53, 605-618.

[4] DEEKS, J. J., ALTMAN, D. G., & BRADBURN, M. J. (2008). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. **Systematic Reviews in Health Care: Meta-Analysis in Context**, **Second Edition,** 285-312.

[5] DOEBLER, P., HOLLING, H., & BÖHNING, D. (2012). A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. **Psychological Methods**, 17, 418.

[6] EGGER, M., DAVEY-SMITH, G., & ALTMAN, D. (Eds.). (2008). **Systematic reviews in health care: meta-analysis in context**. John Wiley & Sons, N. York.

[7] GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. **Biometrika**,*58*, 453-467.

[8] GALINDO VILLARDÓN, M. (1986). Una alternativa de representación simultanea: Hj-biplot. **Qüestiió. quaderns d'estadística i investigació operativa**, 10(1), 13-23.

[9] HARBORD, R. M., DEEKS, J. J., EGGER, M., WHITING, P., & STERNE, J. A. (2006). A unification of models for meta-analysis of diagnostic accuracy studies. **Biostatistics,** 8, 239-251.

[10] HARDY, R. J., & THOMPSON, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. **Statistics in Medicine**, 17, 841-856.

[11] HASSELBLAD, V., & HEDGES, L. V. (1995). Meta-analysis of screening and diagnostic tests. **Psychological bulletin**, 117(1), 167.

[12] HIGGINS, J., & THOMPSON, S. G. (2002). Quantifying heterogeneity in a meta-analysis. **Statistics in medicine**, 21, 1539-1558.

[13] HIGGINS, J. P., THOMPSON, S. G., DEEKS, J. J., & ALTMAN, D. G (2003). Measuring inconsistency in meta-analyses. **BMJ*:* British Medical Journal**, 327, 557.

[14] HUEDO-MEDINA, T. B., SÁNCHEZ-MECA, J., MARÍN-MARTÍNEZ, F., & BOTELLA, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index?. **Psychological methods**, 11, 193.

[15] KRISTON, L., H ÖLZEL, L., WEISER, A.K., BERNER, M. M., & H ÄRTER, M. (2008). Meta-analysis: are 3 questions enough to detect unhealthy alcohol use?. **Annals of Internal Medicine**, 149, 879-888.

[16] 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, 982-990.

[17] RUTTER, C. M. & GATSONIS, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. **Statistics In Medicine**, 20, 2865-2884.

[18] SACKETT, P. R., HARRIS, M. M., & ORR, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error.

[19] SCHULZE, R. E., HOLLING, H. E., & BÖHNING, D. E. (2003). Meta-analysis: **New Developments and Applications in Medical and Social Sciences**. Hogrefe & Huber Publishers.
[20] SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A., & SONG, F. (2003). Methods for meta-analysis in medical research. **Statistics in Medicine**, 22(19), 3111-3114.

Appendix

```
install.packages("mada")
library("mada")
data("AuditC")
AuditC

forest(madad(AuditC), type = "sens")
forest(madad(AuditC), type = "spec")

rs <- rowSums(AuditC)
weights <- 4 * rs / max(rs)
crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1),col = 1:14, lwd = weights)

ROCellipse(AuditC, pch = "")
points(fpr(AuditC), sens(AuditC))

library(tcltk)
library(tcltk2)
library(rgl)
library(tkrplot)
library(shapes)
library(cluster)
library(dendroextras)
library(MASS)
tclRequire(BWidget)
library(biplotbootGUI)
biplotboot(AuditC)
```