

## Review Article

# Modeling for Meta-Analysis of Diagnostic Tests of Binary Responses: An Approach from Copulas and Hierarchical Models

Pambabay Calero Johnny J<sup>1</sup>, Bauz Olvera Sergio A<sup>1</sup>, Sánchez García Ana B<sup>2</sup>, Nieto-Libreros Ana B<sup>3</sup> and Galindo Villardón P<sup>3</sup>

<sup>1</sup>ESPOL, Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, Guayaquil, Ecuador

<sup>2</sup>INICO, University of Salamanca, Spain

<sup>3</sup>Department of Statistics-University of Salamanca, Spain

**\*Corresponding author**

Sánchez-García Ana B, Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, Guayaquil, Ecuador University of Salamanca, Spain, Tel: +34923294630; Email: asg@usal.es

**Submitted:** 08 August 2020

**Accepted:** 11 September 2020

**Published:** 16 September 2020

**ISSN:** 2333-7109

**Copyright**

© 2020 Pambabay Calero JJ, et al.

OPEN ACCESS

**Keywords**

- Copula function
- Binary covariance
- Bayesian hierarchical model
- Markov chain monte carlo
- Meta-analysis
- Diagnostic test

**Abstract**

Current statistical procedures implemented in statistical software packages for the aggregation of diagnostic test accuracy data include hierarchical regression and the bivariate random-effects meta-analysis model. However, these models do not report the overall mean, but rather the mean of a central study and have differences in estimating the correlation between sensitivity and specificity when the number of studies in the meta-analysis is small and/or when the variance between studies is relatively large.

Generally, to capture the possible dependence between diagnostic test results, a binary covariance structure is assumed, but in this article, we consider the use of structures that can be modeled using copula functions as an alternative to dependence modeling with software that is readily available. A posteriori statistics of interest are obtained using Markov Chain Monte Carlo. The results obtained using our approach are compared with those obtained using models that assume bivariate structure and with those obtained using models under the assumption of independence between test results for clinical diagnosis. To illustrate the application of the method and to make comparisons, data from a study published in the literature were used.

**ABBREVIATIONS**

MCMC: Markov Chain Monte Carlo; AUC: Area Under the Curve; HSROC: Hierarchical Summary Receiver Operating Characteristic; BRMA: Bivariate Random-effects Meta-analysis; Se: Sensitivity; Sp: Specificity; TP: True Positives; FP: False Positives; FN: False Negatives; TN: True Negatives; AUDIT-C: Alcohol Use Disorder Identification Test - Consumption

**INTRODUCTION**

The exponential growth of medical literature and the increasingly widespread use of information and communication technologies, together with the dispersion of scientific literature, make it difficult for researchers, and above all health professionals, to access relevant information. Meta-analysis is a systematic review that incorporates a statistical strategy to integrate the results of several studies into a single estimate.

The correct diagnosis of an alteration is of eminent and principal interest in psychology and medicine. Diagnostic tests are common tools to discern the presence or absence of a condition or to examine patients who are at risk of developing a disease.

Many diagnostic tests are based on scores obtained from brief questionnaires or are based on a single biomarker. Therefore, they will not always give a correct diagnosis. When primary studies are available that evaluate the quality of a diagnostic test, performing a diagnostic meta-analysis has become a key tool for investigating available information about a diagnostic test [1,2]. In a primary diagnostic study, the quality of a diagnostic test is often measured in terms of the sensitivity (true positive rate) and specificity (true-negative rates = 1 - false-positive rate) of the test. In order to be able to operate with likelihood ratios in the calculation of probabilities, these must be transformed into advantages (odds), these calculations are simplified by using nomograms [3]. Sensitivity and specificity vary depending on the cut-off point used to separate the presence or absence of the condition of interest. In these cases, the results can be represented graphically as a ROC curve, which allows the characteristics of the test to be known according to different cutting points and which can be used to choose the most appropriate one. An estimator of the overall relevance of a test can be the Area under the Curve (AUC). It must be borne in mind that the information available on the validity of diagnostic tests comes from different populations.

Therefore, the estimates obtained in these studies are subject to random variability and if the studies have been designed incorrectly, to biases. The variability of the measurements will be influenced by multiple factors that are of interest to know and control. Among them, it is especially important to distinguish the intra and inter-study variations.

Due to this particularity, the bivariate nature of the data must be preserved, modeling sensitivity and specificity together. Two models have been established in recent years: a hierarchical model [4], a bivariate model [5]. These models, when focusing on estimating meta-analytic sensitivities and specificities, has the advantages of accessing the individual data and of allowing unexplained heterogeneity as well as correlation between sensitivity and specificity. Moreover, it can be generalized to model covariates and can cope with extreme values of 100% for sensitivity and specificity without applying artificial continuity corrections, and standard software (e.g., R) can be used for analysis. However, there are also some disadvantages of these models. It does not operate on the original scale of sensitivity and specificity, but on the corresponding logit scale and by generally relying on the bivariate normal distribution for the random effects, it only allows one single correlation structure. Recently, a mixed copula model has been proposed as an extension of the generalized linear mixed model-GLMM [6], using a copula representation of the distribution of random effects with normal marginal and beta, respectively.

We propose here a copulas approach for the meta-analysis for diagnostic accuracy studies, which avoids the previously mentioned problems, while keeping all advantages of the Hierarchical Summary Receiver Operating Characteristic (HSROC) and Bivariate Random-Effects Meta-Analysis (BRMA) models. It uses the idea of having marginal beta distributions for sensitivity and specificity, resulting in corresponding marginal beta-binomial distributions for true positives and true negatives, and linking the marginal by copula distributions. Copulas are a useful way to model multivariate data as they explain the dependence structure between sensitivity and specificity by providing a flexible representation of a multivariate distribution. The theory and application of copulas have become important in finance, insurance, and other areas. In this article, with the help of the statistical software R [7] (R Development Core Team 2013), we compare the performance measures of the Bayesian approach against the classic bivariate model [5] using the copulation functions.

## STATISTICAL METHODS

In general, meta-analysis is a two-stage process. In a first step, the results of each study are estimated, although, in the case of the evaluation of diagnostic tests, each study is summarized by a pair of indices that describe the validity of the test.

Typically, these two indices are either sensitivity and specificity, or the positive and negative likelihood ratios. In a second step, global validity indices must be calculated for which various methods have been proposed, which will be explained later. In diagnostic tests, the assumption of methodological homogeneity in the studies are not met and it is therefore important to evaluate heterogeneity.

## Evaluation of heterogeneity

Assessing the possible presence of statistical heterogeneity in the results can be done (in a classical way) by presenting the sensitivity and specificity of each study in a forest plot. In these graphs, the estimators of the indices are represented together with their confidence intervals, and they are usually presented paired and ordered according to one of the indices. A characteristic source of heterogeneity is that which arises because the included studies may have used different thresholds to determine what a positive result is. This effect is known as the threshold effect.

## Threshold effect

To explore this source of variation it is useful to represent in a graph the pairs of sensitivity and specificity of each study in a ROC plane. In this plane, the zone closest to the upper left corner assumes good diagnostic performance while the central zone, the diagonal in which sensitivity and specificity are equal, represents a null diagnostic capability. If there were a threshold effect, the points in the ROC space would reflect a curvilinear curve. Changing the positivity threshold of a test would result in a higher (or lower) sensitivity with the consequent opposite effect on specificity.

The most robust statistical methods proposed for meta-analysis take into account this correlation between sensitivity and do so by estimating a summary ROC curve (SROC) of the included studies. However, on limited occasions the results of the primary studies are homogeneous and the presence of both threshold effect and other sources of heterogeneity can be ruled out. In this situation, the overall result of the review could be obtained from the weighted combination of the indices of the individual studies. As always, this combination can be done using either a fixed-effect model or a random-effects model, depending on the magnitude of heterogeneity. Several statistical methods for estimating the SROC curve have been proposed. The first, proposed by [8] (Moses, Shapiro & Littenberg 1993), is based on estimating a linear regression between two variables created from the validity indices of each study.

## Bivariate and hierarchical approach

The Moses model has some limitations. On the one hand, it does not take into account the different precision with which sensitivity and specificity were estimated in each study, nor does it incorporate heterogeneity between studies. To overcome these limitations, more complex regression models have been proposed. The first is a bivariate random-effects model [5] (BRMA), which assumes that the logit of sensitivity and specificity follow a bivariate normal distribution.

The model contemplates the possible correlation between both indices, models the different precision with which sensitivity and specificity have been estimated, and incorporates a source of additional heterogeneity due to variance between studies. At the study level, this model assumed that the true positives and false positives within study  $i, i = 1, \dots, I$ , follow binomial distributions. At the between-studies level, a bivariate random effects model is assumed for  $\text{logit}(Se_i)$  and  $\text{logit}(Sp_i)$ , where normal priors are assumed for the study-specific parameters:

$$\begin{pmatrix} \text{logit}(Se_i) \\ \text{logit}(Sp_i) \end{pmatrix} \sim \text{Normal} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right) \tag{1}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \tag{2}$$

The second model refers to the Hierarchical Summary Receiver Operating Characteristic (HSROC) or hierarchical model [4]. It is similar to the previous model, except that it makes explicit the relationship between sensitivity and specificity across the threshold. The model included random effects for the cut-off and the test accuracy and focused on estimating the SROC curve. At the study level, it is assumed that within study  $i, i = 1, \dots, I$ , the True Positives (TP) and the False Positives (FP) follow binomial distributions:

$$\begin{aligned} TP &\sim \text{Binomial}(n_{i1}; Se_i) \\ FP &\sim \text{Binomial}(n_{i2}; Sp_i) \end{aligned} \tag{3}$$

where index 1 denotes diseased individuals and index 2 denotes non-diseased. The authors parameterized the sensitivities and specificities as follows:

$$\begin{aligned} \text{logit}(Se_i) &= \left( \theta_i + \frac{\alpha_i}{2} \right) e^{-\beta/2} \\ \text{logit}(1 - Sp_i) &= \left( \theta_i - \frac{\alpha_i}{2} \right) e^{\beta/2} \end{aligned} \tag{4}$$

where  $\theta_i$  is the random threshold in study  $i$ ,  $\alpha_i$  is the random accuracy in study  $i$ , and  $\beta$  is a shape (asymmetry) parameter. Normal distributions are used to model variation in the study-specific parameters across studies:

$$\begin{aligned} \theta_i &\sim \text{Normal}(\theta; \tau_\theta^2) \\ \alpha_i &\sim \text{Normal}(\lambda; \tau_\lambda^2) \end{aligned} \tag{5}$$

The BRMA and HSROC are adjusted using the likelihood method, assuming that the transformed data are approximately normal with constant variance; however, for sensitivity and specificity and proportions in general, the values of mean and variance depend on the underlying probability.

Both hierarchical models involve statistical distributions at two levels. At the lower level, the count of the values taken from the tetrachoric tables extracted from each study is modeled using the binomial distributions and the log-odds of the proportions. At the upper level, it is assumed that the random effects of the study explain the heterogeneity in the accuracy of the diagnostic test among studies beyond what is explained by the variability of sampling at the lower level.

The Bivariate model and the HSROC model are mathematically equivalent when covariates are not available [9,10] but differ in their parameterizations. Bivariate parameterization models sensitivity, specificity, and the correlation between them directly, while HSROC model parameterization models positivity, precision, and curve shape threshold functions to define an SROC curve. Therefore, any factor affecting the probability will change the mean and variance. This implies that models in which predictors affect the mean but assume a constant variance will not be adequate.

A third model is a mixed copula model that is an extension of the general linear mixed model of [6], using rather a copula representation of the distribution of random effects with normal and beta marginal distributions [11]. (Nikolouloupoulos 2015) demonstrates that the meta-analysis of precision studies of the diagnostic test is a primary field of application for copula models since the traditional assumption of multivariate normality is invalid in this context. Joint modelling of study-specific sensitivity and specificity using existing bivariate or copula-based beta distributions overcome the difficulties mentioned above. Since both sensitivity and specificity take values in the interval space (0,1), it is a more natural option to use a beta distribution to describe their distribution across studies, without the need for any transformation. The beta distribution is conjugated with the binomial distribution and therefore it is easy to analytically integrate the random effects giving rise to the marginal beta-binomial distributions.

There are a variety of statistical packages that can be used to perform the analyses described, such as SAS and Stata, which through a series of user-programmed macros make it easier to obtain the models described. The macros of Stata are called Midas, and the macro developed for SAS is called MetaDas [12]. Other specific programs for meta-analysis of diagnostic test studies are the mada, CopulaREMADA [11] (Nikolouloupoulos 2015) and HSROC libraries of the R statistical package.

The methods below are demonstrated using a data set previously published by [13], which analyzed alcohol problems in certain individuals, considering both the 10-item AUDIT and its 3-item abbreviated version (Alcohol Use Disorder Identification Test - Consumption [AUDIT-C]) accurately detect unhealthy drinking, to examine whether the AUDIT-C is as accurate as the full AUDIT in detecting unhealthy alcohol use in adults. The study estimated sensitivity and specificity values around 0.86 and 0.78 respectively.

Fourteen studies evaluating individuals with 18 332, of whom 2 071 had alcohol problems, were included in the review. The prevalence of studies ranged from 5% to 37%, with a median of 11%, (Table 1).

### A statistical method for meta-analysis of diagnostic tests from a copula approach

Copula is a well-known statistical concept for modeling dependence on random variables. A copula is a joint distribution function whose marginal are all uniformly distributed and can be used to model the dependence separately from the marginal distributions, i.e., it is a function that approximates the set behavior of random variables from their individual behaviors [14]. Two random variables  $x_1$  and  $x_2$  with distribution function  $f(x_1)$  and  $f(x_2)$ , (Sklar 1959) demonstrated that there is a C function can be written as:

$$H(x_1, x_2) = C(F(x_1), F(x_2)) = C(u_1, u_2) \tag{6}$$

where  $C(u_1, u_2)$  is itself a distribution function for a bivariate pair of uniform random variables and  $H(x_1, x_2)$  is a distribution function for the original variables  $x_1$  and  $x_2$ . Therefore, a bivariate copula is a function that fulfills the following properties [15] (Nelsen 2007):  $c(u, 0) = 0, C(0, v) = 0, C(u, 1) = u$  and  $C(1, v) = v$ .

**Table 1:** AuditC Data.

ID	TP	FP	TN	FN
1	47	101	738	9
2	126	272	1543	51
3	19	12	192	10
4	36	78	276	3
5	130	211	959	19
6	84	68	89	2
7	68	112	423	0
8	752	3226	2977	0
9	59	55	136	5
10	142	571	2788	50
11	137	107	358	24
12	57	103	437	3
13	34	21	56	1
14	152	88	264	51

The concept of copula allows modeling the dependence between marginal distributions

through the copula parameters. That is to say, in the present paper we will use only a parametric model, despite the fact that there are non-parametric copulas that can be used in order to model the dependence between random variables.

These association parameters can be transformed in correlation measurements, such as the Spearman or Kendall correlation coefficients. [16] (Schweizer, Wol et al. 1981), showed that both measurements can be used for the description of dependence.

**The Hierarchical Copula Model**

It is assumed that each individual study ( $i = 1, \dots, J$ ) in the meta-analysis reports a 2x2 table with values of  $TP_i, TN_i, FP_i$  and false negatives  $FN_i$  and also with summary measures established by the sensitivity  $Se_i$  and specificity  $Sp_i$ . It is further assumed that  $TP_i$  and  $TN_i$  are distributed according to a binomial distribution, [17], see Equation 7.

$$\begin{aligned}
 TP_i &\sim \text{Binomial}(TP_i + FN_i; Se_i) \\
 TN_i &\sim \text{Binomial}(TN_i + FP_i; Sp_i)
 \end{aligned}
 \tag{7}$$

The prior marginal distributions for sensitivity and specificity are assumed as beta distributions with parameters;  $\alpha$  and  $\beta$ , i.e. making an assumption about Se and Sp in their original scale [0; 1], [17], see Equation 8.

$$\begin{aligned}
 Se_i &\sim \text{Beta}(\alpha_{se}, \beta_{se}) \\
 Sp_i &\sim \text{Beta}(\alpha_{sp}, \beta_{sp})
 \end{aligned}
 \tag{8}$$

The corresponding density function for the sensitivity is defined by the Equation 9.

$$f(\mu, \alpha_{se}, \beta_{se}) = \begin{cases} \frac{\tilde{\Delta}(\alpha_{se} + \beta_{se}) \mu^{\alpha_{se}-1}}{0} & ; 0 < \mu < 1 \\ \text{otherwise} & \end{cases}
 \tag{9}$$

Specificity is defined analogously, where  $\Gamma(\alpha)$  is the function gamma with  $\Gamma(n+1) = n!$  for  $n \in \mathbb{Z}^+$ . The expected  $E(u)$  values for the beta distributions are defined as  $\frac{\alpha_*}{\alpha_* + \beta_*}$  and describe

the meta-analytical parameters of interest (Se; Sp). In addition, the  $\frac{\alpha_* \beta_*}{(\alpha_* + \beta_* + 1)(\alpha_* + \beta_*)^2}$  variance estimates explain the variability due to the heterogeneity of the studies. The beta distribution is used due to its flexibility and the fact that it is conjugated to the binomial distribution.

To model the potential dependence between TP and TN, we apply the concept of copulas. In our case, a two-dimensional Cumulative Distribution Function (cdf) is constructed. Therefore, by Sklar's Theorem.

$$F(x_1, x_2) = C(F_{se}(x_1), F_{sp}(x_2))
 \tag{10}$$

The probability density function (pdf) is now determined simply by differentiating the joint distribution function. We get:

$$h(x_1, x_2) = f_{se} f_{sp} c_{12}(F_{se}(x_1), F_{sp}(x_2))
 \tag{11}$$

where  $c_{12}$  denotes the density of a bivariate copula. Density is treated as a plausibility function. Note that in the case of copula, the likelihood is determined analytically and has a closed form. Therefore, the standard methods of maximum likelihood are used for parameter estimation. An advantage of the copula model is that, in principle, there are a large number of copula that allow for different association structures. This is in contrast to the standard model where a bivariate normal distribution is used as a single correlation structure. There are a rich number of copula that can be used, each resulting in a new model for the meta-analysis of precision diagnostic studies. To be specific, we use the following copulas, the bivariate copula Gaussian, Clayton and Frank, these are copulas that can be adjusted to Equation 11 [11] (Nikoloulopoulos 2015).

**Selection of a Model Copula**

The aim is to identify among the copula families set out in previous section, which family or families best fit the joint distribution of  $Se_i$  and  $Sp_i$ . This is done using the Cramér-von Mises goodness-of-fit test [18] (for details of this test).

**Simulation Study**

A simulation study was carried out to compare the adjustment of the different copulas. The scenario specifications were designed to reproduce realistic situations found in the meta-analysis of diagnostic accuracy studies.

**Generation of simulated data and goodness-of-fit of copulas models**

Meta-analyses were investigated with different numbers of studies at random ( $k = 5; 10; 20; 35$ ). The size of a study in each meta-analysis,  $n_p$ , were randomly sampled from a uniform distribution,  $U(30; 1\ 255)$ ; given an underlying prevalence,  $p$ , individuals within each study were randomly classified as sick or undiagnosed, and a continuous test result value,  $x$ , was assigned and randomly sampled [19]. To determine the corresponding



tetrachoric tables, the package Copula Remada of the statistical program R was used, using the function rCopula REMADA. beta.

To create the 2x2 table for each study, individuals were classified as a true positive, false negative, false positive, or true negative based on the test result and disease status. The sensitivity and specificity of the studies was adjusted by a uniform distribution with parameters 0:8 and 1. We generated 21 011 studies in 10 58 independent meta-analysis datasets to allow a reliable analysis of the goodness of fit of previously analyzed models. The goodness-of-fit of copulas models was assessed by examining the test statistician’s estimates and the p-value of the statistical test proposed by [18].

**Adjusting The Hierarchical HSROC and Copula Model**

We will use the CopulaDta package [20] (Nyaga, Arbyn & Aerts 2016) to adjust the copulas analyzed in Section 2.5. The package facilitates the application of complex models and their visualization within the Bayesian framework with short execution times and provides functions of binomial beta bivariate distributions built as a product of two beta marginal distributions and copula densities discussed in the previous Section.

The HSROC package is used to estimate the parameters of a hierarchical summary receiver operating characteristic (HSROC) model allowing for the reference standard to be possibly imperfect, and assuming it is conditionally independent from the test under evaluation. The estimation is carried out using a Gibbs sampler. The package consists of 4 functions. The two main functions are HSROC and HSROCSummary. HSROC, which must be run first, is used to implement a Gibbs sampler while HSROC Summary produces summaries for the HSROC model parameters. The remaining 2 functions are secondary functions: sim data simulates a dataset based on the HSROC diagnostic meta-analysis

model, while beta. parameter returns the shape parameters of the  $Beta(\alpha; \beta)$  probability corresponding to a given range.

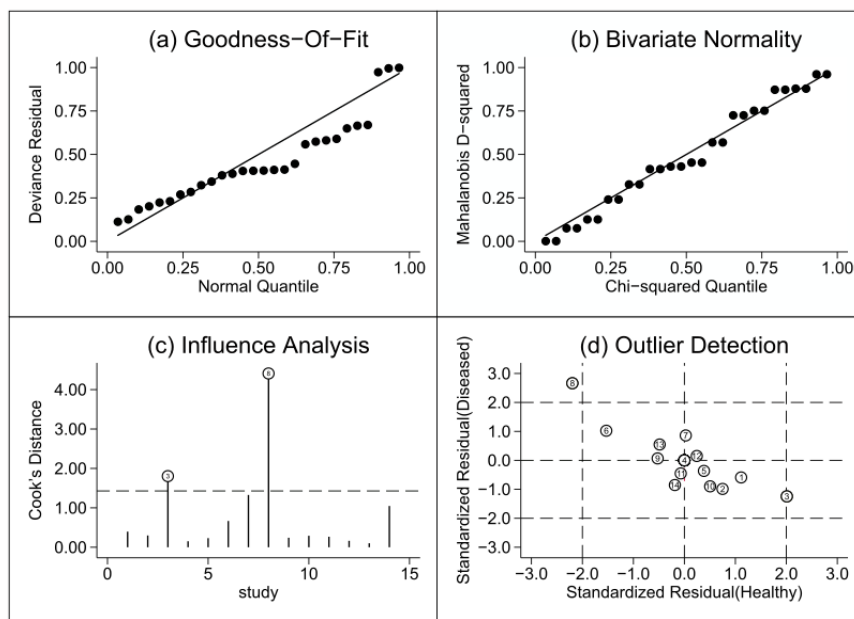
**Results of the adjustment to AuditC data and simulated data**

It is important to analyze the behavior and distribution of the data, before fitting any model, for our analysis, we will use statistical tests based on measures such as: chi-square, Cook distance and Mahalanobis, [21,22] Comrey 1985, Cook 1977, Rasmussen 1988). In that sense, Figure 1, shows that, there are influential studies in the BRMA modeling and that the goodness of the adjustment and bivariate normality is not adequate, particularly due to studies 3 and 8 from Table 1 (Figure 1).

On the other hand, Figure 2 shows how the different values of sensitivity and specificity are located below or above the dotted line (average), indicating a strong presence of heterogeneity in both measures. That is, the graph demonstrates an asymmetry of the test performance means towards higher sensitivity with lower specificity, providing indirect evidence of certain heterogeneity threshold variability (Figure 2).

As mentioned above, the HSROC and BRMA models are equivalent when there is no inclusion of covariates, so we will analyze the outputs produced by the HSROC package of R. The HSROC Summary function creates plots to help evaluate whether the Gibbs sampler has converged. Each trace plot is a scatter plot of the posterior sample of a single parameter vs the iteration number of the Gibbs sampler. The trace plots for the sensitivity and specificity are shown in Appendix 1. For our example, we can see that convergence seems to be achieved fairly quickly. Another graphical summary produced by the HSROC Summary function is the density plot.

It plots a smoothed posterior kernel density function for each parameter. Appendix 2 shows density plots for some of



**Figure 1** Goodness-of-t, bivariate normality, influential and atypical data.

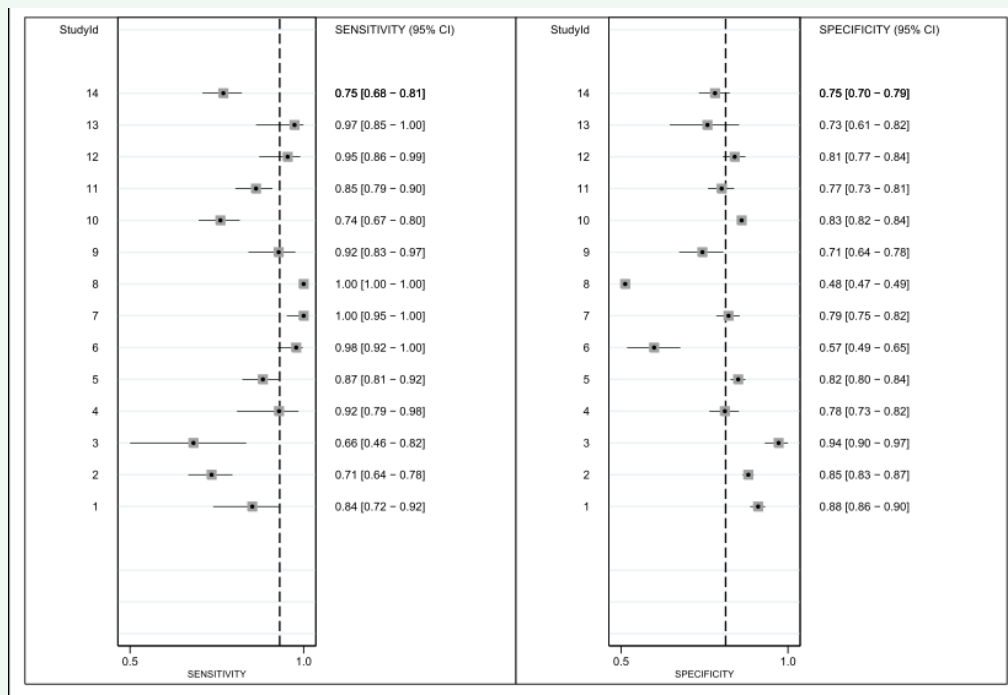


Figure 2 Foresplot.

the between-study parameters. The function also produces a summary receiver operating characteristic (SROC) curve.

The SROC curve summarizes the relationship between sensitivity and (1 - specificity) across studies, taking into account the between-study heterogeneity. The SROC curve for the Audit C data is shown in Figure 3, where, it can be noted that the values for sensitivity and specificity are: 0.430 CI95% (0-0.913) and 0.982 CI95% (0.734-1) respectively (Figure 3).

Individual studies are depicted by a clear circle. The radius of the circle is proportional to the sample size of the study. The black circle marks the pooled sensitivity and specificity across the 14 studies in this meta-analysis. The 95% prediction region is defined by the black dotted-curve. The black dot-dashed-curve marks the boundary of the 95% credible region for the pooled estimates of sensitivity and specificity across the 14 studies. Finally, the correlation between sensitivity and (1-specificity) is 0.854. Table 2 shows that Frank's copula is the most extreme, with a correlation of -0.600, although the Gauss and C90 models also have a high correlation whose values are -0.570 and -0.466 respectively. It is important to note that the sensitivity and specificity results for the Auditc data in the copula models (sensitivity between 0.85-0.87 and specificity between 0.75-0.76) are very close to the values provided by [13]. That is, the estimates made by copulas are more reliable than the estimates made by the HSROC model (very low sensitivity 0.430 and very high specificity 0.98) (Table 2) (Figure 4).

Since the data is not distributed as a bivariate normal, and the value of the correlation between performance measures is positive, modelling of dependency through copulas should be used. In that sense, you have the results for the values

of sensitivity, specificity and correlation  $\tau_{x,y}$ , table 2. The algorithms used can be seen in Appendix 3.

For each copula described above, the respective goodness-of-fit tests are performed. The results are shown in Table 3. To link marginal beta distributions with a bivariate dependent distribution, the best fit is observed when working with FGM copula. This is seen in the value of the statistic and p-value which leads to the acceptance of the hypothesis of equality of empirical and parametric distributions. For FGM copulation, the value of the statistic and the p-value are 0.022 and 0.903. Note that, the p-value for FGM copula is the highest value in relation to the other copulas Table 3.

Table 4 shows the average estimates with their respective 95% confidence intervals of the statistic and the p-value of the hypothesis test for each copula model according to the number of studies simulated for each meta-analysis. FGM copula best fits the simulated data for numbers of studies between 5 and 22, as the average value of the statistic is the lowest compared to the other copula models. Between 23 and 28 studies the best fit was achieved with Gaussian Copula and for studies between 29 and 35 the best fit was achieved with Clayton Copula (Table 4).

## CONCLUSION AND REMARKS

[In this paper, definition, implications and methodologies were presented for the development of models in the meta-analysis of diagnostic tests associated with copulas, applied to a particular data set, although [23,24] pointed out that it might be difficult to estimate copulas from a data set, especially for count data. This problem may not occur under the conditions indicated in the article, because the parameter  $p \in (0;1)$  is modeled from a continuous distribution. In general terms, copulas are functions

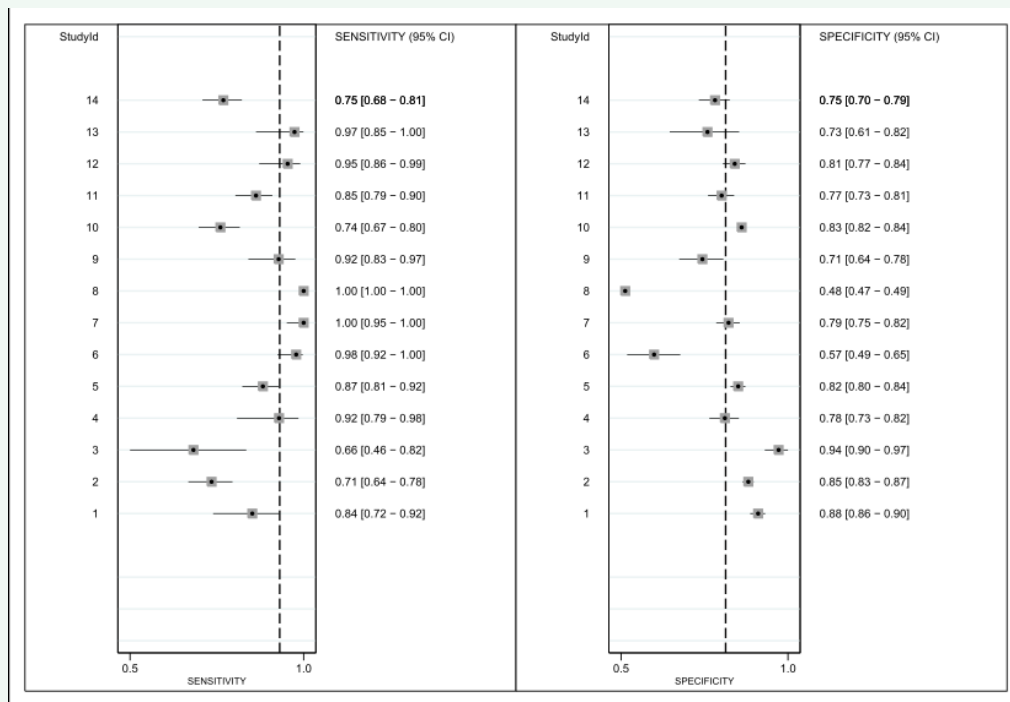


Figure 3 Foresplot.

Table 2: A posteriori mean, with 95% credibility interval, for the marginal means and the estimated correlation parameters for the different copulas for the AuditC data.

Copulas	Parameter	Mean	Lower	Upper
Gauss	Se	0.862	0.766	0.920
	Sp	0.755	0.6898	0.811
	correlation	-0.570	-0.799	-0.289
C90	Se	0.865	0.777	0.922
	Sp	0.760	0.695	0.813
	correlation	-0.466	-0.801	-0.0005
C270	Se	0.854	0.743	0.920
	Sp	0.752	0.680	0.810
	correlation	-0.324	-0.758	-2.219e-17
FGM	Se	0.871	0.780	0.932
	Sp	0.756	0.692	0.812
	correlation	-0.214	-0.222	-0.121
Frank	Se	0.858	0.754	0.929
	Sp	0.751	0.773	0.808
	correlation	-0.600	-0.767	1.000

Abbreviations: C90: Clayton 90; C270: Clayton 270; FGM: Farlie-Gumbel Morgenstern; Se: Sensitivity; Sp: Specicity

that approximate the set behavior of random variables, based on their individual (marginal) behaviors [14].

Sensitivity and specificity measurement models that make more realistic assumptions about marginal distribution functions (e.g., those that deviate from the assumption of normality), although they may be more expensive in computational terms, are best measured by summaries of a meta-analysis of diagnostic tests. In this sense, models derived from copula offer a structure

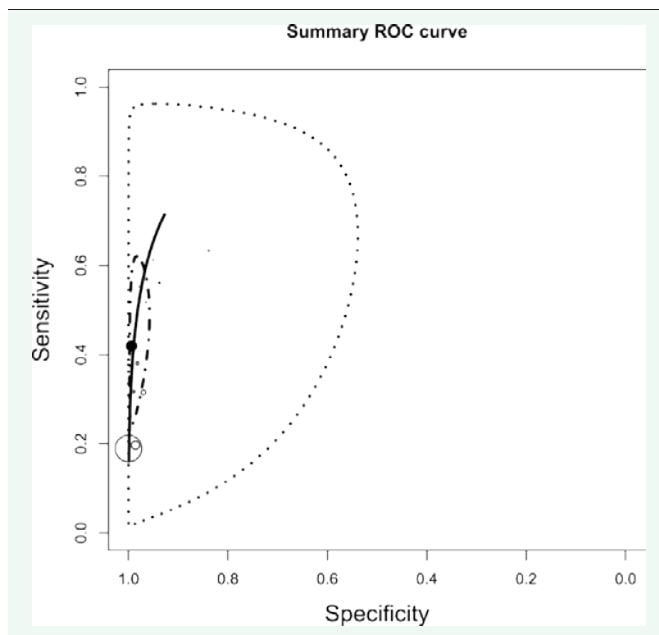
analytical flexible that is appropriate for the measurement of sensitivity and specificity.

In the meta-analysis of diagnostic tests there is great heterogeneity [25,26], due to prevalence, the number of studies, individuals analyzed, laboratory procedures, etc. For this reason, bivariate copulas can be expanded and modeled in conjunction with sensitivity and specificity using multivariate copulas.

**Table 3:** Goodness-of-t tests for Copula Model.

Copula Model	Statistic	p-value
Gauss	0.06547	0.301
Clayton	0.05539	0.659
FGM	0.02287	0.903
Frank	0.06647	0.296

Abbreviations: FGM: Farlie-Gumbel-Morgenstern

**Figure 4** SROC for AuditC data.

The HSROC model does not operate on the original scale of sensitivity and specificity, but on the corresponding logit scale and by generally relying on the bivariate normal distribution for the random effects, it only allows one single correlation structure. So we propose the use of copulas for the meta-analysis for diagnostic accuracy studies, which avoids the previously mentioned problems.

The use of copulas allows the study of dependencies with structures that are not necessarily linear which is possible in diagnostic situations whose results are obtained after dichotomization. Copula models that use beta-binomial distributions to characterize the marginal of true positives and negatives are linked to bivariate or multivariate copulas, i.e., it is a bivariate logistic regression model with random effects, presenting greater flexibility to capture the functional dependence between sensitivity and specificity. In other words, hierarchical models assume a normal bivariate behavior between the logit transformations of sensitivity and specificity. When the bivariate normality assumption is not met, copula modeling is used, which uses beta-binomial distributions to characterize the marginal distributions of the true positive and negative, i.e. it presents a greater flexibility to capture the functional dependence between sensitivity and specificity.

We illustrate the procedure using a dataset published in the literature. The adjustment of the data was obtained by assuming

binary associated tests and taking covariance as a parameter. It is important to note that the five copulas naturally and correctly capture the dependence structure between sensitivity and specificity (inverse relationship). The FGM copula showed a better fit compared to Gauss, C90, C270 and Frank copulas, as it obtained the highest sensitivity value. On the other hand, Frank's copula estimated the highest correlation between sensitivity and specificity and the estimated values for the pair of summary measures (sensitivity and specificity) are very close to the values obtained by [13]. This is not the case with the HSROC method, reflecting the effect that hierarchical modeling has in capturing the relationship between sensitivity and specificity.

Likewise, from the results of the 2x2 table, it can be seen that the best modeling for the study data corresponds to the FGM copula according to the Cramér-von Mises goodness-of-fit test. In the same sense, copula modeling estimated higher sensitivity values than the HSROC model, that is, for the present context, parameter estimates by couples were better than the hierarchical model.

An attractive feature of copulas is that through beta-binomial distributions the measurements of TP and TN are modelled, which is due to the fact that the beta distribution is conjugated with the binomial distribution. An explicit advantage of the beta-binomial distribution is that study sizes are automatically accounted for and the relationship between sensitivity pair and specificity is captured naturally. These characteristics do not happen with the BRMA model, other words, modeling by copulas is a true random effects model, which means that all tests within studies have a single specific parameter (Se; Sp).

For the selection of a dependent bivariate copulation model, the use of hypothesis tests fits highlighted easy to understand. This allows analysts too quickly and identifies a copula model for the construction of a bivariate dependent distribution that fits a data set well.

We have simulated meta-analysis in several scenarios (different numbers of studies, etc) and evaluated the hierarchical model of copulations. The methodology presented allows us to identify the best fit for the copulas analyzed in this work. It suggests that for several studies between 5 and 22, the FGM copula performs the best fit.

To finally conclude, the proposed model with marginal beta-binomial distributions and bivariate copulas seems to be a promising tool for the meta-analysis of diagnostic accuracy studies. What are probably needed in the future are feasible model selection tools that help to assess fits of different models to data sets. These tools should be able to discriminate between the hierarchical models and the copula models but also between different copula models. Until then, it might be a good idea to compute the copula models along with the hierarchical models to check the robustness of the latter.

## ACKNOWLEDGEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper. The first author's research was supported by financial assistance from Escuela Superior Politécnica del Litoral, Ecuador. The second author's



**Table 4:** Goodness-of-t tests for simulated data.

Copulas	Parameter	Number of studies in the Meta-analysis				
		[5-10]	[11-16]	[17-22]	[23-38]	29-35]
Gauss	Statistic	0.118 (0.113-0.123)	0.069(0.066-0.071)	0.050(0.049-0.052)	0.043(0.041-0.044)	0.037(0.036-0.038)
	p-value	0.503(0.464-0.541)	0.490(0.451-0.528)	0.470(0.434-0.506)	0.449(0.405-0.494)	0.490(0.452-0.528)
Clayton	Statistic	0.168(0.104-0.232)	0.169(0.050-0.286)	0.137(0.012-0.262)	0.110(-0.028-0.249)	0.034(0.033-0.035)
	p-value	0.526(0.484-0.569)	0.521(0.480-0.562)	0.527(0.490-0.564)	0.483(0.44-0.526)	0.498(0.461-0.536)
FGM	Statistic	0.064(0.059-0.070)	0.050(0.047-0.053)	0.048(0.045-0.051)	0.061(0.055-0.068)	0.067(0.058-0.075)
	p-value	0.717(0.676-0.758)	0.628(0.586-0.669)	0.487(0.443-0.530)	0.345(0.299-0.391)	0.303(0.262-0.343)
Frank	Statistic	0.111(0.104-0.118)	0.068(0.066-0.071)	0.051(0.049-0.052)	0.043(0.042-0.045)	0.037(0.036-0.038)
	p-value	0.431(0.371-0.491)	0.489(0.441-0.537)	0.465(0.421-0.508)	0.434(0.385-0.482)	0.466(0.425-0.508)

**Abbreviations:** C90: Clayton 90; C270: Clayton 270; FGM: Farlie-Gumbel Morgenstern; Se: Sensitivity; Sp: Specificity

research was supported by financial assistance from Escuela Superior Politécnica del Litoral, Ecuador.

## REFERENCES

1. Egger M, Smith GD, Schneider M. "Systematic Reviews of Observational Studies," in Systematic Reviews in Health Care, London, UK: BMJ Publishing Group. 2001; 211-227.
2. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull.* 1995; 117: 167-178.
3. Fagan. Nomogram for Bayes's theorem. *N Engl J Med.* 1975; 293: 257.
4. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med.* 2001; 20: 2865-2884.
5. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005; 58: 982-990.
6. Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: Methods for the absolute risk difference and relative risk. *Stat Methods Med Res.* 2012; 21: 621-633.
7. RDC Team. R: A language and environment for statistical computing. 2011.
8. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Stat Med.* 1993; 12: 1293-1316.
9. Arends LR, Hamza TH, van Houwelingen JC, Heijnenbrok-Kal MH, Hunink MGM, Stijnen T. Bivariate Random Effects Meta-Analysis of ROC Curves. *Med Decis Mak.* 2008; 28: 621-638.
10. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics.* 2007; 8: 239-251.
11. Nikoloulopoulos AK. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Stat Med.* 2015; 34: 3842-3865.
12. Takwoingi Y, Deeks J. METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies User guide version 1.3. 2010; 37.
13. Kriston L. Meta-analysis: Are 3 Questions Enough to Detect Unhealthy Alcohol Use? *Ann Intern Med.* 2008; 149: 879-888.
14. Ghalibaf MB. Relationship between kendall's tau correlation and mutual information. *Rev Colomb Estad.* 2020; 43: 3-20.
15. Nelsen R. An introduction to copulas. Springer Science & Business Media. 2007.
16. Schweizer B, Wolff EF. On Nonparametric Measures of Dependence for Random Variables. *Ann Stat.* 1981; 9: 879-885.
17. Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Stat Med.* 2014; 33: 17-30.
18. Haigh J, Conover WJ. Practical Nonparametric Statistics. *J R Stat Soc Ser.* 1981; 144: 370.
19. Bachmann LM, Puhan MA, Ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: Literature survey. *Br Med J.* 2006; 332: 1127-1129.
20. Nyaga VN, Arbyn M, Aerts M. CopulaDTA: An R Package for Copula Based Bivariate Beta-Binomial Models for Diagnostic Test Accuracy Studies in a Bayesian Framework. 2016.
21. Cook RD. Detection of Influential Observation in Linear Regression. *Technometrics.* 1977; 19: 15-18.
22. Rasmussen JL. Evaluating Outlier Identification Tests: Mahalanobis D Squared and Comrey Dk. *Multivariate Behav Res.* 1988; 23: 189-202.
23. Genest C, Nešlehová J. A Primer on Copulas for Count Data. *ASTIN Bull.* 2007; 37: 475-515.
24. Mikosch T. Tales and facts-rejoinder. *Extremes.* 2006; 9: 55-62.
25. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection. *Stat Med.* 2009; 28: 2384-2399.
26. Calero JJP, Bauz-Olvera SA, Nieto-Librero AB, Villardón MPG, Hernández-González S. An Alternative to The Cochran-(Q) Statistic For Analysis Of Heterogeneity In Meta-Analysis Of Diagnostic Tests Based On Hj Biplot. *Investig Operacional.* 2018; 39: 536-545.