Revista Colombiana de Estadística - Applied Statistics
Julio 2020, volumen 43, no. 2, pp. 315 a 344
http://dx.doi.org/10.15446/rce.v43n2.80288

Using Copula Functions to Estimate The AUC for Two Dependent Diagnostic Tests

Uso de funciones cópula para estimar el área bajo la curva característica de operación para dos pruebas de diagnóstico dependientes

Luis Carlos Bravo Melo[a], Jennyfer Portilla[b], Jose Rafael Tovar Cuevas[c]

Statistics School, Universidad del Valle, Cali, Colombia

Abstract

When performing validation studies on diagnostic classification procedures, one or more biomarkers are typically measured in individuals. Some of these biomarkers may provide better information; moreover, more than one biomarker may be significant and may exhibit dependence between them. This proposal intends to estimate the Area Under the Receiver Operating Characteristic Curve (AUC) for classifying individuals in a screening study. We analyze the dependence between the results of the tests by means of copula-type dependence (using FGM and Gumbel-Barnett copula functions), and studying the respective AUC under this type of dependence. Three different dependence-level values were evaluated for each copula function considered. In most of the reviewed literature, the authors assume a normal model to represent the performance of the biomarkers used for clinical diagnosis. There are situations in which assuming normality is not possible because that model is not suitable for one or both biomarkers. The proposed statistical model does not depend on some distributional assumption for the biomarkers used for diagnosis procedure, and additionally, it is not necessary to observe a strong or moderate linear dependence between them.

***Key words*:** AUC; Copula function; FGM copula; Gumbel copula; ROC curve; Weak dependence.

[a]Postgraduate student. E-mail: bravo.luis@correounivalle.edu.co
[b]Post graduate student. E-mail: jennyfer.portilla@correounivalle.edu.co
[c]Ph.D. E-mail: jose.r.tovar@correounivalle.edu.co

Resumen

Cuando se realizan estudios de validación en procedimientos de clasificación diagnóstica, normalmente se miden uno o más biomarcadores en los individuos. Algunos biomarcadores pueden proporcionar mejor información que otros y en muchos casos, más de uno puede ser necesario. Cuando se utilizan varios biomarcadores para hacer clasificación, se presenta dependencia entre ellos. En este trabajo se estima el área bajo la curva característica de operación (ABCOR) para establecer la capacidad clasificadora de dos biomarcadores en un procedimiento para diagnóstico clínico. Se estudia mediante copulas (FGM y Gumbel-Barnett) la dependencia entre pruebas y se estima la respectiva área bajo la curva, asumiendo tres niveles para cada estructura de dependencia. En la literatura revisada los autores asumen un modelo normal para representar el comportamiento de los biomarcadores utilizados para el diagnóstico clínico. Hay situaciones en las que no es posible asumir este modelo porque no es adecuado para uno o ambos biomarcadores. El método estadístico propuesto no depende de un supuesto distribucional para los biomarcadores utilizados en el procedimiento de diagnóstico y tampoco es necesario considerar una dependencia lineal fuerte o moderada entre ellos.

***Palabras clave*:** ABCOR; Cópula FGM; Cópula Gumbel Barnett; COR; Dependencia débil.

1. Introduction

The problem of estimating performance parameters and the prevalence in studies for validating diagnostic procedures have been associated with three aspects of interest that are approachable through statistical theory: verification bias, lack of identifiability and the presence of dependence between the test results (Tovar 2011). The last problem has been addressed by different methods such as latent variable models and reparametrizations and many authors have assumed a binary dependence structure using a covariance parameter in the estimation model. Nikoloulopoulos (2018) mentions that the composite likelihood is amongst the computational methods used for estimation of the generalized linear mixed model (GLMM) in the context of bivariate meta-analysis of diagnostic test accuracy studies. To synthesize the diagnostic test accuracy studies, a copula mixed model has been proposed in the biostatistics literature. This general model includes the GLMM as a special case and can also allow for flexible dependence modelling, different from assuming simple linear correlation structures, normality and tail independence in the joint tails. Tovar and Achcar (Tovar & Achcar 2012, Tovar & Achcar 2011*a*, Tovar & Achcar 2013, Tovar & Achcar 2011*b*) addressed the problem of dependence between diagnostic test results by assuming that the dependence structure between the biological traits (biomarkers), measured on an interval or rational scale, can be modeled using a copula function. These authors assumed weak linear dependencies (FGM copula function) and weak, but not necessarily linear dependencies (Gumbel Barnett copula function) between the results of the biomarkers used as diagnostic tests for their approaches. The authors estimated

the performance test parameters and the prevalence, but they did not estimate the area under the receiver operating characteristic (ROC) curve.

To obtain the ROC curve it is necessary to dichotimize the values of the expressions of the biomarkers, establishing a threshold value (cut point), which can be defined using clinical criteria or a statistical methodology. If the cutting point is obtained applying statistical methodology (such as the ROC curve) on the data obtained for the field work, it is necessary to estimate the area under the ROC curve (AUC), in addition to the performance parameters of the test (sensitivity and specificity). The ROC curve is a graph of sensitivity versus 1−specificity for all possible threshold values and it is the most commonly used global index for diagnostic precision. The AUC is also used to choose between two different diagnostic tests. Many authors have studied the statistical properties of the AUC and the methods for estimating them; for example, Faraggi & Reiser (2002) developed and compared some of the processes used for estimating the AUC under parametric and nonparametric assumptions. Zou, O'Malley & Mauri (2007) reviewed and applied the measures of precision used for ROC curves (sensitivity, specificity and AUC) to evaluate diagnostic tests and predictive models.

Relevant works on clinical diagnostic studies agree on the importance of combining the information about the health state contained in different biomarkers used as diagnostic tests because these combinations tend to be more accurate than diagnostic procedures based on single tests (Etzionin, Kooperberg, Pepe, Smith & Gann 2003). Thus a great interest in developing methods to combine multiple tests for disease classification that will result in a deeper and more detailed analysis of the ROC curve (Ma & Huang 2007, Pepe & Thompson 2000, Su & Liu 1993). Generally, the parametric assumptions apply to the distributions of the observed variable in normal and non-normal populations. The maximum likelihood methods to estimate the area under the curve and the relevant parameters under a binormal model assumption have been widely used to estimate this area (DeLong, DeLong & Clarke-Pearson 1988). The normality assumption on the biomarkers or on monotonic transformations of them in both diseased and non diseased populations, in some situations is not true because there exist many biomarkers expressed in a continuous form (Pundir & Amala 2012). Pundir & Amala (2015) consider the use of two continuous biomarkers as clinical diagnostic tests, and they develop a method to estimate AUC under the assumption of correlated tests using a log-normal distribution and the Pearson's correlation coefficient. On the other hand, DeLong et al. (1988) addresses a nonparametric comparison of areas under correlated ROC curves using the theory of generalized U-statistics which takes advantage of the properties of Mann-Whitney statistic to generate an estimated covariance matrix.

The main goal of this work is to estimate the AUC in screening studies to validate procedures for clinical diagnoses, that use two biomarkers expressed in a continuous form. The proposed model assumes a dependence structure between the two biomarkers that can be modeled using copula functions. Given that, both biomarkers are measured in each individual, it is possible that a scatter plot with their data behaves very similar to that observed when the test results are independent between them. We assume that the dependence between diagnostic

tests is linear and weak so it can be modeled using an FGM copula, or the dependence structure is weak but not necessarily linear. We then use a Gumbel Barnett copula structure to model it.

This document is organized as follows: Section 2 presents relevant sections on copula functions and ROC curve and AUC for continuous tests, respectively. We present the analysis of the ROC curve and the AUC, considering copula-type dependencies between diagnostic tests, discusses the steps followed to derive the AUC estimate. Section 3 presents the estimates obtained using the proposed method and the results of a simulation study. In addition, a comparison is shown with the estimates obtained with the Pudir and Amala method (Pundir & Amala 2015). Section 4 we provide a practical example with real data on dengue detection. Finally, Section 5 includes a discussion regarding aspects found during the implementation process. Calculations, simulations, adjustments and ROC curve tracing were performed using the statistical software R.

2. Materials and Methods

2.1. Copula Functions

A copula describes the dependence structure of a multivariate random variable. Using copulas, random variables can be transformed through their cumulative distributions into uniformly distributed variables. The dependence structure is determined by the relationships established between the uniform distributions (Gallardo 2010). The copula functions may use these relationships to link marginal distributions with a joint distribution.

Thus in accordance with Dupuis (2007), a copula is a joint distribution function of random variables with uniform standard distribution as marginals:

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d) \tag{1}$$

where $U_i \sim U(0,1)$, $i = 1, \ldots, d$. Thus copula functions allow the characterization of the dependence structure of a set of random variables independent of the form of the marginal distributions. Random variables with uniform distributions are obtained by applying the probability integral transformation in each of the marginals with distribution $F_1(x), \ldots, F_d(x)$ so that $U_1 = F_1(X), \ldots, U_d = F_d(X)$ (Genest, Quessy & Rémillard 2006). Given a set of random variables $X_1, \ldots, X_d$ with a joint probability distribution H and marginal distribution functions $F_i(x)$, $i = 1, \ldots, d$, one unique copula function C may be written as $C(u_1, \ldots, u_d) = H[F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)]$, where $F_i(u) = inf[x : F_i(X) \geq u]$ defines the quantile function. However, if C is a copula function and $F_1(X), \ldots, F_d(X)$ are arbitrary distribution functions, then function H defined as $H(X_1, \ldots, X_d) = C[F_1(X_1), \ldots, F_d(X_d)]$ is a multivariate distribution function with marginal distribution functions $F_1, \ldots, F_d$.

2.2. ROC Curve and AUC for Continuous Tests

The ROC curve is a graphic in which all sensitivity/specificity pairs resulting from the continuous variation of cutoff points (thresholds) can be found in the full range of the observed results. The proportion of true positives (sensitivity) are located on the y-axis, and the proportion of false positives (1−specificity) are located on the x-axis (Burgueño, García & Gonzáles 1995). Specifically, the use of a threshold (cutoff point) t defines a binary test from one of the continuos biomarkers Z considered in the diagnostic procedure to be evaluated. If $Z \geq t$ then, the individual is classified as positive and if $Z < t$ the individual has a negative result (Pepe 2003). Let X and Y be random variables that represent the values of the biomarkers in the nondiseased and diseased groups. Let $(X_1, X_2, \ldots, X_d)$ be d vectors of values that take related biomarkers measured in individuals among the nondiseased group, and let $(Y_1, Y_2, \ldots, Y_d)$ be the set of vectors associated to the diseased group, where $d = 1, 2, \ldots, k$ is the number of tests. The corresponding rates of the true and false positives for threshold t are $TPR(t)$ and $FPR(t)$, for the diagnostic procedure are given by:

$$TPR(t) = P(Y_1 \geq t_1, Y_2 \geq t_2, \ldots, Y_d \geq t_d | D = 1), \tag{2}$$

$$FPR(t) = P(X_1 \geq t_1, X_2 \geq t_2, \ldots, X_d \geq t_d | D = 0), \tag{3}$$

with D being a dichotomous variable representing the true state of the individual; that is, $D = 1$ for a diseased individual and $D = 0$ for a nondiseased individual, and t corresponds to the threshold vector $t = (t_1, t_2, \ldots, t_d)$, $t_i \in (-\infty, \infty)$. Thus, the ROC curve is the complete set of possible fractions of true and false positives found using the dichotomization of X and Y with different thresholds:

$$ROC(\cdot) = \{(FPR(t), TPR(t))\} \tag{4}$$

A perfect diagnostic test accurately separates diseased subjects from nondiseased subjects. For a given threshold t, we must have $TPR(t) = 1$ and $FPR(t) = 0$ so that the ROC curve is formed over the entire left portion of the positive quadrant.

The AUC is a global measure of accuracy for a diagnostic test and is thus the most commonly used summary index for the ROC curve. The AUC is shown to be the probability of correctly classifying a pair of individuals, selected from the population at random, as healthy or sick, using the results obtained after applying the diagnostic test (Burgueño et al. 1995, Sumi & Hossain 2012).

A perfect test with a perfect ROC curve has a value of $AUC = 1.0$. Likewise, an uninformative test with $ROC(t) = t$ has an $AUC = 0.5$. The majority of tests have values that fall between these two values.

2.3. Analysis of a ROC Curve with Copula Dependence and $d = 2$ Continuos Biomarkers

Sometimes two biomarkers can be associated with the presence of a disease; therefore these biomarkers must be considered in conjunction to classify the subject

(Ma & Huang 2007). Let (X, Y) represent the values of the biomarkers in the nondiseased and diseased groups. Let (X_1, X_2) be two sets of related biomarkers measured in the nondiseased group, and let (Y_1, Y_2) be two related biomarkers taken from the diseased group. Thus (X_1, X_2) and (Y_1, Y_2) are independent pairs of bivariate biomarkers in each group of individuals; and a subject is identified as diseased when the values of Y_1 and Y_2 are sufficiently large (greater than a given threshold or cutoff point) (Wang & Li 2012). The cumulative distribution functions for the random variables that define the biomarker results are defined as $F_Y(t_1, t_2) = P(Y_1 \leq t_1, Y_2 \leq t_2)$ and $F_X(t_1, t_2) = P(X_1 \leq t_1, X_2 \leq t_2)$, respectively, where t_1 and t_2 correspond to the cutoff points for each test. The method develops an iterative procedure taking all the possible permutations between both biomarkres within each group and for each permutation it evaluates the individual's health condition and classifies it as positive or negative. For each pairs t_1, t_2 its possible to obtain a 2×2 table, with the results showed in Table 1.

Table 1: Final classification obtained after to apply the diagnostic procedure. .

	Positive	Negative
True	$Y_1 > t_1 \wedge Y_2 > t_2$	$X_1 < t_1 \wedge X_2 < t_2$
False	$X_1 > t_1 \wedge X_2 > t_2$	$Y_1 < t_1 \wedge Y_2 < t_2$
	$X_1 > t_1 \wedge X_2 < t_2$	$Y_1 < t_1 \wedge Y_2 > t_2$
	$X_1 < t_1 \wedge X_2 > t_2$	$Y_1 > t_1 \wedge Y_2 < t_2$

For the construction of the ROC curve, the false positive rate (FPR) and true positive rate (TPR) can be defined as $P(X_1 > t_1, X_2 > t_2)$ and $P(Y_1 > t_1, Y_2 > t_2)$, respectively, according to this bivariate criterion. We assume that dependence between results of biomarkers can be modeled using copula functions; and that it is possible to estimate the AUC including that fact in the estimation model. Our methodological approach assume two copula functions as candidates for modeling the dependence structure between the biomarkers; the Farlie-Gumbel-Morgenstern (FGM) and the Gumbel-Barnett.

The FGM copula function has the following analytical form:

$$C(u_1, u_2) = u_1 u_2[1 + \varphi(1 - u_1)(1 - u_2)], \quad \varphi \in [-1, 1] \tag{5}$$

where φ is the dependence parameter with $\rho = -\frac{1}{3} \Leftrightarrow \varphi = -1$ and $\rho = \frac{1}{3} \Leftrightarrow \varphi = 1$; ρ is the Pearson correlation coefficient (Nelsen 2006).

The Gumbel Barnett copula function has the form:

$$C(u_1, u_2) = \\ u_1 + u_2 - 1 + (1 - u_1)(1 - u_2) \times \exp\{-\phi \log(1 - u_1) \log(1 - u_2)\}, \quad \phi \in [0, 1] \tag{6}$$

where ϕ is the dependence parameter and $\rho = 0 \Leftrightarrow \phi = 0$, $\rho = -0.41 \Leftrightarrow \phi = 1$, ρ is the Pearson correlation coefficient (Portilla & Tovar 2018).

For the random variables associated with results of the biomarkers, we have that $X_i \sim G_{X_i}(x_i)$ and $Y_i \sim G_{Y_i}(y_i)$ $\forall i = 1, 2$. Once the distribution $G(\cdot)$ has been determined (Goodness-of-fit tests can be performed to determine the

corresponding distribution), we proceed to estimate their respective parameters jointly (see Appendix A). The explicit forms of the TPR and FPR in the bivariate case are analytically complex. The corresponding rates (ratios) of true and false positives for threshold $t = (t_1, t_2)$ are $TPR(t)$ and $FPR(t)$, respectively, with

$$\begin{aligned} TPR(t) &= P(Y_1 > t_1, Y_2 > t_2) \\ &= \int_{t_1}^{1} \int_{t_2}^{1} C_Y(G_{Y_1}(y_1), G_{Y_2}(y_2))\ dG_{Y_1}(y_1)\ dG_{Y_2}(y_2) \end{aligned}$$

and

$$\begin{aligned} FPR(t) &= P(X_1 > t_1, X_2 > t_2) \\ &= \int_{t_1}^{1} \int_{t_2}^{1} C_X(G_{X_1}(x_1), G_{X_2}(x_2))\ dG_{X_1}(x_1)\ dG_{X_2}(x_2) \end{aligned}$$

After applying the probability integral transformation (PIT), the following equations must be true:

$$TPR(t) = \int_{t_1^*}^{1} \int_{t_2^*}^{1} C_Y(v_1, v_2)\ dV_1\ dV_2 \tag{7}$$

and

$$FPR(t) = \int_{t_1^*}^{1} \int_{t_2^*}^{1} C_X(u_1, u_2)\ dU_1\ dU_2 \tag{8}$$

where t_1^* and t_2^* correspond to the cutoff points for each test after we applied the respective PIT.

In accordance with Pundir & Amala (2015), the AUC of a bivariate dependent ROC curve has the form:

$$\begin{aligned} AUC = P(Y_1 > X_1, Y_2 > X_2) = \\ \int\int I(y_1 > x_1, y_2 > x_2)\ dG_X(x_1, x_2)\ dG_Y(y_1, y_2), \end{aligned} \tag{9}$$

where $G_X = P(X_1 \leq x_1, X_2 \leq x_2)$ and $G_Y = P(Y_1 \leq y_1, Y_2 \leq y_2)$.

Given that the cumulative functions G_X and G_Y can be written in terms of copula functions, the AUC for the bivariate dependent copula ROC curve assumes the following form:

$$\begin{aligned} AUC = P(Y_1 > X_1, Y_2 > X_2) = \\ \int\int I(y_1 > x_1, y_2 > x_2)\ dC_X(x_1, x_2)\ dC_Y(x_1, x_2), \end{aligned} \tag{10}$$

where the univariate case, the expression (10) is proportional to the statistics of the traditional non-parametric Mann-Whitney test (Bamber 1975)

Then, considering the uniform variables obtained after to apply the PIT using the marginal distributions (or the empirical cumulated distributions), we have:

$$AUC = \int\int I(v_1 > u_1, v_2 > u_2)\ dC_X(u_1, u_2)\ dC_Y(v_1, v_2) \tag{11}$$

which cannot be expressed in a closed form, its can be estimated using numerical methods, such as the trapezoidal rule or Simpson's rule (Pundir & Amala 2015).

If we assume that the biomarker results have a dependence structure that can be modeled using an FGM copula, we have:

$$\begin{aligned} AUC = P(Y_1 > X_1, Y_2 > X_2) = \int\int \Big\{ & I(v_1 > u_1, v_2 > u_2) \\ & d[u_1u_2[1+\varphi_1(1-u_1)(1-u_2)]]\ d[v_1v_2[1+\varphi_2(1-v_1)(1-v_2)]] \Big\} \end{aligned}$$

Similarly, if we used an estimation model that assumes a Gumbel-Barnett structure for the dependence between the test results, we have:

$$\begin{aligned} AUC =& P(Y_1 > X_1, Y_2 > X_2) \\ =& \int\int \Big\{ I(v_1 > u_1, v_2 > u_2) \\ & d[u_1 + u_2 - 1 + (1-u_1)(1-u_2)\exp\{-\phi_1 \log(1-u_1)\log(1-u_2)\}] \\ & d[v_1 + v_2 - 1 + (1-v_1)(1-v_2)\exp\{-\phi_2 \log(1-v_1)\log(1-v_2)\}] \Big\} \end{aligned}$$

It is possible to use copula functions to model the structure dependence between two random variables in a statistical procedure developed to estimate the AUC curve, when the marginal distributions are or are not known. If the marginal distributions are not known, it is possible to apply the probability integral transformation on observed data using the respective empirical cumulated distribution (Achcar, Tovar & Moala 2019). The ROC curve and its area can be estimated using the transformed data preserving the dependence structure of the biomarkers. The importance of determining the data distribution, is to be able to estimate θ (copula dependence parameter), where θ is φ or ϕ; that is, make use of the data marginal distributions to use all the information when estimating the dependence parameter, using the expression (A3) in Apendix A (Bouyé, Durrleman, Nikeghbali, Riboulet & Roncalli 2000).

We developed the procedure when the validation study includes two continuos biomarkers as validation tests, but, it is possible to generalize our proposed approximation for cases with d biomarkers, and the analytical form of the AUC is as follows:

$$\begin{aligned} AUC =& \\ & \int\int I(v_1 > u_1, v_2 > u_2, \ldots, v_d > u_d)\ dC(u_1, u_2, \ldots, u_d)\ dC(v_1, v_2, \cdots, v_d) \end{aligned}$$

2.4. Simulation Study

We simulated a validation study that includes the use of two biomarkers with continuous expression as indicators of the disease status and a confirmatory test called gold standard, which classifies the individuals without error. Given that we needed to compare our results with those obtained using methods reported in the literature, we generated pairs of observations of variables distributed with a bivariate normal distribution, using the R package.

To have reference values, we used the clinical values reported for triglycerides and LDL cholesterol by the MedlinePlus web page as motivation (National Institutes of Health and others 2004). Then, within the nondisease individuals, data from Test 1 (triglycerides) were simulated using $\mu_{X_1} = 164$ $\sigma_{X_1} = 24.6$, and data from Test 2 (LDL cholesterol) $\mu_{X_2} = 160$ $\sigma_{X_2} = 24$ were used. Within the diseased individuals, data from Test 1 were simulated with $\mu_{Y_1} = 224$ $\sigma_{Y_1} = 33.6$; while for Test 2, the values used were $\mu_{Y_2} = 209$ $\sigma_{Y_2} = 31.35$. The variances were obtained establishing 15% as the coefficient of variation for the normal data. We simulated a validation study carried out using a sample of 10000 individuals in a population with a 10% prevalence (1000 diseased individuals and 9000 nondiseased individuals).

We simulated pairs of data $u_{jk}, v_{jk},\ j = 1, 2;\ k = 1, 2, \ldots, n_j$ with copula dependence (FGM or Gumbel-Barnett). Three different dependence-level values were evaluated ($\theta = 0.2$, $\theta = 0.5$ and $\theta = 0.9$) for each copula function considered. Next, we transformed each pair considering that $\Phi^{-1}(u_{jk}) = z_{jk}$ then $x_{jk} = z_{jk}\sigma_{X_j} + \mu_{X_j}$ where Φ is the cumulative probability normal function. In that way, we simulated data with copula dependence structure and normal marginal distributions.

The data with an FGM copula dependence structure were generated using the algorithm proposed by Johnson (1987) as follows:

1. To generate independent v_1 and v_2 values from a Uniform$(0, 1)$ distribution.
2. Run $u_1 = v_1$.
3. For a given value of θ, compute $A = \theta(2u_1 - 1) - 1$ and $B = \left(1 - 2\theta\,(2u_1 - 1) + \theta^2(2u_1 - 1)^2 + 4\theta v_2\,(2u_1 - 1)\right)^{\frac{1}{2}}$
4. Run $u_2 = \dfrac{2v_2}{B - A}$

The data set of pairs of data with a Gumbel-Barnett dependence structure were simulated using an algorithm based on the inversion of data from a Gumbel Type I distribution (see Gumbel (1960) for details) as follows:

Given V_1 and V_2 with $V_i \sim Uniform(0, 1)$ run

$$w_1 = -\log(1 - v_2)$$

Then, solve the nonlinear equation

$$1 - (1 + \theta w_2)e^{-(1+\theta w_1)w_2} - v_2 = 0$$

In that way, pairs (w_1, w_2) from a Gumbel Type I distribution are obtained (Gumbel 1960). Applying the transformation

$$v_1 = 1 - exp(-w_2) \quad v_2 = 1 - exp(-w_1)$$

it is possible to obtain a vector of pairs of observations (V_1, V_2) with uniform marginal distributions and dependence that can be modeled using a Gumbel-Barnett copula function.

With the simulated data, and with the aim of showing that the simulated data shows a weak or not necessarily linear dependence, we computed the estimate value of the coefficient ρ using the formulas that appears in Appendix B, depended of the chosen copula function. We also estimate Pearson's correlation coefficient directly from the simulated data, where $\hat{\rho}_x$ is the estimate in the group of nondiseased individuals and $\hat{\rho}_y$ is the estimate obtained in the other group. In accordance with the results shown in Table 2, it was possible to conclude that the simulated data complies with the desired characteristics of dependence. The scatter plot of the simulated data for each dependence level in diseased and nondiseased individual groups, are shown in Figures 1 and 2.

Table 2: Pearson's correlation index and dependence parameter values.

	φ	ρ	$\hat{\rho}_x$	$\hat{\rho}_y$
FGM	0.2	0.069756	0.062408	0.061626
	0.5	0.173648	0.175355	0.151183
	0.9	0.309017	0.281865	0.264024
	ϕ	ρ	$\hat{\rho}_x$	$\hat{\rho}_y$
Gumbel	0.2	−0.147889	−0.160861	−0.125373
	0.5	−0.277343	−0.295476	−0.254096
	0.9	−0.383462	−0.470656	−0.476716

Given that the AUC does not have a closed form, this indicator cannot be obtained analytically; however, the form may be approximated using simulation techniques such as the bootstrap or Monte Carlo (MC) processes.

The proposed algorithm was run 1000 times as follows:

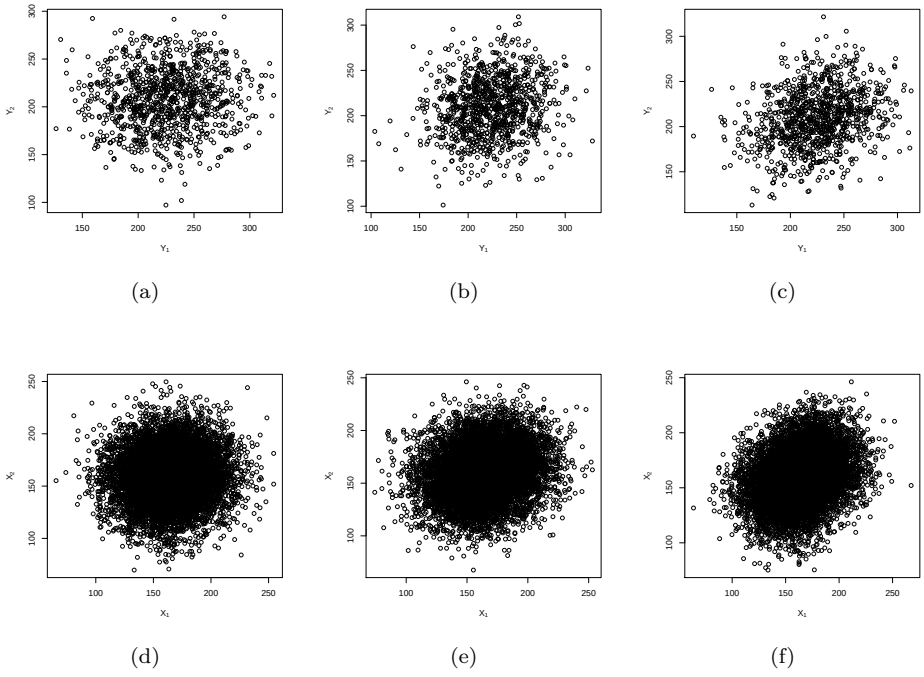
1. Determine the parametric model (i.e., find the distribution functions for each variable that best fits the data). Goodness-of-fit tests can be performed to determine the corresponding distribution. Thanks to Sklar's theorem (Nelsen 2006) and the integral probability transformation, the method is developed with u_i values and copula functions. Once the model is chosen, estimate the parameters related to this model, using maximum likelihood or moments method.

2. To estimate the dependence parameter, use equation (A3) in Apendix A (Bouyé et al. 2000).

3. Using the estimates of parameters in Step 1 as parametric values, generate Markov chain Monte Carlo (MCMC) samples from the copula function (with the respective marginal values) for size m for the nondiseased group and n for the diseased group. Using the generated values, calculate the $\hat{AUC}$ using equation (11).

4. For each sample generated in Step 3, maintain the respective samples for each group (i.e., m and n), perform bootstrapping, and calculate the FPR and TPR. For each stage of resampling, calculate the AUC that will serve as the input for calculating the intervals corresponding to the 2.5 and 97.5 percentiles.

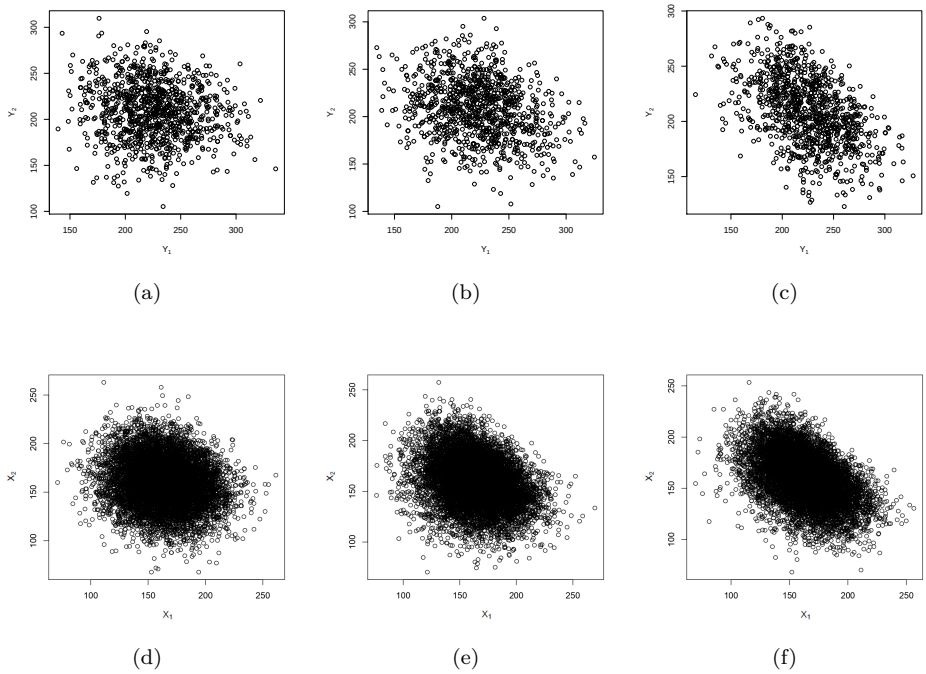
5. Repeat Step 4 many times (at the last 1000 times).



(a) (b) (c)

(d) (e) (f)

FIGURE 1: Scatter plots of health status, assuming an FGM dependence structure. 1(a) Scatter plot for diseased individuals with $\varphi = 0.2$; $\hat{\rho}_y = 0.06$. 1(b) Scatter plot for diseased individuals with $\varphi = 0.5$; $\hat{\rho}_y = 0.15$. 1(c) Scatter plot for diseased individuals with $\varphi = 0.9$; $\hat{\rho}_y = 0.26$. 1(d) Scatter plot for nondiseased individuals with $\varphi = 0.2$; $\hat{\rho}_x = 0.06$. 1(e) Scatter plot for nondiseased individuals with $\varphi = 0.5$; $\hat{\rho}_x = 0.17$. 1(f) Scatter plot for nondiseased individuals with $\varphi = 0.9$; $\hat{\rho}_x = 0.28$.



(a) (b) (c)

(d) (e) (f)

FIGURE 2: Scatter plots of health status, assuming a Gumbel Barnett dependence structure. 2(a) Scatter plot for diseased individuals with $\phi = 0.2$; $\hat{\rho}_y = -0.12$. 2(b) Scatter plot for diseased individuals with $\phi = 0.5$; $\hat{\rho}_y = -0.25$. 2(c) Scatter plot for diseased individuals with $\phi = 0.9$; $\hat{\rho}_y = -0.47$. 2(d) Scatter plot for nondiseased individuals with $\phi = 0.2$; $\hat{\rho}_x = -0.16$. 2(e) Scatter plot for nondiseased individuals with $\phi = 0.5$; $\hat{\rho}_x = -0.29$. 2(f) Scatter plot for nondiseased individuals with $\phi = 0.9$; $\hat{\rho}_x = -0.47$.

3. Results

3.1. ROC Curves with Simulated date Assuming FGM Dependence

According to the method proposed in section 2.4, a prevalence of 10% was assumed for this case, which, considering a population of 10000 individuals, determines 9000 non-sick and 1000 sick; this is $(m, n) = (9000, 1000)$. Given the context of this work (clinical diagnostic tests), only the positive part of the FGM copula function should be used (Dendukuri & Joseph 2001, Georgiadis, Johnson, Gardner & Singh 2003). AUC estimation error, bootstrap confidence intervals for AUC were estimated, in addition to the specificity, sensitivity (of the joint test) and cutoff points for each test (with their respective standard error). The performance parameters and cutoff points estimates were obtained using the

Youden index (Specificity+Sensitivity−1) (Youden 1950), taking the respective values that maximized that index.

According to Table 3, the effect of the dependence level on the AUC estimates is not very important although it is possible to observe a trend to decrease when dependence increases.

Table 3: AUC, performance test parameters (specificity and sensitivity) and cutoff point estimates, estimation errors and average length of the AUC confidence interval for the FGM data. $E(\cdot)$ corresponds to the respective expected value, $Se(\cdot)$ to the standard error, $t1$ to the cutoff point of test 1 and $t2$ to the cutoff point of test 2.

	φ		AUC	Interval Length	Specificity	Sensitivity	$t1$	$t2$
FGM	0.2	$E(\cdot)$	0.944770	0.000917	0.907288	0.848352	177.7991	173.0683
		$Se(\cdot)$	0.000522	-	0.002896	0.002929	0.428356	0.377072
	0.5	$E(\cdot)$	0.941367	0.000896	0.903108	0.841207	178.7563	173.9788
		$Se(\cdot)$	0.000509	-	0.003213	0.003197	0.455436	0.396331
	0.9	$E(\cdot)$	0.937002	0.000863	0.899757	0.830758	180.1202	175.2873
		$Se(\cdot)$	0.000455	-	0.003204	0.003347	0.421199	0.414718

The observed values for the AUC estimation errors are small, which is expected because the sample sizes are very large. The AUC values are high and similar to one another, considering the variation in the level of dependence. The interval lengths are very small considering that if the AUC value is between 0 and 1, the maximum length of the interval is 1. In the case of specificity and sensitivity, good values (or at least within those typically expected) are seen for each performance parameter, with lower errors in the estimation. This scenario is very similar to the situation with the AUC estimations. The estimates of the cutoff points show little variability and show the same trend to increase when the dependence increases.

A prevalence of 60% is assumed (unlike the initial proposal of 10%), considering a population of 10000 individuals, 4000 non-sick and 6000 sick are determined; this is $(m, n) = (4000, 6000)$. This to verify that the prevalence does not actually affect the AUC estimate, since the latter depends on the performance parameters (sensitivity and specificity) of the test (Table 4). It's important to consider this situation since it is a similar measure of prevalence to that presented later in the case study (Dengue data).

Table 4: AUC, performance test parameters and cutoff point estimates, estimation errors and average length of the AUC confidence interval for the FGM data.

	φ		AUC	Interval Length	Specificity	Sensitivity	$t1$	$t2$
FGM	0.2	$E(\cdot)$	0.946286	0.001196	0.909553	0.851116	179.3697	171.9851
		$Se(\cdot)$	0.000702	-	0.003025	0.003143	0.478657	0.443158
	0.5	$E(\cdot)$	0.942970	0.001148	0.905660	0.843918	180.4051	172.8766
		$Se(\cdot)$	0.000688	-	0.003167	0.003404	0.494868	0.458196
	0.9	$E(\cdot)$	0.938741	0.001108	0.902748	0.833478	181.9064	174.1380
		$Se(\cdot)$	0.000633	-	0.003555	0.003730	0.516503	0.461524

The previous section, the FGM copula dependence simulation was performed for different sample sizes ($N = 100$, $N = 500$ and $N = 1000$), considering a dependence level $\varphi = 0.2$, and 10% prevalence (see Table 5).

Table 5: AUC, performance test parameters and cutoff point estimates, estimation errors and average length of the AUC confidence interval for the FGM data, considering different population sizes.

FGM function copula $\varphi = 0.2$							
		AUC	Interval Length	Sensitivity	Specificity	$t1$	$t2$
$N = 100$	$E(\cdot)$	0,944458	0,011796	0,851024	0,905970	178,4551	172,1064
	$Se(\cdot)$	0,004677	-	0,015408	0,015568	3,229691	2,458372
$N = 500$	$E(\cdot)$	0,945018	0,005389	0,849577	0,907445	177,9386	172,7651
	$Se(\cdot)$	0,002131	-	0,009165	0,007606	1,460596	1,398408
$N = 1000$	$E(\cdot)$	0,945062	0,003834	0,848222	0,908589	178,0073	173,0137
	$Se(\cdot)$	0,001825	-	0,006846	0,006508	1,158346	0,964591

Finally, Table 6 shows the average and standard error of the AUC estimates, interval length, performance parameters and cut-off point of simulated data with $\varphi = 0$ in the FGM copula function. The above with a population of 1000 individuals, and 10% prevalence.

Table 6: AUC estimates with simulated FGM data, $\varphi = 0$.

	AUC	Intervalo	Espec	Sensi	$t1$	$t2$
$E(\cdot)$	0,94706	0,000904	0,91083	0,8527	177,21	172,50
$Se(\cdot)$	0,00058	-	0,00271	0,0028	0,3884	0,3847

3.2. ROC Curves With Simulated Date Assuming Gumbel Barnett Dependence

Upon adjusting the model with the Gumbel-Barnett copula function, the results are similar to those observed with the other copula function (FGM) in that the AUC estimates are high with low errors, and are nearly constant. A slight increase in the estimates occurred as the level of dependence increased. For the average interval length, we noted that these intervals became increasingly narrow (their amplitude decreased) as the level of dependence increased. The estimates of the performance test parameters and the AUC showed similar behavior. In the same way as for other copula dependence, the estimates of the cutoff point estimates increased slightly when the dependence level increased (Table 7).

Table 7: AUC, performance test parameters and cutoff point estimates, estimation errors and average length of the AUC confidence interval for the Gumbel-Barnett data.

	ϕ		AUC	Interval Length	Specificity	Sensitivity	$t1$	$t2$
Gumbel	0.2	$E(\cdot)$	0.954396	0.013459	0.931458	0.859772	178.4756	173.8466
		$Se(\cdot)$	0.004899	-	0.005068	0.009258	0.481494	0.355141
	0.5	$E(\cdot)$	0.959536	0.007697	0.937323	0.857589	180.0122	175.2528
		$Se(\cdot)$	0.004985	-	0.016997	0.022203	0.854568	0.872031
	0.9	$E(\cdot)$	0.964563	0.006747	0.952738	0.870531	182.2365	178.0468
		$Se(\cdot)$	0.004756	-	0.010828	0.025205	0.789746	0.929957

3.3. Comparison of Estimation Methods Using the Simulated Data

We estimated the parameters using our methodology and the method proposed by Pundir & Amala (2015). It is important to point out that although both methods estimate the AUC in presence of a dependence structure in the data, Pundir and Amala work under the assumption that both biomarkers can be modeled with random variables under a bivariate normal distribution of probabilities, then the dependence between biomarkers is assumed to be linear and can be expressed in the estimation model using Pearson's correlation coefficient. The proposed method does not consider the marginal distributions, and the dependence between biomarkers is not necessarily linear. Thus we needed to compare the methods, we simulated sets of $N_2(\mu, \sigma, \rho)$ data and using the procedure in Section 2.4; and we fitted Pundir and Amala's model and ours. For both cases the Bootstrap confidence intervals were obtained; while the performance parameters and cutoff point estimates were obtained using the Youden index (Youden 1950).

According to the results in Table 8, Pundir and Amala's estimation method (AUC2) presents lower estimates than those obtained using the proposed method (AUC1). The AUC2 estimates in nearly all scenarios are outside the confidence intervals estimated for AUC1, except for of AUC2, considering a Gumbel dependence of $\theta = 0.5$, where the estimate is within the given interval.

Table 8: AUC estimates using the proposed methodology (AUC1) and Pundir and Amala's estimation method (AUC2).

		Proposed Method		Pundir & Amala Method	
	φ	AUC1	Interval1	AUC2	Interval2
FGM	0.2	0.96531	[0.96498, 0.96581]	0.92218	[0.91875, 0.92916]
	0.5	0.96559	[0.96511, 0.96591]	0.92288	[0.91208, 0.92293]
	0.9	0.96556	[0.96536, 0.96612]	0.92415	[0.91490, 0.93491]
	ϕ	AUC1	Interval1	AUC2	Interval2
Gumbel	0.2	0.95565	[0.94565, 0.96565]	0.94024	[0.93383, 0.94056]
	0.5	0.95543	[0.91543, 0.96343]	0.93984	[0.93357, 0.94540]
	0.9	0.96357	[0.95357, 0.98713]	0.94988	[0.94746, 0.95173]

The model considering the Gumbel-Barnett copula function is capable of perceiving nonlinear dependencies and/or negative dependencies, which were considered in the construction of the data used in the estimates (data simulated with Gumbel-Barnett copula-type dependence). This phenomenon may explain the difference between the AUC estimates, given that a linear dependence is assumed for AUC2.

The model considering the FGM copula function notes weak and/or low dependencies. Given that the AUC2 estimates use a linear correlation coefficient to measure the dependence present in the data and the dependencies with which the data are constructed (data simulated with an FGM copula-type dependence) based on this coefficient are very low, these dependencies cannot be clearly perceived. Thus the results are slightly lower than those found for AUC1. On the other hand, the confidence intervals obtained assuming GB dependence are narrower than those observed when we assumed the other dependence structure.

Table 9: Estimates of the performance parameters and cutoff points using the proposed method and that of Pundir and Amala.

			Proposed Method		
	φ	Specificity_1	Sensitivity_1	$t1_1$	$t2_1$
FGM	0.2	0.91494	0.84572	177.4492	172.9329
	0.5	0.91108	0.85137	176.7055	172.5069
	0.9	0.91490	0.85532	177.1139	172.2572
	ϕ	Specificity_1	Sensitivity_1	$t1_1$	$t2_1$
Gumbel	0.2	0.92440	0.86437	177.5931	173.1518
	0.5	0.93701	0.85673	178.6307	175.0469
	0.9	0.95452	0.87538	180.2976	176.2156
			Pundir & Amala's Method		
	φ	Specificity_2	Sensitivity_2	$t1_2$	$t2_2$
FGM	0.2	0.89145	0.80960	158.6583	154.5250
	0.5	0.88617	0.81443	157.4523	153.6407
	0.9	0.88241	0.81933	156.6926	153.0160
	ϕ	Specificity_2	Sensitivity_2	$t1_2$	$t2_2$
Gumbel	0.2	0.90703	0.83917	160.2792	156.0931
	0.5	0.91959	0.83687	160.8038	156.6696
	0.9	0.93279	0.84973	161.8535	157.8930

The observed values of the estimates of the performance test parameters obtained using the proposed method (subindex 1 in Table 9) were higher than those observed using Pundir and Amala's method (subindex 2 in Table 9). If a GB dependence structure is considered, the estimates of the specificity and cutoff points increase for high values of the dependence level when the estimates are obtained using the proposed method. The estimates of the performance test parameter assuming an FGM dependence structure do not show differences when the dependence level changes (see Table 9).

We obtained the AUC estimated for each biomarker assuming independence between tests and considering the dependence structure. In both cases, the estimated AUC was lower than the estimate of the joint AUC (Table 10). The

AUC estimates and their respective 95% confidence intervals were estimated using the R pROC package.

Table 10: Estimation of the AUC for each simulated biomarker.

	φ	AUC test 1	95% Confidence interval 1	AUC test 2	95% Confidence interval 2
FGM	0.2	0.9210	[0.9111, 0.9310]	0.8864	[0.8744, 0.8984]
	0.5	0.9285	[0.9193, 0.9378]	0.8907	[0.8790, 0.9025]
	0.9	0.9325	[0.9232, 0.9418]	0.8939	[0.8820, 0.9057]
	ϕ	AUC test 1	95% Confidence interval 1	AUC test 2	95% Confidence interval 2
Gumbel	0.2	0.9350	[0.9266, 0.9434]	0.8924	[0.8808, 0.9039]
	0.5	0.9236	[0.9143, 0.9330]	0.8899	[0.8781, 0.9018]
	0.9	0.9308	[0.9212, 0.9404]	0.8929	[0.8814, 0.9044]

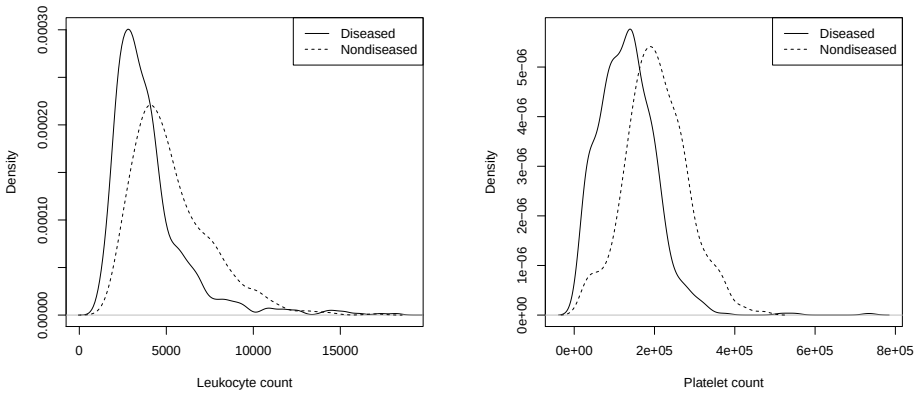
4. Dengue Data

The proposed method was applied to diagnostic tests for detecting dengue; an acute viral disease transmitted by mosquitoes, characterized by high fevers, headaches, pain in muscles and joints, and skin rash. The data set was obtained from the Colombian network for studying dengue (AEDES). In this study, 1380 individuals with symptoms suggestive of dengue were clinically evaluated by a specialist and the results of a hemogram. For each patient an algorithm was run starting with polymerase chain reaction (RT-PCR) test results; the NS1 antigen and antibodies against dengue (IgG and IgM) were applied as the gold standard tests. The tests consist of a leukocyte count (white blood cells: Test 1) and a platelet count (Test 2) in randomly selected individuals. Of these individuals, 744 were diagnosed as having dengue, and 636 individuals were diagnosed as not sick (discarded for dengue but with symptoms of some other condition).

The density plot of both variables had an asymmetric shape in both the diseased and nondiseased groups of individuals, which can be an indicator of the lack of fit to normal distribution (see Figure 3). The scatter plot and the estimates of the correlation or concordance indexes commonly computed for the biomarkers in both groups of patients, showed the presence of a weak dependence structure (see Figures 4 and 5).

The ROC curves of each test, as well as their respective area under the curve (AUC), are presented in Figure 6. The cutoff points found were $t1 = 3590$ for test 1 and $t2 = 158250$ for test 2. The sensitivity and specificity were 0.5362903 and 0.7893082, respectively for test 1. Sensitivity and specificity were 0.6854839 and 0.7374214, respectively for test 2.

Even when we observed an asymmetric form in the distribution of the biomarkers used to diagnose dengue, we decided to assume normal distributions to compare the results between the procedures. We fitted our procedure assuming a gamma(α, β) distribution for each biomarker using the parametrization $1/\beta = \lambda$. The estimates of the FGM dependence parameter and the parameters for the marginal distributions are found in Table 11.



(a) (b)

FIGURE 3: Density plots of the dengue biomarkers in sick and nonsick individuals. 3(a) Density plot for leukocyte count in sick and nonsick groups. 3(b) Density plot for platelet count in sick and nonsick groups.

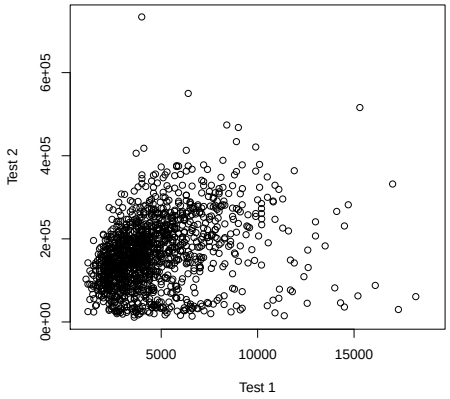
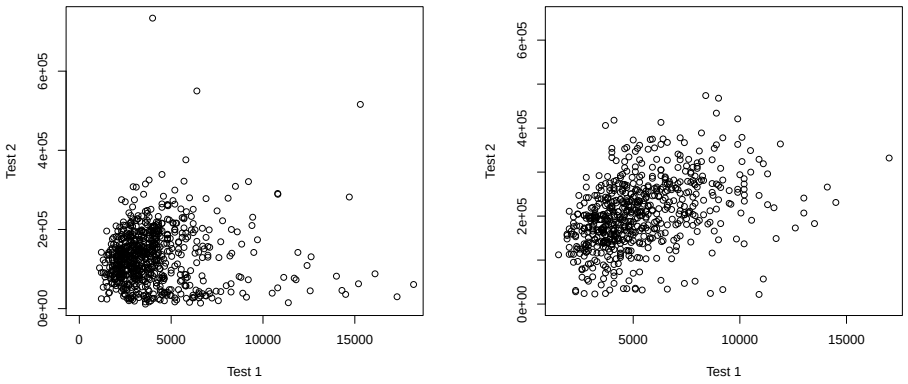


FIGURE 4: Scatter plot of the two biomarkers ($\tau = 0.26367$; $\rho = 0.30512$; $\rho_s = 0.37137$).

TABLE 11: Estimates of the performance test and dependence parameters.

Marginal distribution	Real health status	$\hat{\varphi}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
Normal	Diseased	0.2159	4036.53	133154.90	2720.93	71841.68
	Nondiseased	0.0678	5306.87	204818.60	2274.82	77593.94
	Real health status	$\hat{\varphi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Gamma	Diseased	0.8708	5.1121	5.1682	0.00125	0.000038
	Nondiseased	0.9633	6.1762	5.7512	0.00117	0.000028



(a) (b)

FIGURE 5: Scatter plot of the biomarkers according to health status. 5(a) Scatter plot for the biomarkers in the diseased group ($\tau = 0.07295$; $\rho = 0.05222$; $\rho_s = 0.09113$). 5(b) Scatter plot for the biomarkers in the nondiseased group ($\tau = 0.30525$; $\rho = 0.38312$; $\rho_s = 0.44299$).

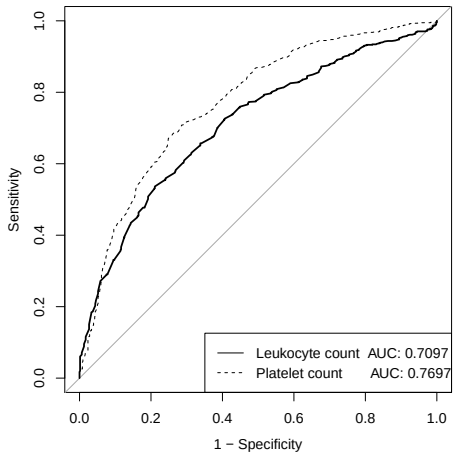


FIGURE 6: ROC curves and AUC of diagnostic tests for dengue.

To evaluate the fit of the dataset to the copula functions, the multiplier method of the goodness of fit (GOF) test was used as introduced by Kojadinovic, Yan & Holmes (2011). This method consists of comparing and validating the distance between the empiric copula function and the copula function under consideration. The FGM copula function shows the better fit.

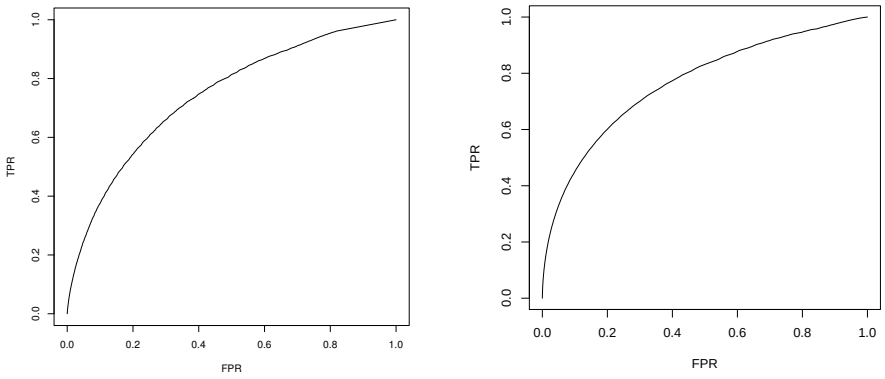
Table 12 shows the AUC estimations, performance parameters and cutoff points obtained using the methods explained in the previous section. The data used for

these estimates were the original data (no transformations) and the data after logarithmic transformation. The latter were applied with the goal of making data available on the same scale and trying to derive the greatest differences between the averages for the sick and not sick populations to obtain the highest performance of the ROC curve (Pundir & Amala 2015). We obtained the estimates of the parameters, assuming Normal(μ, σ) and Gamma(α, β) distributions for marginal distribution and an FGM dependence structure.

According to the results shown in Table 12, the results obtained using the Pundir and Amala's method are quite far from what was expected and if this analysis were done, the same would lead to the conclusion that the two biomarkers together, could only identify nondiseased individuals. The proposed method results in a better approximation for the set of parameters to be estimated when we used the transformed data, assuming normal distributions for the marginals (see Table 12). Figure 7 shows the AUC curves based on the nontransformed data and both marginal distributions after we used the proposed method of estimation.

Table 12: AUC, performance parameters and cutoff point estimates for the dengue data, assuming an FGM structure of dependence.

Marginal distribution	Data set	AUC	Specificity	Sensitivity	t1	t2
FGM-Normal	Originals	0.77496	0.75109	0.66831	4800.00	195000.0
	Logarithm	0.82647	0.74702	0.75104	5340.00	209000.0
FGM-Gamma	Originals	0.72981	0.69544	0.63872	4600.00	186000.0
	Logarithm	0.75160	0.65982	0.71314	5199.98	202999.7
Pundir & Amala	Originals	0.23644	1	0	1099.00	11500.0
	Logarithm	0.19531	1	0	404.665	4230.60



(a) (b)

Figure 7: ROC curves assuming FGM dependence structure with Normal and Gamma marginal distributions for dengue data set. 7(a) ROC curve FGM-Normal for original data $\hat{AUC}$ 0.74123. 7(b) ROC curve FGM-Gamma for original data $\hat{AUC}$ 0.72981.

5. Conclusion and Remarks

Many procedures for the clinical classification of individuals include two biological traits (biomarkers whose natural behavior is modified in the presence of disease), and an error-free test known as the gold standard (which classifies individuals without error). Given that both biomarkers are measured in the same individual, it is necessary to include a dependence structure in the statistical model associated with the situation. It is possible that this dependence structure will not be perceptible using scatter plots of the data or commonly used indexes such as Pearson's rho, Spearma's rho or Kendall's tau. In this paper we studied the situation where we have two biomarkers expressed on a continuous scale and are assumed to have a very weak linear dependence structure or a very weak, but not necessarily linear dependence structure. To model the dependence structure, we used two copula functions: the FGM and the Gumbel Barnett copulas, within iterative procedure that allows to obtain the AUC for joint ROC curve.

It is important to point out that under the bivariate normality assumption, the Pundir and Amala's method works very well; but when the marginal distributions are not normal, this aproach does not permit reliable results; wherever inside that scenario the proposed methodology allowed us to obtain good quality estimates, because the method does not need the marginal distributions, a feature of the copula functions. Then the proposed method performs the estimation procedure using the normalized data obtained after we apply the inverse probability transformation, which eliminates the need to have normally-distributed marginals.

Our simulation study allowed us to see in a general way, the effect of the dependence structure between the biomarkers on the AUC estimates, controlling the marginal distributions effect. The FGM dependency does not really change the AUC estimates much, that is, the dependency effect is very weak. For GB-type dependencies, the effect is more evident and the specificity and AUC estimates are modified slightly.

Given that, the purpose is to estimate the joint AUC for the ROC curve, it is necessary to estimate the dependence parameter using the data set and add the estimate obtained to the algorithm developed to estimate the AUC. In this work we first obtained the maximum likelihood estimate; but other estimates obtained using the moments or bayesian methods could be considered.

It is important to note that when working with two biomarkers, the diagnostic procedure must be developed by joining these tests and observing the joint AUC (probability of properly classifying an individual with both tests at the same time). The proposed method is presented as an aid to this process, since once it detects the dependency between the biomarkers, it works correctly. Furthermore, it is proposed as an alternative to the study of weak or not necessarily linear dependency for this type of case.

It was observed that the biomarkers in the dengue detection study case have a weak but positive correlation, so FGM is a better option than the Gumbel-Barnett copula, since the latter considers weak but negative linear correlations. Generally, as mentioned by several authors (Achcar et al. 2019, Dendukuri &

Joseph 2001, Georgiadis et al. 2003, Pundir & Amala 2015, Tovar & Achcar 2011*a*), the correlation between the biomarkers used for clinical diagnosis shows a positive correlation. However, in terms of the method developed in this work of how to fit a copula-type dependency model, positive and negative correlation scenarios were considered, so the copula functions already presented were chosen.

Acknowledgements

We thank the program "Development of applied research to contribute to an effective and sustainable model of dengue intervention in Santander, Casanare and Valle del Cauca" of the AEDES Knowledge and Cooperation Network (RedAedes), Comfandi and the clinical and epidemiological experts participating in the consultations. This work was partially supported by the Science, Technology and Innovation Fund - FCTeI of SGR, Colombia - BPIN 2013000100011 and the AEDES and Comfandi Network. The participation of the second author was partially financed by a scholarship from the Virginia Gutierrez de Pineda Program for Young Researches and Innovators of the Administrative Department of Science, Technology and Innovation (Colciencias) in Colombia.

Appendix A. Estimation of Parameters Jointly

Let ϑ be a vector of parameters to be estimated and Θ the parameter space. The likelihood for observation q is denoted $L_q(\vartheta)$. Let $\ell_q(\vartheta)$ be the log-likelihood of $L_q(\vartheta)$. Given Q observations the log-likelihood function is defined as (Bouyé et al. 2000):

$$\ell(\vartheta) = \sum_{q=1}^{Q} \ell_q(\vartheta) \tag{A1}$$

For the case of the copula function, the log-likelihood function is given by:

$$\ell(\vartheta) = \sum_{q=1}^{Q} \ln\left(c_{\mathbf{X}}(G_{X_1}(x_{1q}), G_{X_2}(x_{2q}))\right) + \sum_{q=1}^{Q}\sum_{r=1}^{2} \ln\left(g_{X_r}(x_{rq})\right) \tag{A2}$$

Hence, $\hat{\vartheta}_{MLE}$ is given by:

$$\hat{\vartheta}_{MLE} = arg \max_{\vartheta \in \Theta} \ \ell(\vartheta) \tag{A3}$$

Appendix B. Relationship Between the Correlation Coefficients and Dependency Copula Parameters

The Spearman's correlation coefficient can be written in terms of the Pearson's correlation coefficient (Kruskal 1958):

$$\rho_s = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right),$$

where ρ_s represents the Spearman's correlation coefficient and ρ denotes the Pearson's correlation coefficient. Thus

$$\rho = 2 \sin\left(\frac{\pi \rho_s}{6}\right)$$

From Nelsen (2006) it follows that for FGM:

$$\begin{aligned}
\rho_s &= \left[12 \int_0^1 \int_0^1 uv[1 + \varphi(1-u)(1-v)] \; du \; dv\right] - 3 \\
&= 12\left(\frac{1}{4} + \frac{\varphi}{36}\right) - 3 \\
&= \frac{\varphi}{3}
\end{aligned}$$

So,

$$\hat{\rho} = 2 \sin\left(\frac{\pi \hat{\varphi}}{18}\right) \tag{B1}$$

Similarly (Kruskal 1958):

$$\tau = \frac{2}{\pi} \arcsin \rho,$$

where τ represents the Kendall's correlation coefficient. Thus

$$\rho = \sin\left(\frac{\pi \tau}{2}\right)$$

It follows that for FGM (Nelsen 2006):

$$\begin{aligned}
\tau &= 12 \left[\int_0^1 \int_0^1 [uv + \varphi uv(1-u)(1-v)][1 + \varphi(1-2u)(1-2v)] \; du \; dv\right] - 1 \\
&= \frac{2\varphi}{9}
\end{aligned}$$

So,

$$\hat{\rho} = \sin\left(\frac{\pi \hat{\varphi}}{9}\right) \tag{B2}$$

The relationship showed in the formulas: B1 and B2, is valid only for bivariate normal distributions.

For the Gumbel-Barnett copula function, the above calculations show extensive and unwieldy forms that depend on the mathematical function $Ei(\cdot)$ known as *exponential integral*, as well

$$\tau = \left[7 - \frac{2}{\phi}\right] e^{\frac{2}{\phi}} Ei\left(-\frac{2}{\phi}\right) - 4e^{\frac{1}{\phi}} Ei\left(-\frac{1}{\phi}\right) + 4e^{\frac{4}{\phi}} Ei\left(-\frac{4}{\phi}\right) - 1$$

and

$$\rho_s = 12\left[-\frac{e^{\frac{4}{\phi}}}{\phi} Ei\left(-\frac{4}{\phi}\right)\right] - 3$$

For this reason, as a moment estimator, the relationship between the Pearson correlation coefficient and the bivariate exponential distribution function Type I is used, from which the Gumbel-Barnett copula function is generated, which offers greater mathematical simplicity. This relationship is (Portilla & Tovar 2018):

$$\hat{\rho} = -1 + \int_0^{\infty} \frac{e^{-z}}{1 + \phi z}\, dz \tag{B3}$$

Appendix C. R Codes (FGM Code)

```
###Tasa de Falsos positivos
fpr <-function(t1,t2,mx1,mx2,sx1,sx2,Maxc1){
ax <-array(dim = length(t1)); bx <-array(dim = length(t1));
fpr <-array(dim = length(t1));
fhi<-Maxc1
for(i in 1:length(t1))
{
  ax[i] <-pnorm(t1[i],mx1,sx1); bx[i] <-pnorm(t2[i],mx2,sx2);
  fpr[i] <-integrate(
    function(x){
      sapply(x,function(x){
        integrate(function(y){1+(fhi*(1-2*x)*(1-2*y))},
        ax[i],1)$value})},bx[i],1)$value
}
return(fpr);
}

###Tasa de verdaderos positivos
tpr <-function(t1,t2,my1,my2,sy1,sy2,Maxc2){
ay <-array(dim = length(t1)); by <-array(dim = length(t1));
```

```
tpr <-array(dim = length(t1));
fhi<-Maxc2
for(i in 1:length(t1))
{
ay[i] <-pnorm(t1[i],my1,sx1); by[i] <-pnorm(t2[i],my2,sy2);
tpr[i] <-integrate(
function(x){
sapply(x,function(x){
  integrate(function(y){1+(fhi*(1-2*x)*(1-2*y))},
  ay[i],1)$value})}
,by[i],1)$value
}
return(tpr)
}

###Generacion de datos
FGM<-function(n,fi){
  ###Generacion de variables uniforme
  V1<-runif(n,0,1)
  V2<-runif(n,0,1)
  U1<-V1
  A=fi*((2*U1)-1)-1
  B_0=1-(2*fi*((2*U1)-1))+ ((fi^2)*((2*U1)-1)^2)
   + (4*fi*V2)*((2*U1)-1)
  B=sqrt(B_0)
  U2<-(2*V2)/(B-A)
  BD<-matrix(cbind(U1,U2),n,2)
  return(BD)
}
col1=colors()
phi=0.2
cv<-.15
mu1<-164##Sanos prueba 1
mu2<-160##Sanos prueba 2
mu3<-224##Enfermos prueba 1
mu4<-209##Enfermos prueba 2
sigma1<-cv*mu1 ;sigma2<-cv*mu2;
sigma3<-cv*mu3; sigma4<-cv*mu4

M=1000
Espe=0
Sensi=0
t11=0
t21=0
aucF=0
jen=0
```

```
B=1000
Intervalos<-matrix(0,M,2)

Prevalencia=0.1
Sa=round(10000*(1-Prevalencia))
En=10000-Sa

for(k in 1:M){
aucB=0
test.san<-FGM(Sa,phi)
x1<-qnorm(test.san[,1],mu1,sigma1)
x2<-qnorm(test.san[,2],mu2,sigma2)

test.enf<-FGM(En,phi)
y1<-qnorm(test.enf[,1],mu3,sigma3)
y2<-qnorm(test.enf[,2],mu4,sigma4)

c2=rbind(x1,x2);c1=rbind(y1,y2)

###construye la curva ROC
datos_FGM<-test.enf
LogVerosimi<-function(fi){Like1<-(-sum(log(1+
  (fi*(((2*datos_FGM[,1])-1)*((2*datos_FGM[,2])-1))))))}
Maxc2<-suppressWarnings(optim(c(runif(1)),LogVerosimi,
method = "BFGS",lower=-1,upper=1)$par)

datos_FGM<-test.san
LogVerosimi<-function(fi){Like1<-(-sum(log(1+
   (fi*(((2*datos_FGM[,1])-1)
*((2*datos_FGM[,2])-1))))))}
Maxc1<-suppressWarnings(optim(c(runif(1)),LogVerosimi,
method = "BFGS",lower=-1,upper=1)$par)

mx1= mu1;mx2=mu2;my1=mu3;my2=mu4;
sy1 =sigma3;sy2 = sigma4;sx1 =sigma1;sx2 =sigma2
t1<-sort(c(x1,y1)); t2 <-sort(c(x2,y2));
FPR <-fpr(t1,t2,mx1,mx2,sx1,sx2,Maxc1);
TPR <-tpr(t1,t2,my1,my2,sy1,sy2,Maxc2);
FPR1<-FPR;TPR1<-TPR
FPR = c(0,sort(FPR),1);TPR = c(0,sort(TPR),1);
library(bitops);library(caTools);
auc = trapz(FPR,TPR);dt1 <-data.frame(FPR,TPR)

plot(FPR,TPR,type="l")
lines(FPR,TPR,col=col1[k])
```

```
Indice<-((1-FPR1)+TPR1-1)
Espe[k]=1-FPR1[which.max(Indice)]
Sensi[k]=TPR1[which.max(Indice)]
t11[k]=t1[which.max(Indice)]
t21[k]=t2[which.max(Indice)]
aucF[k]=auc

for(kl in 1:B){
  elegir=sample(1:length(test.san[,1]),replace=T,size=length
     (test.san[,2])) BMuestra=test.san[elegir,]
  elegir=sample(1:length(test.enf[,1]),replace=T,size=length
     (test.enf[,2])) BMuestra1=test.enf[elegir,]

  x1<-qnorm(BMuestra[,1],mu1,sigma1)
  x2<-qnorm(BMuestra[,2],mu2,sigma2)
  y1<-qnorm(BMuestra1[,1],mu3,sigma3)
  y2<-qnorm(BMuestra1[,2],mu4,sigma4)

  c2=rbind(x1,x2);c1=rbind(y1,y2)

  datos_FGM<-test.enf
  LogVerosimi<-function(fi){Like1<-(-sum(log(1+(fi*
  (((2*datos_FGM[,1])-1)*((2*datos_FGM[,2])-1))))))}
  Maxc2<-suppressWarnings(optim(c(runif(1)),
  LogVerosimi,method = "BFGS",lower=-1,upper=1)$par)

  datos_FGM<-test.san
  LogVerosimi<-function(fi){Like1<-(-sum(log(1+(fi*
  (((2*datos_FGM[,1])-1)*((2*datos_FGM[,2])-1))))))}
  Maxc1<-suppressWarnings(optim(c(runif(1)),
  LogVerosimi,method = "BFGS",lower=-1,upper=1)$par)

  mx1= mu1;mx2=mu2;my1=mu3;my2=mu4;
  sy1 =sigma3;sy2 = sigma4;sx1 =sigma1;sx2 =sigma2
  t1<-sort(c(x1,y1)); t2 <-sort(c(x2,y2));
  FPR <-fpr(t1,t2,mx1,mx2,sx1,sx2,Maxc1);
  TPR <-tpr(t1,t2,my1,my2,sy1,sy2,Maxc2);
  FPR1<-FPR;TPR1<-TPR
  FPR = c(0,sort(FPR),1);TPR = c(0,sort(TPR),1);
  library(bitops);library(caTools);
  auc = trapz(FPR,TPR);dt1 <-data.frame(FPR,TPR)
  aucB[kl]<-auc
}

Intervalos[k,(1:2)]<-c(quantile(aucB,0.025),
               quantile(aucB,0.975))
```

```
lia<-lia+1
print(lia)
}
```

[Recibido: octubre de 2019 — Aceptado: mayo de 2020]

References

Achcar, J., Tovar, J. & Moala, F. (2019), 'Use of graphical methods in the diagnostic of parametric probability distributions for bivariate lifetime data in presence of censored data', *Journal of Data Science* **17**(3), 445–480.

Bamber, D. (1975), 'The area above the ordinal dominance graph and the area below the receiver operating characteristic graph', *Journal of Mathematical Psychology* **12**(4), 387–415.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. & Roncalli, T. (2000), 'Copulas for finance-a reading guide and some applications', *SSRN Electronic Journal*. 10.2139/ssrn.1032533.

Burgueño, M., García, J. & Gonzáles, J. (1995), 'Las curvas ROC en la evaluación de las pruebas diagnósticas', *Medicina Clínica* **104**, 661–670.

DeLong, E., DeLong, D. & Clarke-Pearson, D. (1988), 'Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach', *Biometrics* **44**(3), 837–845.

Dendukuri, N. & Joseph, L. (2001), 'Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests', *Biometrics* **57**(1), 158–167.

Dupuis, D. J. (2007), 'Using copulas in hydrology: Benefits, cautions, and issues', *Journal of Hydrologic Engineering* **12**(4), 381–393.

Etzionin, R., Kooperberg, C., Pepe, M., Smith, R. & Gann, P. (2003), 'Combining biomarker to detect disease with application to prostate cancer', *Biostatistics* **4**(4), 523–538.

Faraggi, D. & Reiser, B. (2002), 'Estimation of the area under the ROC curve', *Statistics in Medicine* **21**, 3093–3106.

Gallardo, B. (2010), Teoría de cópulas y aplicaciones en simulación de riesgos financieros y en ingeniería civil, Master Thesis, Estadística Aplicada, Universidad de Granada, Granada, España.

Genest, C., Quessy, J.-F. & Rémillard, B. (2006), 'Goodness-of-Fit Procedures for Copula Models Based on the Probability Integral Transformation', *Scandinavian Journal of Statistics* **33**(2), 337–366.

Georgiadis, M., Johnson, W., Gardner, I. & Singh, R. (2003), 'Correlation-Adjusted Estimation of Sensitivity and Specificity of Two Diagnostic Tests', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **52**(1), 63–76.

Gumbel, E. J. (1960), 'Bivariate Exponential Distributions', *Journal of the American Statistical Association* **55**(292), 698–707.

Johnson, M. E. (1987), *Multivariate statistical simulation*, first edn, John Wiley & Sons, Inc., Wiley Series in Probability and Statistics.

Kojadinovic, I., Yan, J. & Holmes, M. (2011), 'Fast large sample godness-of-fit test for copulas', *Statistica Sinica* **21**(2), 841–871.

Kruskal, W. H. (1958), 'Ordinal Measures of Association', *Journal of the American Statistical Association* **53**(284), 814–861.

Ma, S. & Huang, J. (2007), 'Combining multiple markers for classification using ROC', *Biometrics* **63**(3), 751–757.

National Institutes of Health and others (2004), US National Library of Medicine. https://medlineplus.gov/.

Nelsen, R. B. (2006), *An Introduction to Copulas*, second edn, Springer, New York.

Nikoloulopoulos, A. K. (2018), 'On composite likelihood in bivariate meta-analysis of diagnostic test accuracy studies', *AstA Adv Stat Anal* **102**, 211–227.

Pepe, M. S. (2003), *The Statistical Evaluation of Medical Tests for Classification and Prediction*, first edn, Oxford University Press, Oxford University, New York.

Pepe, M. S. & Thompson, M. L. (2000), 'Combining diagnostic test results to increase accuracy', *Biostatistics* **1**(2), 123–140.

Portilla, J. & Tovar, J. (2018), 'Estimating the Gumbel-Barnett copula parameter of dependence', *Revista Colombiana de Estadística* **41**(1), 53–73.

Pundir, S. & Amala, R. (2012), 'A Study on the Bi-Rayleigh ROC Curve Model', *Bonfring International Journal of Data Mining* **2**(2), 42–47.

Pundir, S. & Amala, R. (2015), 'Detecting diagnostic accuracy of two biomarkers through a bivariate log-normal ROC curve', *Journal of Applied Statistics* **42**(12), 2671–2685.

Su, J. Q. & Liu, J. S. (1993), 'Linear Combinations of Multiple Diagnostic Markers', *Journal of the American Statistical Association* **88**(424), 1350–1355.

Sumi, N. & Hossain, M. (2012), 'A study on parametric approaches to compare areas under two correlated ROC curves', *Bangladesh Journal of Scientific Research* **25**(1), 61–71.

Tovar, J. R. (2011), 'Métodos estadísticos desarrollados para la estimación de la prevalencia y parámetros de desempeño de tests para diagnóstico clínico: una revisión de literatura', *Revista Investigaciones Andina* **13**(23), 338–351.

Tovar, J. R. & Achcar, J. A. (2011*a*), 'Dependence between three diagnostic tests in presence of verification bias: a copula function approach', *Revista Brasileira de Biometria* **49**(1), 74–90.

Tovar, J. R. & Achcar, J. A. (2011*b*), 'Indexes to measure dependence between clinical ciagnostic tests: a comparative study', *Revista Colombiana de Estadística* **34**(3), 433–450.

Tovar, J. R. & Achcar, J. A. (2012), 'Two depedent diagnostic tests: use of copula functions in the estimation of the prevalence and performance test parameters', *Revista Colombiana de Estadística* **35**, 331–347.

Tovar, J. R. & Achcar, J. A. (2013), 'Dependence between two diagnostic tests with copula functions approach: a simulation study', *Communications in Statistics: Simulation and Computation* **42**(2), 454–475.

Wang, M. C. & Li, S. (2012), 'Bivariate maker measurements and ROC analysis', *Biometrics* **68**(4), 1207–1218.

Youden, W. J. (1950), 'Index for rating diagnostic tests', *Cancer* **3**(1), 32–35.

Zou, K., O'Malley, A. & Mauri, L. (2007), 'Receiver-Operating characteristics analysis for evaluating diagnostic tests and predictive models', *Circulation* **115**(5), 654–657.