Research Article

ROC curve generalization for non-monotone relationships

Pablo Martínez-Camblor,[*†‡] Norberto Corral,[‡] Corsino Rey[¶‡] Julio Pascual,[§‡] Eva Cernuda-Morollón[§‡]

The ROC curve is a popular graphical method frequently used in order to study the diagnostic capacity of continuous markers. It represents in a plot the true-positive against the false-positive rates. Both the practical and theoretical aspects of the ROC curve have been extensively studied. Conventionally, it is assumed that the considered marker has a monotone relationship with the studied characteristic; i.e., the upper (lower) values of the (bio)marker are associated with a higher probability of a positive result. However, there exist real situations where both the lower and the upper values of the marker are associated with higher probability of a positive result. We propose a ROC curve generalization, $\mathcal{R}_g$, useful in this context. All pairs of possible cut-off points, one for the lower and another one for the upper marker values, are considered selecting the best of them. The natural empirical estimator for the $\mathcal{R}_g$ curve is considered and its uniform consistence and asymptotic distribution are derived. Finally, two real-world applications are studied.

Keywords: Area under the curve, Asymptotic distribution, Resampling methods, ROC curve.

1. Introduction

Receiver operating characteristic (ROC) curve is a popular and valuable tool commonly used in order to study the diagnostic performance of a particular marker. It displays in a plot the false-positive rate (FPR) or 1-specificity ($1 - S_P$) i.e., the inability of the test to recognize a negative subject (frequently healthy) as negative against the sensitivity (S_E) or true-positive rate (TPR) i.e., the capacity of a diagnostic criterion/test to identify a positive subject (frequently diseased) as positive.

As classifying individuals into groups is a common problem in a variety of contexts, the ROC curve has received great attention in the specialized literature. There exists a huge number of papers which deal with both the theoretical and practical aspects of the ROC curve and its related problems (see, for instance, Martínez-Camblor [1] for a recent review). The book of Zhou, Obuchowski and McClish [2] provide a deep study of different ROC curve issues.

Conventionally, without loss of generality (wlg), it is assumed that larger values of the marker indicate the larger confidence that a given subject is positive (diseased). Therefore, let χ and ξ be two continuous random variables representing the values of the diagnostic test for the negative (without the studied characteristic/normal) and the positive (with the characteristic/diseased) subjects, respectively, for a fixed point t (FPR), the ROC curve is defined by

$$\begin{aligned}\mathcal{R}(t) =& 1 - F_\xi(F_\chi^{-1}(1-t)) \\ =& \mathcal{P}\{\xi > F_\chi^{-1}(1-t)\} = \mathcal{P}\{1 - F_\chi(\xi) \leq t\} = F_{1-F_\chi(\xi)}(t),\end{aligned} \tag{1}$$

[†] Oficina de Investigación Biosanitaria de Asturies (OIB-FICYT)
[‡] Universidad de Oviedo
[¶] UCI Pediátrica, Departamento de Pediatría, Hospital Universitario Central de Asturias (HUCA)
[§] Área de Neurociencias, Servicio de Neurología, Hospital Universitario Central de Asturias (HUCA)
[*] *Correspondence to: Pablo Martínez-Camblor. Oficina de Investigación Biosanitaria de Asturies. Calle Matemático Pedrayes, 25, Entresuelo 33005 Oviedo. Asturies, Spain. Tel. +0034 985109805. E-mail: pablomc@ficyt.es*

where F_χ and F_ξ denote the cumulative distribution functions (CDF) for the variables χ and ξ, respectively. Although other different parametric (Hanley [3]), semi-parametric (Hsieh and Turnbull [4]) and non-parametric estimators (Zou, Hall and Shapiro [5]) have been proposed, the empirical estimator, based on replacing the involved unknown CDFs for their respective empirical cumulative distribution functions (ECDF) is the most popular one. Hence, let X and Y be two random samples drawn from ξ and χ respectively, for each $t \in [0,1]$ the empirical ROC curve estimator is defined by

$$\hat{\mathcal{R}}(t) = 1 - \hat{F}_n(X, \hat{F}_m^{-1}(Y, (1-t)), \tag{2}$$

where $\hat{F}_n(X,\cdot)$ is the ECDF referred to the sample X (with size n) and $\hat{F}_m^{-1}(Y,\cdot) = \inf\{s \in \mathbb{R} \,/\, \hat{F}_m(Y,s) \geq \cdot\}$ (with size m). Both the uniform consistence and the asymptotic behaviour of $\hat{\mathcal{R}}$ have been investigated under weak and standard assumptions (see, for instance, Hsieh and Turnbull [4]). In particular, if *(i)* both F_χ and F_ξ have continuous densities, f_χ and f_ξ, $f_\xi(F_\chi(t))/f_\chi(F_\chi(t))$ is bounded in any subinterval (a,b) of $(0,1)$ and *(ii)* $n/m \to \lambda$ as $min(n,m) \to \infty$, there exists a probability space on which one can define independent sequences of Brownian bridges, $\{\mathcal{B}_1^{(m)}(t)\}_{\{0\leq t\leq 1\}}$ and $\{\mathcal{B}_2^{(n)}(t)\}_{\{0\leq t\leq 1\}}$, such that, for each $t \in [0,1]$,

$$\sqrt{n}\cdot(\hat{\mathcal{R}}(t) - \mathcal{R}(t)) = \lambda^{1/2}\cdot r(t)\cdot \mathcal{B}_1^{(m)}(1-t) + \mathcal{B}_2^{(n)}(1-\mathcal{R}(t)) + o(1) \quad a.s. \tag{3}$$

where $r(t) = \partial\mathcal{R}(t)/\partial t$. However, sometimes not only the larger (lower) marker values are associated with disease, but both the lower and larger values are related with the presence of the studied feature. For instance, in haemodialysis population, both the high and low levels of serum iPTH, calcium and phosphate are associated with higher risk of mortality [6]. Also, in the intensive care units, leukocyte counts greater than 20,000 (leukocytosis) or below 5,000 (leukopenia) are associated with bad prognosis in critically ill patients.

Obviously, this topic has already been pointed out. Hilden [7] exposed the following theoretical example: the (bio)marker values (T) for half of the diseased subjects are less than 80 and the other half have values greater than 120 while the (bio)marker values for the healthy patients are always between 80 and 120. Zhou, Obuchowski and McClish [2] reduced this problem to the right-size one by performing the transformation $T' = |T - 100|$. In this case, this transformation solves the problem. However, it is not always possible to find, easily, a suitable transformation, specially, when the marker distribution for the diseased or for the healthy subjects is skewed.

Figure 1 depicts the density functions for the negative and the positive populations (left) and the resulting right-side ROC curve (right). Note that, in this case, the sole difference between the negative and the positive subjects is the variability. Although not so extreme, the depicted situation is similar to that described by Hilden [7]. From the usual right-side approach, the classification based on the point $\boldsymbol{B}$ obtains a higher percentage of false-positives (97.7%), than true-positives (81.4%). However, if a subject with a marker value lower than -2 or greater than 2 is declared as positive and it is declared as negative otherwise, the false-positive and the true-positive percentages are 4.6% and 37.1%, respectively. This criterion is equivalent to use a threshold of 2 on the transformed marker $T' = |T|$.

In this paper, the authors propose a ROC curve generalization useful in the above context. It considers all possible pairs of cut-off points, one for the lower bound and another one for the upper bound; those obtaining the same specificity and achieve the largest sensitivity, are selected. The uniform consistence and the asymptotic distribution for its direct empirical estimator are derived in Section 2. Section 3 is devoted to study the properties of the associated area under the curve,

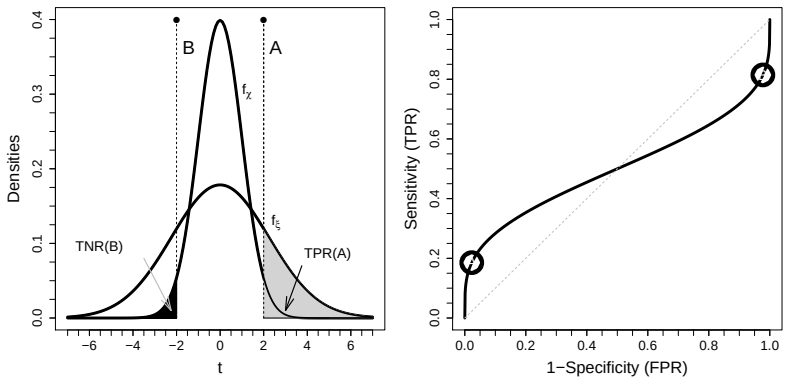

Figure 1: Density functions for the negative, f_χ, and the positive, f_ξ, populations (left) and the resulting ROC curve (right).

AUCg. In Section 4, this methodology is applied to two real data sets. On one hand, we study the association between the leukocyte count and the mortality risk in critically ill children. And, on the other hand, we study the relationship between the Calcitonin gene related peptide (CGRP) levels and the response to the *OnabotulinumtoxinA* treatment for the chronic migraine in adults. The results which guarantee the uniform consistence and the asymptotic distribution for the empirical general ROC curve estimator and a `R` function (`Rg`) to compute the proposed general ROC curve are provided as supplementary material.

2. ROC generalization: Definition and properties

For each fixed threshold, $x \in \mathbb{R}$, the right-side ROC curve is the pair $\{1 - S_P(x), S_E(x)\} = \{1 - F_\chi(x), 1 - F_\xi(x)\}$. If $t = 1 - F_\chi(x)$, the well-known ROC curve expression $\{t, \mathcal{R}(t)\}$ is obtained. When different thresholds drive to the same specificity, the one which maximizes the sensitivity is taken (in the usual right-side convention, the minimum of all possible thresholds). When both the larger and lower values of the marker are associated with a higher probability of disease, two thresholds $x_l, x_u \in \mathbb{R}$ $(x_l \leq x_u)$ are necessary in order to make a diagnosis. A subject is classified as positive when its marker value is below x_l or greater than x_u; therefore, the specificity is defined by $S_P(x_l, x_u) = F_\chi(x_u) - F_\chi(x_l)$. For arbitrary $t \in [0, 1]$, let be $\mathcal{F}_t = \{(x_l, x_u) \in \mathbb{R}^2$ such that $x_l \leq x_u$ and $F_\chi(x_l) + 1 - F_\chi(x_u) = t\}$, the generalized ROC curve is defined by

$$\mathcal{R}_g(t) = \sup_{(x_l, x_u) \in \mathcal{F}_t} \{F_\xi(x_l) + 1 - F_\xi(x_u)\}.$$

Obviously, for each $(x_l, x_u) \in \mathcal{F}_t$ there exists a $\gamma \in [0, 1]$ such that $F_\chi(x_l) = \gamma \cdot t$, and then $x_l = F_\chi^{-1}(\gamma \cdot t)$. Due to $(x_l, x_u) \in \mathcal{F}_t$, then $F_\chi(x_u) = 1 - [1 - \gamma] \cdot t$, and $x_u = F_\chi^{-1}(1 - [1 - \gamma] \cdot t)$. Therefore,

$$\mathcal{R}_g(t) = \sup_{\gamma \in [0,1]} \{F_\xi(F_\chi^{-1}(\gamma \cdot t)) + 1 - F_\xi(F_\chi^{-1}(1 - [1 - \gamma] \cdot t))\},$$

or equivalently,

$$\mathcal{R}_g(t) = \sup_{\gamma \in (0,1)} \{1 - \mathcal{R}(1 - \gamma \cdot t) + \mathcal{R}([1 - \gamma] \cdot t)\}. \tag{4}$$

If $\gamma_t = \arg\sup_{\gamma \in [0,1]} \{1 - \mathcal{R}(1 - \gamma \cdot t) + \mathcal{R}([1 - \gamma] \cdot t)\}$ (if it is not a unique value, we take the minimum among all possibilities) the above expression is equivalent to,

$$\mathcal{R}_g(t) = \{1 - \mathcal{R}(1 - \gamma_t \cdot t) + \mathcal{R}([1 - \gamma_t] \cdot t)\}.$$

The γ_t value determines the optimum proportion of false positives, $\gamma_t \cdot t$, committed by the left (negative subjects with a marker below x_l), and by the right, $[1 - \gamma_t] \cdot t$ (negative subjects with a marker larger than x_u). Logically, under symmetry (Figure 1), the optimum value is $1/2$ for all $t \in [0, 1]$.

Directly, $\mathcal{R}_g$ generalizes the usual right-side (left-side) ROC curve. When lower (upper) marker values are associated with normal subjects, for each $t \in [0, 1]$, $\gamma_t = 0$ $(\gamma_t = 1)$ and $\mathcal{R}_g = \mathcal{R}$ $(\mathcal{R}_g(t) = 1 - \mathcal{R}(1 - t))$. Note that, in the *Hilden* case, enunciated above, $\mathcal{R}_g = 1$ when $\gamma_t = 1/2$ for all $t \in [0, 1]$. In addition, taking into account $\mathcal{R}_g$ is mainly based on the usual ROC curves, it inherits most of its properties, for instance, the invariance for monotone increasing transformations [4].

Figure 2 represents the $\mathcal{R}_g$ construction for the populations considered in Figure 1. In this case, $\mathcal{R}(t) = 1 - \mathcal{R}(1 - t)$ and the optimum value for γ is $1/2$ for all $t \in [0, 1]$. The set of all possible trajectories of the curve, $1 - \mathcal{R}(1 - \gamma \cdot t) + \mathcal{R}([1 - \gamma] \cdot t)$ for $\gamma \in [0, 1]$, defines the gray zone. In this case, this area coincides with the difference between the area under the right-side (left-side) ROC curve and the area under the $\mathcal{R}_g$ curve.

Conventionally, let X and Y be two independent random samples drawn from the random variables ξ and χ respectively, then for each $t \in [0, 1]$, the empirical general ROC curve estimator is,

$$\hat{\mathcal{R}}_g(t) = \sup_{\gamma \in [0,1]} \left\{1 - \hat{\mathcal{R}}(1 - \gamma \cdot t) + \hat{\mathcal{R}}([1 - \gamma] \cdot t)\right\}, \tag{5}$$

where $\hat{\mathcal{R}}$ is the empirical ROC curve estimator defined in equation 2. The $\hat{\mathcal{R}}_g$ estimator has good asymptotic properties. In particular, in the current manuscript, results to guarantee both the uniform consistency and the asymptotic distribution are included as supplementary material.

Figure 3 is similar to Figure 2 but empirical curves based on independent random samples of positive and negative populations (sharing the same sample size, 100) were used instead of the theoretical ones. Here and onwards, the values of γ_t were computed numerically. For $t = 1/2$, the observed $\hat{\mathcal{R}}_g$ value was 0.89. This value was obtained for $\gamma_t = 0.481$ ($\hat{\mathcal{R}}(1 - \gamma_t \cdot t) = \hat{\mathcal{R}}(0.759) = 0.60$ and $\hat{\mathcal{R}}([1 - \gamma_t] \cdot t) = \hat{\mathcal{R}}(0.259) = 0.49$). In this case, the minimum observed value for $\{1 - \hat{\mathcal{R}}(1 - \gamma \cdot t) + \hat{\mathcal{R}}([1 - \gamma] \cdot t)\}$ was 0.42, obtained for $\gamma = 1$ (usual left-side ROC curve).

Additionally, empirical results, based on simulations (results not shown), suggest that the convergence ratio of the $\hat{\mathcal{R}}_g$ to its target is similar to the convergence ratio for the empirical ROC curve estimator. However, it is worth to point out that, when the diagnostic capacity of the marker is poor (close to the main diagonal), $\hat{\mathcal{R}}_g$ over-estimates the marker diagnostic capacity.

3. Area under the curve

The area under the ROC curve (AUC) is the most commonly used global index of diagnostic accuracy (Fluss, Faraggi and Reiser [8]). It ranges between 0.5 and 1 and represents the overall performance of a diagnostic test, in terms of its accuracy, at all relevant diagnostic thresholds (cut-off points) used to discriminate subjects with or without the considered characteristic. Moreover, in the usual right-side (left-side) assumption the AUC (labeled by $\mathcal{A}$) is

$$\mathcal{A} = \int_0^1 \mathcal{R}(t)dt = 1 - \int F_\xi(u)dF_\chi(u) = \mathcal{P}\{\xi > \chi\}.$$

Therefore, in this case, the AUC has a direct probabilistic interpretation. In particular, the AUC is the probability that the value of the marker in a randomly chosen positive subject will be higher than the value of the marker in a randomly chosen

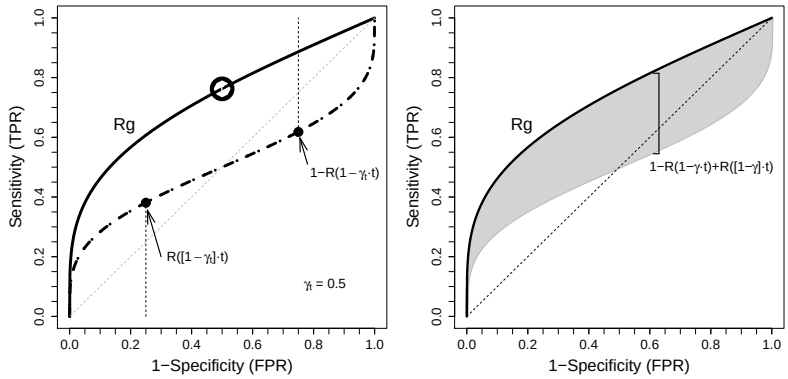

Figure 2: $\mathcal{R}_g$ construction for the populations considered in the Figure 1 (left). In gray, the set of all possible trajectories of $\{1 - \mathcal{R}(1 - \gamma \cdot t) + \mathcal{R}([1 - \gamma] \cdot t)\}$ for $\gamma \in [0, 1]$ and the $\mathcal{R}_g$ curve, in black (right).

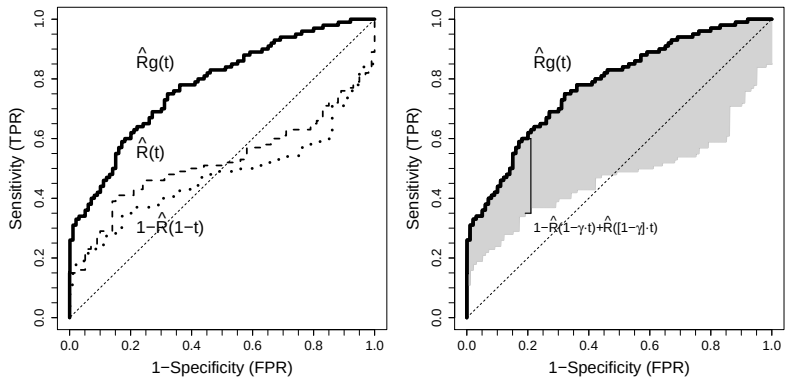

Figure 3: $\hat{\mathcal{R}}_g$ construction for two samples of negative and positive populations ($n = 100$, $m = 100$) drawn from the populations considered in the Figure 1 (left). In gray, all possible trajectories of $\{1 - \hat{\mathcal{R}}(1 - \gamma \cdot t) + \hat{\mathcal{R}}([1 - \gamma] \cdot t)\}$ for $\gamma \in [0, 1]$ and the $\hat{\mathcal{R}}_g$ curve, in black (right).

Table 1: Descriptive statistics for the leukocyte count in the low and high mortality risk (MR) groups. In particular, sample size (**N**), mean and standard deviation (**mean** $\pm$ **sd**), percentile 25 ($\boldsymbol{P}_{25}$), 50 ($\boldsymbol{P}_{50}$) and 75 ($\boldsymbol{P}_{75}$) are provided.

	N	**Mean** $\pm$ **sd**	$\boldsymbol{P}_{25}$	$\boldsymbol{P}_{50}$	$\boldsymbol{P}_{75}$
Low MR	176	14,290.7 $\pm$ 6,588	9,725	13,400	17,875
High MR	12	18,325.8 $\pm$ 20,343	7,250	11,400	22,850

negative subject. Of course, the area under the general ROC curve can also be directly computed and it can be read as the sensitivity average for the different specificities. Its value can still be used in order to summarize the accuracy of the studied test by a single number; but, unfortunately, it loses the probabilistic interpretation, even for γ_t constant, due to the relationship between the marker and the studied characteristic is not monotone.

The direct non-parametric estimator for the area under the generalized ROC curve, AUCg, is

$$\hat{\mathcal{A}}_g = \int_0^1 \hat{\mathcal{R}}_g(t)dt. \tag{6}$$

Theorem 2 (see supplementary material) guarantees that $\sqrt{n}\cdot(\hat{\mathcal{A}}_g - \mathcal{A}_g)$ ($\mathcal{A}_g = \int_0^1 \mathcal{R}_g(t)dt$) is, asymptotically, normal distributed with mean zero. The variance depends on the first derivative of the γ_t function and, in general, it is not possible to derive expressions useful in practice; therefore, some resampling method must be used in order to perform inferences. By using the traditional naive bootstrap method [9], the precision of the results, obtained via Monte Carlo simulations, was similar to that of the usual AUC estimator (results not shown). In the supplementary material, the explicit expression for the variance of $\sqrt{n}\cdot(\hat{\mathcal{A}}_g - \mathcal{A}_g)$, when $\gamma_t' = 0\,\forall t \in (0,1)$, is derived. Unfortunately, the simulation results suggest that this expression is not useful for general situations. In particular, when γ_t is replaced by $\int \gamma_t dt$ and a plug-in method is employed for estimating all the unknown parameters, the observed estimations for the variances were really unstable (results not shown).

4. Real data applications

4.1. Association between leukocyte count and mortality risk in critically ill children

Having available tools to determine the risk of mortality at admission to the Paediatric Intensive Care Unit (PICU), or during the first 24 hours after admission, is a clinical necessity [10]. Leukocyte count measurement constitutes a routinely determination when a patient is admitted to the PICU. Classically, low leukocyte or high leukocyte counts are described as one of the criteria for the diagnosis of sepsis, a severe systemic infection [11]. Recently, low leukocyte count was found as an independent risk factor of mortality in critically ill children with sepsis [12] as well as in adults with necrotizing pneumonia [12]. Therefore, the objective of our study was the investigation of leukocyte count in a well defined cohort of consecutive PICU patients to test the hypothesis that low or high leukocyte counts would be associated with increased prediction of mortality risk scores in critically ill children. We designed a prospective observational study set in two PICUs of university hospitals (8-bed PICU of Hospital Universitario Central de Asturias in Oviedo and 11-bed PICU of Hospital Universitario Gregorio Marañón in Madrid). The study protocol was approved by the Hospital Ethics Committee of Hospital Universitario Central de Asturias. The study was conducted in a number of consecutive patients, age below 18 years, who were admitted to one of these PICUs. Patients were divided in two groups according to mortality risk scores. Higher score risk mortality group (high MR) included patients with a PIM 2 and PRISM III scores greater than percentile 75 ($N = 12$); lower score risk mortality group (low MR) included patients with a PIM 2 and/or PRISM score below percentile 75 ($N = 176$). The interested reader is referred to Rey, García-Hernández, Concha et al. [10] for more information about this cohort. Leukocyte count routine determinations were performed during the first 12 hours after admission.

A total of 188 (61.7% women) patients were included in the study. Twelve of them were classified within the high mortality risk group (high MR) and 176 in the low mortality risk group (low MR). Patients in the high MR were younger (mean $\pm$ standard deviation: 31.67$\pm$ 42.38 months) than patients in the low MR, 67.32 $\pm$ 59.88 (the Student-Welch test provided a p-value of 0.016). Table 1 shows some descriptive statistics for the leukocyte count in both the low and high mortality risk groups.

Note that, in the high MR group the observed variability of the leukocyte count was really large. In addition, in the low MR group, P_{25} was larger and P_{75} lower than in the high MR group. When an usual right-side (left-side) ROC curve is

plotted, it crosses the diagonal line and obtains a low AUC value. In particular, the observed AUC for the left-side ROC curve was 0.520 with a 95% (naive) bootstrap (based on 5,000 iterations) confidence interval of (0.275-0.693). The general ROC curve detects that both the lower and the higher values of the marker are associated with the high MR group and achieved an $\mathcal{A}_g$ of 0.740 with a 95% (naive) bootstrap (based on 5,000 iterations) confidence interval of (0.658-0.884). If the children with leukocyte count lower than 24,900 (10.12 in logarithmic scale) and greater than 10,200 (9.23 in logarithmic scale) were classified within the low MR group (optimal thresholds in the Youden index sense) the observed TPR and the TNR were 0.750 and 0.636, respectively. Figure 4 depicts the kernel density estimation for the leukocyte count in logarithmic scale for both the high (black) and low (gray) mortality risk groups and the right-side, the left-side and the general ROC curves.

A moderate increase in the number of leukocytes is the normal *stress response* of the body when it suffers an injury being considered one of the main elements in the systemic response syndrome. This inflammatory response favours the recovery of the patient. However, when the response is very intense (large increase in leukocyte count) or very poor (decrease in leukocyte count) the recovery of the patient is compromised. Therefore, mortality risk, especially in critically ill patients, increases.

4.2. Relationship between the response to the OnabotulinumtoxinA treatment for the chronic migraine headaches and the CGRP levels in women

The *OnabotulinumtoxinA* is the first and only FDA-approved (United State food and drug administration), preventive treatment for chronic migraine in adults (see, for instance, Dodick et al. [14]). On the other hand, serum CGRP levels are increased in chronic migraineurs indicating a chronic activation of the trigemino-vascular system, and it is proposed as the first biomarker for this entity [15]. For this reason, we explored the relationship between basal levels of CGRP, determined as it was previously described, and the response to the *OnabotulinumtoxinA* in 70 women (age ranges between 20 and 63 years, mean $\pm$ standard deviation: 46.4 $\pm$ 10.3) meeting chronic migraine criteria. The study was approved by the ethic review board of the Hospital Universitario Central de Asturias (HUCA) and the patients gave written consent. Migraine patients are usually considered as responders when attack frequency is decreased by 50%, so we adopted this criterion and we checked it by the use of monthly headache calendars in all patients.

The observed response percentage was 78.6% (55/70). Table 2 shows the main descriptive statistics for the CGRP levels in the response and non-response groups. The values in the response group were, in general, higher than in the non-response group. However, the minimum value in the response group (11.44) was lower than the minimum value in the non-response group (27.44).

Figure 5 depicts, at left, the kernel density estimations for both the response (black) and the non-response (gray) groups for the CGRP levels in logarithmic scale. At right, the right-side ($\hat{\mathcal{R}}(t)$), the left-side ($1-\hat{\mathcal{R}}(1-t)$) and the general ROC curves. The area under the right-side ROC curve was 0.619 (0.473-0.765) (between brackets a 95% (naive) bootstrap (based on 5,000 iterations) confidence intervals) while $\hat{\mathcal{A}}_g$ was 0.731 (0.639-0.881). Optimal thresholds (maximizing TPR-FPR) were (36.51, 66.97); i.e, women with CGRP levels below 36.51 (3.60 in logarithmic scale) and larger than 66.97 (4.20 in logarithmic scale) are the optimal group. Particularly, with these cut-off points, the FPR and the TPR were 0.267 and 0.673, respectively.

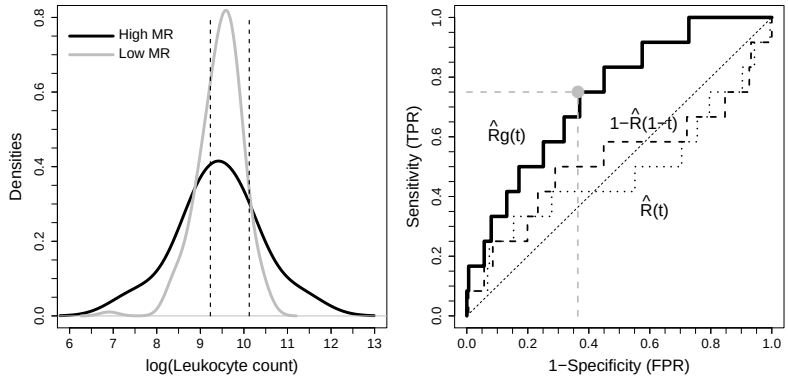

Figure 4: At left, kernel density estimation for the leukocyte count in logarithmic scale for both the high (black) and low (gray) mortality risk groups. Vertical lines stand for the optimal thresholds. At right, right-side ($\hat{\mathcal{R}}$), left-side ($1-\hat{\mathcal{R}}(1-t)$) and general ROC curves. In gray, TPR and FNR achieved by the optimal threshold (in the Youden index sense).

Table 2: Descriptive statistics for the CGRP levels in the response and non-response groups. In particular, sample size (**N**), mean and standard deviation (**mean** $\pm$ **sd**), percentile 25 ($\boldsymbol{P}_{25}$), 50 ($\boldsymbol{P}_{50}$) and 75 ($\boldsymbol{P}_{75}$) are provided.

	N	**Mean** $\pm$ **sd**	$\boldsymbol{P}_{25}$	$\boldsymbol{P}_{50}$	$\boldsymbol{P}_{75}$
No response	15	57.68 $\pm$ 22.01	38.71	50.45	66.91
Response	55	70.97 $\pm$ 33.01	45.01	71.04	88.34

The remarkable response in a subset of patients with high CGRP levels supports its value as a biomarker of chronic migraine and suggests that the inhibition of CGRP release could be a mechanism accounting for the *OnabotulinumtoxinA* positive effect in chronic migraine. There are several potential explanations for the fact that a subset of patients with low CGRP levels seemed to respond to botulinum toxin A treatment. The first could simply be a placebo response, considering that about one-third of patients with frequent migraine respond to placebo in clinical trials and that this response increases with injections versus oral medications. Second, some patients with a chronic migraine phenotype and low CGRP levels could, in fact, suffer from a psychogenic headache instead of a true chronic migraine. Finally, it is also possible that; for some patients, other pain-producing peptides, such as VIP or PACAP, and not CGRP are released by the activated trigemino-vascular system.

5. Conclusions

In this paper, we explored a ROC curve generalization for the case where the relationship between the marker and the studied feature are not monotone i.e., extreme values of the marker are associated with high probability of having the considered characteristic. In spite of the proposed ROC curve generalization, $\mathcal{R}_g$, does not have a direct probabilistic interpretation; for a considered (bio)marker, the $\mathcal{R}_g(t)$ is (can be read as) the greatest reached sensitivity when the specificity is, at most, $1-t$. The explored approach has the advantage of solving the non-monotonicity problem without transforming the original data. Note that, a suitable transformation not always exists and, in addition, these transformations are usually made taking into account information extracted from the sample and then, the same sample is used in order to calculate the ROC curve. In the cases where such transformation exists, both methods result in the same ROC curve. In the explored datasets, working with the transformed markers $|T-\mu_0|$, $|T-m_0|$ and $|T-p_0|$ where μ_0, m_0 and p_0 stand for the mean, the median, and the middle point ($0.5\cdot[max-min]$) in the control group, respectively, and T is the original marker (leukocyte count and CGRP level, in the considered examples), the obtained curves improved the original right-side ones but both were worse than $\hat{\mathcal{R}}_g$. The best AUCs were 0.69 and 0.67 for the leukocyte and the CGRP data, respectively.

Additionally, its direct non-parametric empirical estimator is studied. It is the result of replacing the unknown CDFs for their usual non-parametric estimators, the respective ECDFs. From the properties of the traditional right-side ROC curve, both the uniform convergence and the asymptotic distribution are derived. Our simulations (not shown here) suggest that the quality of estimations provided by $\hat{\mathcal{R}}_g$ are similar to the usual $\hat{\mathcal{R}}$; although, logically, the new estimator overestimates

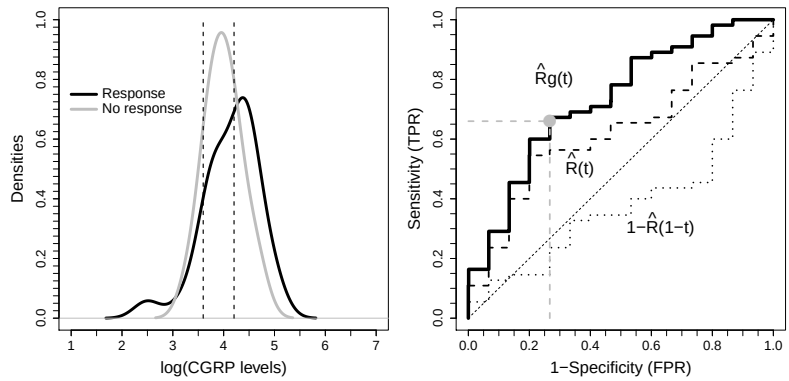

Figure 5: Left, kernel density estimation for the CGRP levels (logarithmic scale) for both the response (black) and non-response (gray) mortality risk groups. Vertical lines stand for the optimal thresholds. Right, right-side ($\hat{\mathcal{R}}$), left-side ($1-\hat{\mathcal{R}}(1-t)$) and general ROC curves. In gray, TPR and FNR achieved by the Youden-threshold.

the diagnostic capacity when that is poor. This problem is shared with the traditional ROC curve analysis if the desired direction (left or right) is not previously defined.

Probably, the main handicap of the proposed curve is that its associated area under the curve loses the direct probability interpretation. The AUCg can still be used in order to summarize the diagnostic capacity of the studied biomarker. It still ranges between 0.5 and 1 and higher values are associated with better diagnostic capacity. In addition, it can be read as the average of the achieved sensitivity for all possible specificities. The asymptotic normality of the area under the proposed general ROC curve estimator, $\hat{\mathcal{A}}_g$, has also been derived. However, the explicit expression for the variance depends on the first derivative of the function γ_t and it is not useful in practice. Unfortunately, the result obtained in Theorem 3 (see supplementary material) is not robust against the unrealistic assumption that $\gamma_t' = 0 \, \forall t \in (0,1)$; hence, in order to make inference, the use of some resampling method is advisable.

Finally, the proposed methodology is used in two real datasets. In the first one, we explored the use of the leukocyte count in order to predict the mortality risk in critically ill children. In the second one, it is studied the relationship between the response to the *OnabotulinumtoxinA* treatment for the chronic migraine headaches and the CGRP levels in 70 Spanish women. In this problem both, the larger and the lower values of CGRP are associated with higher probability of response. The proposed general ROC curve allows to detect a potential placebo effect in those women with low CGRP levels.

The customer software used for developing the computes involved in this work (in particular, the `R` function called `Rg`) is included as supplementary material. The authors expect developing additional functions which perform ROC curve comparison, ROC meta-analysis and other proposed solutions for different problems related with the ROC curve and uploading all together in a `R package`.

Acknowledgements

The authors are grateful to Susana Diaz-Coto for the manuscript revision. We want to thank to the PICUs of the Hospital Universitario Central de Asturias (HUCA) and of the Hospital Universitario Gregorio Marañón for the permission to use the PICU data. This work was supported by the Grant MTM2011-23204 of the Spanish Ministry of Science and Innovation (FEDER support included) and PI11/00889 FISSS grant (Fondos Feder, ISCIII, Ministry of Economy, Spain) and Allergan grant to JP.

References

1. Martínez-Camblor P, Area under the ROC curve comparison in the presence of missing data, *Journal of the Korean Statistical Society*, 2013, **42**(4), 431-442.
2. Zhou XH, Obuchowski NA, McClish DK, *Statistical methods in diagnostic medicine*, 2002, Wiley, New York.
3. Hanley JA, The use of the binormal model for parametric ROC analysis of quantitative diagnostic tests, *Statistics in Medicine*, 1996, **15**, 1575-1585.
4. Hsieh F, Turnbull BW, Nonparametric and semiparametric estimation of the receiver operating characteristic curve, *Annals of Statistics*, 1996, **24**(1), 25-50.
5. Zou KH, Hall WJ, Shapiro DE, Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests, *Statistics in Medicine*, 1997, **16**(19), 2143-2156.
6. Floege J, Kim J, Ireland E, Chazot C, Drueke T, de Francisco A, Kronenberg F, Marcelli D, Passlick-Deetjen J, Schernthaner G, Fouqueray B, Wheeler DC, Serum iPTH, calcium and phosphate, and the risk of mortality in a European haemodialysis population, *Nephrology Dialysis and Transplant*, 2011, **26**, 2948-1955.
7. Hilden J, The area under the ROC curve and its competitors, *Medical Decision Making*, 1991, **11**, 95-101.
8. Fluss R, Faraggi D, Reiser B, Estimation of the Youden Index and it's associated cutoff point, *Biometrical Journal*, 2005, **47**(4), 458-472.
9. Efron B, Tibshirani RJ, *An introduction to the bootstrap*, Chapman & Hall, 1993, London, England.
10. Rey C, García-Hernández I, Concha A, Martínez-Camblor P, Botrán M, Medina A, Prieto B, López-Herce J, Pro-adrenomedullin, pro-endothelin-1, procalcitonin, C-reactive protein and mortality risk in critically ill children: a prospective study, *Critical Care*, 2013, **17**:R240.
11. Goldstein B, Giroir B, Randolph A, International paediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in paediatrics, *Pediatric Critical Care Medicine*, 2005, **6**, 2-8.
12. Pérez DV, Jordan I, Esteban E, García-Soler P, Murga V, Bonil V, Ortiz I, Flores C, Bustinza A, Cambra FJ, Prognostic factors in paediatric sepsis study, from the Spanish society of paediatric intensive care, Pediatric Infection Disease Journal, 2014, **33**, 152-157.
13. Khanafer N, Sicot N, Vanhems P, Dumitrescu O, Meyssonier V, Tristan A, Bs M, Lina G, Vandenesch F, Gillet Y, Etienne J. Severe leukopenia in Staphylococcus aureus-necrotizing, community-acquired pneumonia: risk factors and impact on survival, *BMC Infectiosus Diseases*, 2013, 13:359.
14. Dodick DW, Turkel CC, DeGryse RE, Aurora SK, Silberstein SD, Lipton RB, Diener HC, Brin MF, OnabotulinumtoxinA for treatment of chronic migraine: Pooled results from the double-blind, randomized, placebo-controlled phases of the PREEMPT clinical program, *Headache: The Journal of Head and Face Pain*, 2010, **50** 921-936.
15. Cernuda-Morollón E, Larrosa D, Ramón C, Vega J, Martínez-Camblor P, Pascual J, Interictal increase of CGRP levels in peripheral blood as a biomarker for chronic migraine, *Neurology*, 2013, **81**(14), 1191-1196.