REVSTAT – Statistical Journal
Volume 12, Number 1, March 2014, 21–41

A REVIEW ON ROC CURVES IN THE PRESENCE OF COVARIATES

Authors: Juan Carlos Pardo-Fernández
– Department of Statistics and Operations Research, Universidade de Vigo, Spain
juancp@uvigo.es

María Xosé Rodríguez-Álvarez
– Department of Statistics and Operations Research, Universidade de Vigo, Spain
mxrodriguez@uvigo.es

Ingrid Van Keilegom
– Institute of Statistics, Université catholique de Louvain, Belgium
ingrid.vankeilegom@uclouvain.be

Abstract:

- In this paper we review the literature on ROC curves in the presence of covariates. We discuss the different approaches that have been proposed in the literature to define, model, estimate and do asymptotics for ROC curves that incorporate covariates. For reasons of brevity, we mostly focus on nonparametric approaches, although some parametric and semiparametric methods are also discussed. We also analyze endocrinological data on the body mass index to illustrate the methodology. Finally, we mention some research topics that need further investigation or that are still unexplored.

Key-Words:

- *covariate; kernel smoothing; location-scale model; nonparametric inference; regression; ROC curve.*

AMS Subject Classification:

- 62-02, 62G08, 62H30.

1. INTRODUCTION

ROC curves are a very useful instrument to measure how well a variable or a diagnostic test is able to distinguish two populations from each other. It is therefore an essential element in the classification and discrimination literature, and it has interested and still interests many statisticians from a theoretical as well as from an applied point of view.

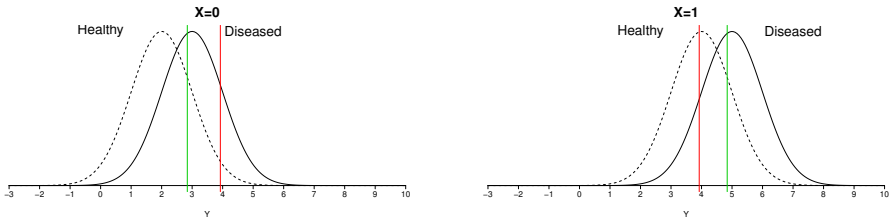
When covariates are present, it might be advisable to incorporate them in the ROC curve in order to make use of the additional information. In fact, in many situations the performance of a diagnostic test and, by extension, its discriminatory capacity can be affected by covariates. Pepe (2003, pp. 48–49) provides several examples of covariates that can affect a test result. For instance, patient characteristics, such as age and gender, are important covariates to be considered. Furthermore, when a diagnostic test is performed by a tester (*e.g.*, a radiologist engaged in interpreting images), a characteristic of the tester, such as experience or expertise, will often affect the test result. The incorporation of covariates into the ROC curve might be done for two purposes: (a) obtain covariate-specific ROC curves, or ROC curves that condition on a specific value of a covariate vector; and (b) get some kind of average ROC-curve, or covariate-adjusted ROC curve, which takes the covariate information of each data point into account in order to obtain a better measure of the discriminatory capacity than the rude 'marginal' or 'pooled' ROC curve.

In this paper we first explain in Section 2 why it is important to take covariate information into account by giving some concrete examples of situations where the covariates have an impact on the performance of the diagnostic test and/or its discriminatory capacity. We next consider in Section 3 both the covariate-specific and the covariate-adjusted ROC curve, and we give an overview of estimation methods that have been proposed for both concepts in Section 4. The focus lies on reviewing the literature and not on giving detailed derivations or lengthy discussions. They can be found in the respective papers. For reasons of brevity, we mostly focus on nonparametric approaches, although some parametric and semiparametric methods are also discussed. In Section 5 we analyze endocrinological data on the body mass index to illustrate the methodology. Finally, in Section 6 we mention some research topics that need further investigation or that are still unexplored.

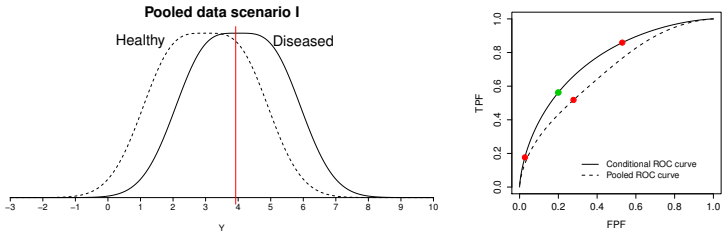
2. MOTIVATION

This section is devoted to motivating the need for incorporating covariates into the ROC analysis by means of illustrating the consequences that ignoring such information may have on the practical conclusions drawn from the study at hand. In brief, there are two different scenarios on which covariate information will have to be incorporated into the ROC analysis: (a) when the performance of the diagnostic test is affected by covariates, but not its discriminatory capacity; and, (b) when the discriminatory capacity itself is affected. A good overview of this aspect can be found in Janes and Pepe (2008, 2009a) and in fact, the examples given here are partially based on both papers. For a more detailed review of the subject, readers are urged to consult these references.

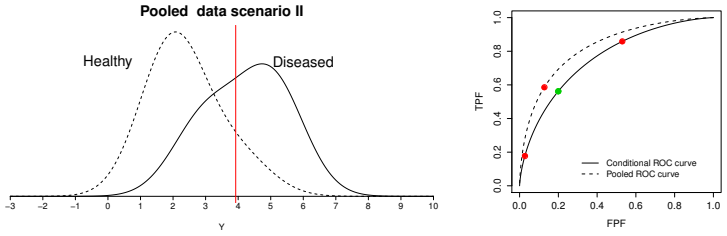
On the one hand, let us start with those situations in which the performance of the diagnostic test is affected by covariates, even where the discriminatory capacity of the test is unaffected. This situation is depicted in Figure 1(a), in which a covariate $\boldsymbol{X}$ (*e.g.*, patient gender) affects the result but not the discriminatory capacity of diagnostic test Y. In other words, the separation between the conditional distributions of the diagnostic test result in both healthy and diseased populations is the same, irrespective of the values of covariate $\boldsymbol{X}$. In Figure 1(b), covariate $\boldsymbol{X}$ is independent of disease status, which will be denoted by D (diseased) and $\bar{D}$ (healthy), *i.e.*, the result of the diagnostic test changes according to the gender of the patient but the prevalence of the disease is the same for both genders. In such a case, when data are pooled regardless of the gender of the patient, the obtained ROC is attenuated with respect to the ROC curve in each of the populations determined by covariate $\boldsymbol{X}$. However, if covariate $\boldsymbol{X}$ is associated with disease status, the pooled ROC curve will also 'incorporate' the portion of discrimination attributable to the covariate. This situation can lead to a pooled (or marginal) ROC curve that lies above or below the conditional ROC curve (see Figures 1(c) and 1(d)). It should be noted that, despite the fact that in the previous examples the discriminatory capacity of the diagnostic test is the same for both populations defined by covariate $\boldsymbol{X}$, the threshold that gives rise to a pair of values for the FPF (false positive fraction) and the TPF (true positive fraction) could not coincide in each population. This is also illustrated in Figure 1. The red lines and dots represent a common threshold used to define test positivity. As can be observed, this threshold provides a different pair of FPF and TPF on $\boldsymbol{X}=1$ and $\boldsymbol{X}=0$, as well as on the pooled data. On the other hand, the green lines and dots depict the threshold to be used to ensure a FPF$=0.2$ in both populations. Accordingly, studying the effect of covariates on the distribution of a diagnostic test in the healthy/diseased population will enable assessment of which factors affect the FPF/TPF when a specific threshold value is set. Conversely, different threshold values can be chosen for each of the populations determined by the covariates, in order to ensure that the FPF/TPF remains constant across all of them.



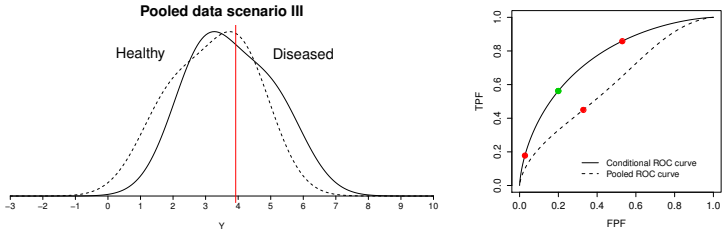
(a)



(b) Scenario I



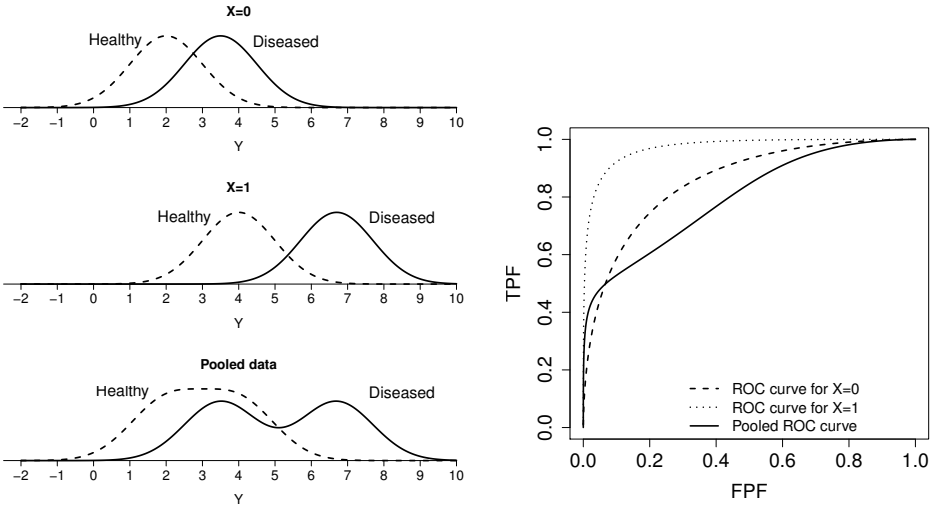
(c) Scenario II



(d) Scenario III

Figure 1: (a) Probability distributions of a hypothetical diagnostic test Y in diseased (solid line) and healthy (dashed line) populations conditional on a binary covariate $\boldsymbol{X} = 0, 1$.
Shown in (b), (c) and (d) are the pooled probability distributions (left panel), and the corresponding pooled ROC curves, along with the common conditional ROC curves (right panel).
Scenario I: disease status and covariate are independent, $P(\text{status } D \mid \boldsymbol{X} = 0) = 0.5$ and $P(\text{status } D \mid \boldsymbol{X} = 1) = 0.5$.
Scenario II: $P(\text{status } D \mid \boldsymbol{X} = 0) = 0.2$ and $P(\text{status } D \mid \boldsymbol{X} = 1) = 0.8$.
Scenario III: $P(\text{status } D \mid \boldsymbol{X} = 0) = 0.6$ and $P(\text{status } D \mid \boldsymbol{X} = 1) = 0.4$.
In all cases $P(\text{status } D) = 0.5$ and $P(\boldsymbol{X} = 1) = 0.5$ were considered. The performance of the common threshold 3.9 is also indicated (red lines and dots), as well as the common conditional threshold that gives rise to a FPF $= 0.2$ in both the populations determined by covariate X (green lines and dots).

On the other hand, in those situations where the accuracy of a diagnostic test is affected by covariates, failure to incorporate information furnished by them may lead, as in the previous cases, to erroneous conclusions. For instance, let us consider the example shown in Figure 2, where the accuracy of a diagnostic test changes according to a binary covariate $\boldsymbol{X}$ (with $\boldsymbol{X}$ again assumed to be patient gender). The conditional ROC curve shows that test Y is more accurate when $\boldsymbol{X} = 1$ than when $\boldsymbol{X} = 0$, though discriminatory capacity is high in both cases. Nevertheless, pooling the data regardless of the values of the covariate yields a ROC curve that is below the specific ROC curves for each of the populations determined by covariate $\boldsymbol{X}$. Taking into account the possible modifying effect of covariates on the accuracy of a diagnostic test, *i.e.*, on the ROC curve, will help identifying the optimal populations to whom or conditions under which the test should be applied, or alternatively, those where the test is unlikely to be useful. Furthermore, different thresholds for defining test positivity can be chosen to vary with covariate values.



(a) (b)

Figure 2: (a) Probability distributions of Y in diseased (solid line) and healthy (dashed line) populations conditional on $\boldsymbol{X}$ and pooled probability distributions.
(b) Conditional ROC curve in each of the populations determined by covariate $\boldsymbol{X}$, together with the pooled ROC curve.
The shown results were obtained assuming that the performance and discriminatory capacity of the diagnostic test depend on $\boldsymbol{X}$, but $\boldsymbol{X}$ is independent of true disease status: $P(\text{status } D \mid \boldsymbol{X} = 1) = P(\text{status } D \mid \boldsymbol{X} = 0) = 0.5$. Moreover, $P(\text{status } D) = 0.5$ and $P(\boldsymbol{X} = 1) = 0.5$ were considered.

Summarising, both in situations where the result of a diagnostic test, though not necessarily its discriminatory capacity, is affected by covariates, and in those where the discriminatory capacity itself is affected by covariates, this information must be incorporated into the ROC analysis. Failure to do so, by

pooling the data regardless of the values of the covariates and using a classification rule that relies on a common threshold value, will result in the test having a discriminatory capacity that is biased compared to its 'true potential' discriminatory capacity. Accordingly, optimistic or pessimistic results may be obtained and, by extension, erroneous conclusions with respect to the real discriminatory capacity of the diagnostic test, which in turn entails an 'incorrect' choice of threshold values to be used in practice.

The previous explanations motivate two possibilities when estimating ROC curves under the presence of covariates. If the discriminatory capacity of the diagnostic test is affected by covariates, then conditional or covariate-specific ROC curves must be considered. When the test diagnostic varies with the covariates, but its discriminatory capacity is not affected by them, then the covariate-adjusted ROC curve, introduced by Janes and Pepe (2009a), is recommended. Both concepts will be defined in the next Section.

3. NOTATION AND DEFINITIONS

Let us assume that along with the continuous diagnostic variables in the diseased population, Y_D, and in the healthy population, $Y_{\bar{D}}$, covariate vectors $\boldsymbol{X}_D$ and $\boldsymbol{X}_{\bar{D}}$ are also available. For the sake of clarity, in this paper we will further assume that the covariates of interest are the same in both healthy and diseased populations. It should be noted, however, that this is not always so. In some circumstances, it could be of interest to evaluate the discriminatory capacity of a diagnostic test with respect to population-specific covariates, as for instance disease stage.

As a natural extension of the ROC curve for continuous diagnostic tests, the conditional or covariate-specific ROC curve, given a covariate value $\boldsymbol{x}$, is defined as

$$\text{(3.1)} \qquad \text{ROC}_{\boldsymbol{x}}(p) = 1 - F_D\big(F_{\bar{D}}^{-1}(1-p \mid \boldsymbol{x}) \mid \boldsymbol{x}\big)\ , \qquad 0 \le p \le 1\ ,$$

where

$$F_D(y \mid \boldsymbol{x}) = P\big(Y_D \le y \mid \boldsymbol{X}_D = \boldsymbol{x}\big)\,,$$
$$F_{\bar{D}}(y \mid \boldsymbol{x}) = P\big(Y_{\bar{D}} \le y \mid \boldsymbol{X}_{\bar{D}} = \boldsymbol{x}\big)\,.$$

Note that in this case, a number of possible different ROC curves can be obtained for each value $\boldsymbol{x}$ in the range of the common part of the supports of $\boldsymbol{X}_D$ and $\boldsymbol{X}_{\bar{D}}$. Associated with the conditional ROC curve, some other measures of discriminatory performance can also be defined. The most widely used one is the area under the ROC curve (AUC), which in the conditional case is defined as $\text{AUC}_{\boldsymbol{x}} = \int_0^1 \text{ROC}_{\boldsymbol{x}}(p)\,\mathrm{d}p$. As for the unconditional case, the $\text{AUC}_{\boldsymbol{x}}$ takes values between 0.5 (for an uninformative test) and 1 (for a perfect test).

Both, the conditional ROC curve and the conditional AUC defined above, depict the discriminatory capacity of a test but for specific values of the covariate vector. It would nevertheless be of great interest to have global discriminatory measures that also take into account covariate information. In this context, the so-called covariate-adjusted ROC curve is defined as an average of conditional ROC curves weighted according to the distribution of the covariate in the diseased population, that is

$$\text{AROC}(p) = \int \text{ROC}_{\boldsymbol{x}}(p)\, \mathrm{d}H_D(\boldsymbol{x}) \ , \tag{3.2}$$

were $H_D(\boldsymbol{x}) = P(\boldsymbol{X}_D \leq \boldsymbol{x})$ is the multivariate distribution function of the vector $\boldsymbol{X}_D$. Despite of the intuitive definition given in the expression above, the covariate-adjusted ROC curve admits other equivalent representations. For instance, in Janes and Pepe (2009a) it is also expressed as

$$\text{AROC}(p) = P\Big(Y_D > F^{-1}_{\bar{D}}(1-p \,|\, \boldsymbol{X}_D)\Big) , \tag{3.3}$$

which means that the covariate-adjusted ROC curve for a FPF $= p$ can be seen as the overall TPF when the threshold used to define test positivity is covariate-specific. This latter expression will be useful when it comes to construct estimators for $\text{AROC}(p)$. Note that based on (3.2), in those situations where the accuracy of a diagnostic test is not affected by covariates, the covariate-adjusted ROC curve coincides with the covariate-specific ROC curve which is common for all covariate values.

4. ESTIMATION PROCEDURES

In order to introduce the estimators, let us assume that we have two independent samples of i.i.d. observations $(\boldsymbol{X}_{\bar{D}1}, Y_{\bar{D}1}), ..., (\boldsymbol{X}_{\bar{D}n_{\bar{D}}}, Y_{\bar{D}n_{\bar{D}}})$ from population $(\boldsymbol{X}_{\bar{D}}, Y_{\bar{D}})$ and $(\boldsymbol{X}_{D1}, Y_{D1})$, ..., $(\boldsymbol{X}_{Dn_D}, Y_{Dn_D})$ from population $(\boldsymbol{X}_D, Y_D)$. Some of the estimators that will be presented below only apply for one-dimensional covariates. However, by a slight abuse of notation, even in those cases we will keep the bold typography to denote the covariates.

4.1. Estimation of the conditional ROC curve

Several proposals for estimating the conditional ROC curve have been given in the statistical literature. Estimators can immediately be obtained by estimating the conditional distribution functions involved in the definition given in (3.1). Besides, other approaches within the general regression framework have been studied, namely the so-called induced and direct ROC-regression methodologies

(see, *e.g.*, Pepe, 1998, 2003; Rodríguez-Álvarez *et al.*, 2011). In this section, we will first present the general ideas behind both approaches, and then focus our attention on nonparametric estimation techniques.

Estimators based on conditional distribution functions. An obvious estimator of the conditional ROC curve follows directly from its definition. Given a covariate value, $\boldsymbol{x}$, the estimator can be constructed as

$$\widehat{\mathrm{ROC}}_{\boldsymbol{x}}(p) \,=\, 1 - \hat{F}_D\big(\hat{F}_{\bar{D}}^{-1}(1-p \mid \boldsymbol{x}) \mid \boldsymbol{x}\big)\,, \tag{4.1}$$

where $\hat{F}_D(\cdot \mid \boldsymbol{x})$ and $\hat{F}_{\bar{D}}(\cdot \mid \boldsymbol{x})$ are estimators of the conditional distributions $F_D(\cdot \mid \boldsymbol{x})$ and $F_{\bar{D}}(\cdot \mid \boldsymbol{x})$, respectively. When we restrict our attention to one-dimensional covariates, the conditional distributions can be estimated nonparametrically, for instance, by kernel-based estimators given in Stone (1977):

$$\hat{F}_{j,h_j}(y \mid \boldsymbol{x}) \,=\, \frac{\sum_{i=1}^{n_j} k\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_{ji}}{h_j}\Big)\, I(Y_{ji} \leq y)}{\sum_{i=1}^{n_j} k\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_{ji}}{h_j}\Big)}\,,$$

with $j \in \{\bar{D}, D\}$, where $I(\cdot)$ denotes the indicator function and where k is the kernel (usually a symmetric density) and h_D and $h_{\bar{D}}$ are the smoothing parameters. Under this approach, the estimator of the conditional ROC curve at a specific covariate value uses the information corresponding to individuals whose covariate values are close to $\boldsymbol{x}$.

The estimator given in (4.1) is of an empirical type, and therefore has discontinuities. In López-de Ullibarri *et al.* (2008) a nonparametric smooth estimator of the conditional ROC curve is obtained by applying the methodology that Peng and Zhou (2004) proposed in the unconditional case. The key idea of this method consists of smoothing the empirical ROC curve by means of kernel techniques. In the conditional case, the smoothed version of (4.1) given in López-de Ullibarri *et al.* (2008) is

$$\widehat{\mathrm{ROC}}_{\boldsymbol{x},h}(p) \,=\, 1 - \int \hat{F}_{D,h_D}\big(\hat{F}_{\bar{D},h_{\bar{D}}}^{-1}(1-p+hu \mid \boldsymbol{x}) \mid \boldsymbol{x}\big)\, k(u)\, \mathrm{d}u\,, \tag{4.2}$$

where the parameter h controls the amount of smoothing and k is a kernel function. The authors propose a bootstrap method to choose the smoothing parameters involved in (4.1) and (4.2).

Very recently, Inácio de Carvalho *et al.* (2013) presented a nonparametric Bayesian model to estimate the conditional distribution functions involved in (3.1). The main advantage of their approach, in contrast to the proposal of López-de Ullibarri *et al.* (2008), is the possibility of studying the effect of multidimensional covariates. Specifically, covariate-dependent Dirichlet processes (DDP) (MacEachern, 2000) defined in terms of i.i.d. Gaussian processes are proposed to

estimate $F_D(\cdot \mid \boldsymbol{x})$ and $F_{\bar{D}}(\cdot \mid \boldsymbol{x})$. Moreover, the computational burden associated with the proposal is overcome by approximating the Gaussian processes by B-splines basis functions, yielding the so-called B-splines DDP mixture model. The authors show by means of simulation the better performance of the proposed model in complex scenarios when compared to other nonparametric estimators of the conditional ROC curve (González-Manteiga *et al.*, 2011; Rodríguez-Álvarez *et al.*, 2011a).

Estimators based on induced-regression methodology. An alternative way to incorporate information from covariates to the ROC analysis is through regression models. The induced methodology in ROC analysis consists of modelling the effect of the covariates through regression models linking the classification variable and the covariates in each population separately. The regression models will then be used to compose the conditional ROC curve. In a general framework, the relationship between the covariate and the classification variable in each population is given by location-scale regression models

$$Y_{\bar{D}} = \mu_{\bar{D}}(\boldsymbol{X}_{\bar{D}}) + \sigma_{\bar{D}}(\boldsymbol{X}_{\bar{D}})\,\varepsilon_{\bar{D}}\ , \tag{4.3}$$

$$Y_D = \mu_D(\boldsymbol{X}_D) + \sigma_D(\boldsymbol{X}_D)\,\varepsilon_D\ , \tag{4.4}$$

where, for $j \in \{\bar{D}, D\}$, $\mu_j(\boldsymbol{x}) = E(Y_j \mid \boldsymbol{X}_j = \boldsymbol{x})$ and $\sigma_j^2(\boldsymbol{x}) = \mathrm{var}(Y_j \mid \boldsymbol{X}_j = \boldsymbol{x})$ are the conditional mean and the conditional variance of Y_j given $\boldsymbol{X}_j = \boldsymbol{x}$, respectively, and the error ε_j is independent of the covariate $\boldsymbol{X}_j$. The independence between the error and the covariate in the location-scale regression model allows us to rewrite the conditional distribution function of the classification variable in terms of the distribution of the regression error as follows:

$$\begin{aligned} F_j(y \mid \boldsymbol{x}) &= P\Big(Y_j \le y \mid \boldsymbol{X}_j = \boldsymbol{x}\Big) \\ &= P\Big(\mu_j(\boldsymbol{X}_j) + \sigma_j(\boldsymbol{X}_j)\,\varepsilon_j \le y \mid \boldsymbol{X}_j = \boldsymbol{x}\Big) \\ &= P\left(\varepsilon_j \le \frac{y - \mu_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})}\right) = G_j\left(\frac{y - \mu_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})}\right), \end{aligned}$$

where, for $j \in \{\bar{D}, D\}$, $G_j(y) = P(\varepsilon_j \le y)$ is the distribution function of the regression error. An analogous relationship can be established between the conditional quantile function of Y_j given $\boldsymbol{X}_j = \boldsymbol{x}$, $F_j^{-1}(\cdot \mid \boldsymbol{x})$, and the quantile function of ε_j, $G_j^{-1}(\cdot)$, through the expression $F_j^{-1}(p \mid \boldsymbol{x}) = \mu_j(\boldsymbol{x}) + \sigma_j(\boldsymbol{x})\,G_j^{-1}(p)$. Therefore, for a fixed covariate value $\boldsymbol{x}$, and for $0 < p < 1$, the conditional ROC curve can be expressed as

$$\begin{aligned} \mathrm{ROC}_{\boldsymbol{x}}(p) &= 1 - F_D\Big(F_{\bar{D}}^{-1}(1-p \mid \boldsymbol{x}) \mid \boldsymbol{x}\Big) \\ &= 1 - F_D\Big(\mu_{\bar{D}}(\boldsymbol{x}) + \sigma_{\bar{D}}(\boldsymbol{x})\,G_{\bar{D}}^{-1}(1-p) \mid \boldsymbol{x}\Big) \\ &= 1 - G_D\left(\frac{\mu_{\bar{D}}(\boldsymbol{x}) + \sigma_{\bar{D}}(\boldsymbol{x})\,G_{\bar{D}}^{-1}(1-p) - \mu_D(\boldsymbol{x})}{\sigma_D(\boldsymbol{x})}\right) \\ &= 1 - G_D\Big(G_{\bar{D}}^{-1}(1-p)\,b(\boldsymbol{x}) - a(\boldsymbol{x})\Big)\,, \end{aligned} \tag{4.5}$$

where $a(\boldsymbol{x}) = \big(\mu_D(\boldsymbol{x}) - \mu_{\bar{D}}(\boldsymbol{x})\big)/\sigma_D(\boldsymbol{x})$ and $b(\boldsymbol{x}) = \sigma_{\bar{D}}(\boldsymbol{x})/\sigma_D(\boldsymbol{x})$. This formulation allows us to express the conditional ROC curve in terms of the distribution function and quantile function of the regression errors, which are not conditional. Hence, from an estimation point of view, instead of estimating the conditional distribution of Y_D and $Y_{\bar{D}}$ given $\boldsymbol{x}$, one only needs to estimate the error distribution in each population. This is a main advantage with respect to the estimator given in (4.2).

The induced ROC methodology described above has been presented for the most general case. In fact, only particular cases have been addressed in the literature. In a parametric or semiparametric framework, Faraggi (2003) assumes an additive parametric model for the conditional means, with homoscedastic variances and normal errors, in both healthy and diseased populations. Pepe (1998) relaxes the distributional assumptions by not assuming a known probability distribution for the error terms, although the same distribution is considered for both populations. Zhou *et al.* (2002) extend the model in Pepe (1998) by allowing for heteroscedasticity. Finally, Zheng and Heagerty (2004) propose a semiparametric estimator for the conditional ROC curve, in which the distribution of the error terms is unknown and allowed to depend on the covariates, but, as in the previous articles, the effect of the covariates on the conditional means and variances is modelled parametrically. Very recently, Rodríguez and Martínez (2014) presented a Bayesian semiparametric model, where the error terms are assumed to be normally distributed, but nonparametric specifications of the conditional means and variances are allowed.

A different line of research has led to estimation in a fully nonparametric framework, although so far only one-dimensional covariates have been considered. We focus now on those approaches, introduced by González-Manteiga *et al.* (2011) and Rodríguez-Álvarez *et al.* (2011a). When models (4.3) and (4.4) are nonparametric, the estimator of the conditional ROC curve involves the following steps. First, for $j \in \{\bar{D}, D\}$, we need to estimate nonparametrically the location and scale functions in the regression models, say $\hat{\mu}_j(\boldsymbol{x})$ and $\hat{\sigma}_j(\boldsymbol{x})$ by means, for example, of Nadaraya–Watson or local-linear estimators (see, for example, Fan and Gijbels, 1996). Then the distribution of the errors in the two regression models are estimated by the corresponding empirical distribution function of the estimated residuals, *i.e.*, $\hat{G}_j(y) = n_j^{-1} \sum_{i=1}^{n_j} I(\hat{\varepsilon}_{ji} \leq y)$, where, for $j \in \{\bar{D}, D\}$, $\hat{\varepsilon}_{ji} = \big(Y_{ji} - \hat{\mu}_j(\boldsymbol{X}_{ji})\big)/\hat{\sigma}_j(\boldsymbol{X}_{ji})$, $i = 1, ..., n_j$. Finally, given the covariate value $\boldsymbol{x}$, an empirical estimator of the conditional ROC curve is

$$\widehat{\mathrm{ROC}}_{\boldsymbol{x}}(p) \;=\; 1 - \hat{G}_D\Big(\hat{G}_{\bar{D}}^{-1}(1-p)\,\hat{b}(\boldsymbol{x}) - \hat{a}(\boldsymbol{x})\Big)\,, \tag{4.6}$$

where $\hat{a}(\boldsymbol{x}) = \big(\hat{\mu}_D(\boldsymbol{x}) - \hat{\mu}_{\bar{D}}(\boldsymbol{x})\big)/\hat{\sigma}_D(\boldsymbol{x})$ and $\hat{b}(\boldsymbol{x}) = \hat{\sigma}_{\bar{D}}(\boldsymbol{x})/\hat{\sigma}_D(\boldsymbol{x})$. As in the case of (4.1), the previous estimator of the conditional ROC curve is not continuous. In order to obtain a smooth version, González-Manteiga *et al.* (2011) also apply

the methodology in Peng and Zhou (2004), which yields

$$\widehat{\mathrm{ROC}}_{\boldsymbol{x},h}(p) \;=\; 1 - \int \hat{G}_D\Big(\hat{G}_{\bar{D}}^{-1}(1-p+hu)\,\hat{b}(\boldsymbol{x}) - \hat{a}(\boldsymbol{x})\Big)\,k(u)\,\mathrm{d}u\;. \tag{4.7}$$

The authors show that the former estimator also admits the following explicit expression:

$$\widehat{\mathrm{ROC}}_{\boldsymbol{x},h}(p) \;=\; \frac{1}{n_D}\sum_{i=1}^{n_D} K\left(\frac{\hat{G}_{\bar{D}}\Big(\big\{\hat{\varepsilon}_{Di} + \hat{a}(\boldsymbol{x})\big\}/\hat{b}(\boldsymbol{x})\Big) - 1 + p}{h}\right),$$

where K is the distribution function corresponding to the density kernel k.

A detailed study of the asymptotic properties of the estimators given in (4.6) and (4.7) is provided in González-Manteiga *et al.* (2011). In Rodríguez-Álvarez *et al.* (2011a), a bootstrap-based test to check for the effect of the covariate over the conditional ROC curve is proposed. Although both papers focus on the estimation of the conditional ROC curve, an estimator of the conditional AUC is also presented, $\widehat{\mathrm{AUC}}_{\boldsymbol{x}} = \int_0^1 \widehat{\mathrm{ROC}}_{\boldsymbol{x}}(p)\,\mathrm{d}p$, with the integral being approximated by numerical integration methods. In that sense, the paper by Yao *et al.* (2010) goes one step further in proposing a nonparametric estimator for $\mathrm{AUC}_{\boldsymbol{x}}$ based also on induced modelling and local linear kernel smoothers. The authors exploit the relation between the Mann–Whitney statistic and the empirical estimator of the unconditional AUC (see, *e.g.*, Bamber, 1975) and propose a covariate-specific Mann–Whitney estimator for $\mathrm{AUC}_{\boldsymbol{x}}$.

Estimators based on direct-regression methodology. In contrast to the induced methodology, in the direct methodology the effect of the covariates is directly evaluated on the ROC curve. To motivate the standard formulation of direct methodology, let us re-express the conditional ROC curve as follows:

$$\begin{aligned}
\mathrm{ROC}_{\boldsymbol{x}}(p) \;&=\; 1 - F_D\Big(F_{\bar{D}}^{-1}(1-p \mid \boldsymbol{x}) \mid \boldsymbol{x}\Big) \\
&=\; 1 - P\Big(Y_D \le F_{\bar{D}}^{-1}(1-p \mid \boldsymbol{x}) \mid \boldsymbol{X}_D = \boldsymbol{x}\Big) \\
&=\; 1 - P\Big(F_{\bar{D}}(Y_D \mid \boldsymbol{x}) \le 1-p \mid \boldsymbol{X}_D = \boldsymbol{x}\Big) \\
&=\; P\Big(1 - F_{\bar{D}}(Y_D \mid \boldsymbol{x}) < p \mid \boldsymbol{X}_D = \boldsymbol{x}\Big) \qquad (4.8) \\
&=\; E\Big[I\Big(1 - F_{\bar{D}}(Y_D \mid \boldsymbol{x}) < p\Big) \mid \boldsymbol{X}_D = \boldsymbol{x}\Big]\,. \qquad (4.9)
\end{aligned}$$

As can be observed, the conditional ROC curve may be seen as: (a) the conditional distribution function of the random variable $1 - F_{\bar{D}}(Y_D \mid \boldsymbol{x})$ in expression (4.8), or (b) the conditional expected value of the binary variable $I\big(1 - F_{\bar{D}}(Y_D \mid \boldsymbol{x}) < p\big)$ in expression (4.9). The random variable $1 - F_{\bar{D}}(Y_D \mid \boldsymbol{x})$ is called 'placement value' in related literature (see, for example, Hanley and Hajian-Tilaki, 1997) and represents the standardization of the classification variable in the diseased population to the conditional distribution of the non-diseased population.

These two interpretations justify to express the conditional ROC curve as a sort of regression model of the form

$$\text{ROC}_{\boldsymbol{x}}(p) = g\big(\mu(\boldsymbol{x}), \gamma(p)\big)\,, \tag{4.10}$$

where g is a bivariate function on $[0,1]$ and γ is a function defined on the interval $[0,1]$. The function μ collects the effect of the covariates on the conditional ROC curve, and γ is a baseline function related to the shape of the ROC curve. In order to obtain a valid model of ROC curves, some restrictions need to be imposed on the elements of model (4.10). In particular, the function g needs to be monotone increasing in p, with $g(\mu(\boldsymbol{x}), \gamma(0)) = 0$ and $g(\mu(\boldsymbol{x}), \gamma(1)) = 1$ for all $\boldsymbol{x}$. As in the case of the induced methodology presented above, model (4.10) represents the most general formulation of the direct methodology. In fact, only the additive specification

$$\text{ROC}_{\boldsymbol{x}}(p) = g\big(\mu(\boldsymbol{x}) + \gamma(p)\big) \tag{4.11}$$

has been addressed in the statistical literature. Different proposals have been suggested, which differ in the assumptions made about the functions g, μ and γ. In Pepe (1997, 2000) and Alonzo and Pepe (2002), g is assumed to be known, the effect of the covariates on the conditional ROC curve is assumed to be linear, *i.e.*, $\mu(\boldsymbol{x}) = \boldsymbol{\beta}^T\boldsymbol{x}$, and the baseline function γ is assumed to have a parametric form. Cai and Pepe (2002) and Cai (2004) leave γ completely unspecified, but the function μ is linear as well. In general, models such as (4.11) with parametric specifications for μ define the so-called class of ROC-GLMs due to its similarity with a generalized linear model (GLM, McCullagh and Nelder, 1989) in regression (Pepe, 2003). In all the aforementioned papers, the function g is assumed to be known. Huazhen *et al.* (2012) relax this assumption, by allowing a completely unknown function g. As for the approaches in Cai and Pepe (2002) and Cai (2004), the function γ remains unspecified and μ is assumed to have a parametric form. In a completely nonparametric framework, Rodríguez-Álvarez *et al.* (2011b) extend the class of ROC-GLM regression models, by assuming a generalized additive model (GAM, Hastie and Tibshirani, 1990) for the ROC curve, that is

$$\mu(\boldsymbol{x}) = \mu(x_1, ..., x_d) = \alpha + \sum_{k=1}^{d} f_k(x_k)\,,$$

where $f_1, ..., f_d$ are unknown nonparametric functions, and γ also remains unspecified.

Either if the specifications in (4.11) involve a GLM structure (as in Alonzo and Pepe, 2002) or a GAM structure (as in Rodríguez-Álvarez *et al.*, 2011b), the estimation process is similar and can be described as given in the following steps. First, choose a set of FPFs $0 \leq p_l \leq 1$, $l = 1, ..., n_P$, where the conditional ROC curves will be evaluated. Second, estimate $F_{\bar{D}}(\cdot \,|\, \boldsymbol{x})$, say $\hat{F}_{\bar{D}}(\cdot \,|\, \boldsymbol{x})$, on the basis of the sample $(\boldsymbol{X}_{\bar{D}i}, Y_{\bar{D}i})$, $i = 1, ..., n_{\bar{D}}$. Third, for each observation in the diseased population, calculate the estimated placement value $1 - \hat{F}_{\bar{D}}(Y_{Di} \,|\, \boldsymbol{x})$, $1 \leq i \leq n_D$.

Fourth, calculate the binary indicators $I\big(1-\hat{F}_{\bar{D}}(Y_{Di}\,|\,\boldsymbol{x})\leq p_l\big)$, for $1\leq i\leq n_D$ and $1\leq l\leq n_P$. And finally, fifth, fit the model $g\big(\mu(\boldsymbol{x})+\gamma(p)\big)$ as a regression model with the indicators $I\big(1-\hat{F}_{\bar{D}}(Y_{Di}\,|\,\boldsymbol{x})\leq p_l\big)$ as response and covariates $\boldsymbol{X}_{Di}$ and p_l, $i=1,...,n_D$, $l=1...,n_P$.

Depending on the chosen specifications for μ and γ, GLM or GAM techniques will be employed for fitting the model (4.11). For instance, in Rodríguez-Álvarez *et al.* (2011b) the proposed estimation procedure is based on a combination of local scoring and backfitting algorithms (Hastie and Tibshirani, 1990), and the nonparametric functions $f_1,...,f_d$ and γ are estimated using local linear kernel smoothers (see Fan and Gijbels, 1996). Note that in contrast to the nonparametric approaches based on induced modelling presented above, this proposal allows for the possibility of incorporating multidimensional covariates. However, the study of the theoretical properties of the estimator is so far lacking in the literature.

Throughout the above outline of induced and direct modelling, the covariates (whose effect on the ROC curve we seek to evaluate) were assumed to be common to both the healthy and the diseased population. As mentioned before, in practice this is not necessarily so. For instance, it may be of interest to evaluate the performance of the diagnostic variable with respect to disease stage. Induced methodology poses no problem when it comes to incorporating specific covariates of healthy or diseased populations, or both. On the other hand, direct methodology—as presented here—accepts no specific covariates of the healthy population. Yet, even in cases where this may seem a restriction, the need arises in few situations in practice.

4.2. Estimation of the covariate-adjusted ROC curve

As explained in the introduction, in some practical cases, although the diagnostic test varies along with the covariates, its discriminatory capacity may remain unalterable. In such a situation, instead of considering the conditional ROC curve, the covariate-adjusted ROC curve is more convenient. The definition given in (3.3)

$$\text{AROC}(p) \;=\; P\Big(Y_D > F_{\bar{D}}^{-1}\big(1-p\mid \boldsymbol{X}_D\big)\Big)$$

suggests estimating the covariate-adjusted ROC curve as sample proportion of individuals in the diseased population that exceed a certain covariate-specific threshold calculated with the conditional quantile function in the healthy population. Note that the conditional quantile function is an unknown function and therefore needs to be estimated. Janes and Pepe (2009a) propose estimators of the form

$$\widehat{\text{AROC}}(p) \;=\; \frac{1}{n_D}\sum_{i=1}^{n_D} I\Big(Y_{Di} > \hat{F}_{\bar{D}}^{-1}\big(1-p\mid \boldsymbol{X}_{Di}\big)\Big)\,,$$

where $\hat{F}_{\bar{D}}^{-1}(1-p \mid \boldsymbol{X}_{Di})$ can be estimated semiparametrically or nonparametrically. In the context of the induced methodology described in Subsection 4.1, Rodríguez-Álvarez *et al.* (2011a) used the relation between the conditional quantile and the quantile of the regression errors to obtain the following nonparametric estimator:

$$\widehat{\mathrm{AROC}}(p) = \frac{1}{n_D} \sum_{i=1}^{n_D} I\left(\frac{Y_{Di} - \hat{\mu}_{\bar{D}}(\boldsymbol{X}_{Di})}{\hat{\sigma}_{\bar{D}}(\boldsymbol{X}_{Di})} > \hat{G}_{\bar{D}}^{-1}(1-p) \right),$$

where $\hat{\mu}_{\bar{D}}$ and $\hat{\sigma}_{\bar{D}}$ are nonparametric estimators of $\mu_{\bar{D}}$ and $\sigma_{\bar{D}}$ in model (4.3), and $\hat{G}_{\bar{D}}^{-1}$ is the empirical quantile function of the estimated residuals. The theoretical properties of this estimator have not been studied yet.

5. ILLUSTRATION WITH REAL DATA

In this section, a real data illustration of the importance of including covariates into the ROC framework is presented. The data set comes from a cross-sectional study carried out by the Galician Endocrinology and Nutrition Foundation (FENGA), consisting of 2860 individuals representative of the adult population of Galicia (northwest of Spain). A detailed description of this data set can be found in Tomé *et al.* (2008). For confidentiality reasons, only a subsample of the global sample was used in this paper, where we aimed at assessing the performance of the body mass index (BMI) for predicting clusters of cardiovascular disease (CVD) risk factors. Accordingly, diseased subjects were defined as those having two or more CVD risk factors (raised triglycerides, reduced high-density lipoprotein cholesterol, raised blood pressure and raised fasting plasma glucose), following the International Diabetes Federation criteria (International Diabetes Federation, 2006). For the study here presented, a total of 1419 individuals were selected from the original data set, with an age range between 18 and 85 years. From those, 46.4% are men (449 healthy and 209 diseased) and the remaining 53.6% are women (625 healthy and 136 diseased). An in-depth study of the global data set is presented in Rodríguez-Álvarez *et al.* (2011a,b).

It is well known that anthropometric measures behave differently according to both age and gender. This can be observed in Table 1, where some summary statistics of the BMI for men and women, as well as for different age strata, are presented. As illustrated in Section 2, it is therefore advisable to incorporate both covariates into the ROC analysis. In this paper, we applied the nonparametric induced approach proposed by González-Manteiga *et al.* (2011) and Rodríguez-Álvarez *et al.* (2011a) and presented in Section 4.1. Since this proposal only admits one continuous covariate, separate analyses were conducted on men and women respectively.

Table 1: Median and interquartile range of the BMI for the global sample, for men and women, and for different age strata.

	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**
Global sample	22.84	25.91	29.34
Gender			
Female	22.00	24.69	25.91
Male	24.16	26.88	27.14
Age strata			
<30 years	21.28	22.85	25.83
30–39 years	22.66	25.40	28.08
40–49 years	24.18	26.77	29.74
50–59 years	25.84	28.65	31.46
≥ 60 years	26.62	29.38	31.72

In addition to the estimated conditional ROC curves, other summary measures of accuracy, the conditional AUC and the age-adjusted ROC curve, were also obtained. In Figure 3, the estimated age-adjusted ROC curve for both men and women is shown, jointly with the estimated pooled ROC curve. As can be observed, in both cases the age-adjusted ROC curve lies below the pooled ROC curve, especially for men. It is worth remembering that the covariate-adjusted ROC curve is an average of conditional ROC curves, and can therefore be interpreted as a covariate-adjusted global discriminatory measure. Thus, for the endocrinology data, pooling the data regardless of age and gender would lead to an optimistic conclusion about the discriminatory capacity of the BMI when predicting the presence of CVD risk factors.

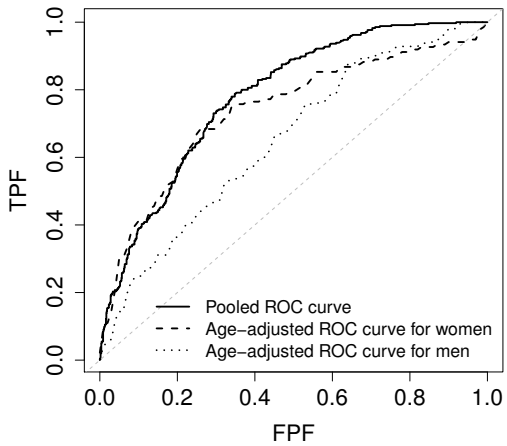


Figure 3: Estimated pooled ROC curve for the endocrinology data (solid line). The dashed and dotted lines represent the estimated age-adjusted ROC curve for women and men, respectively.

In Figure 4 the estimated conditional ROC curve and AUC for different age values are depicted, for both men and women. Note that whereas for men the accuracy of the BMI is more or less constant along age, for women, age displays a relevant effect on the discriminatory capacity of this anthropometric measure. This graphical conclusion was confirmed by applying the bootstrap-based test presented in Rodríguez-Álvarez *et al.* (2011a). The test enabled a significant age effect to be detected in the case of women. In the case of men, however, there was no evidence to suggest such an effect.

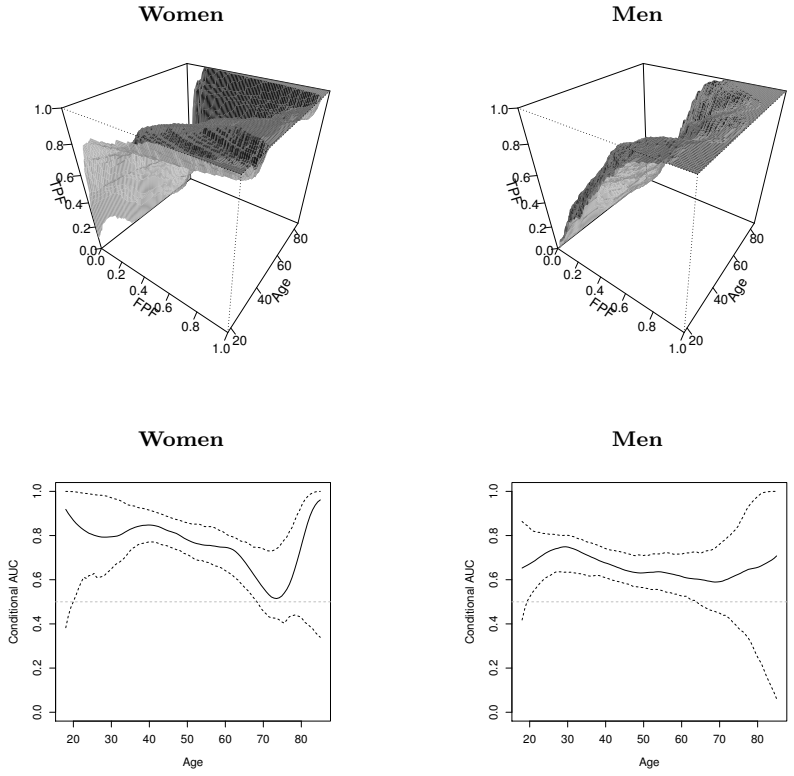


Figure 4: Estimated conditional ROC curves and AUCs for the endocrinology data for women and men. The dashed lines represent the 95 per cent pointwise bootstrap confidence interval.

The results presented in this section emphasize once again the importance and consequences of including the information provided by covariates when evaluating the discriminatory capacity of a diagnostic test. In the case of women, the conditional ROC curve should be reported since it has been proved that age has an effect on the accuracy of the BMI. For men, however, no age effect was detected. Nevertheless, even in this case, reporting the discriminatory capacity of the pooled data would lead to an optimistic conclusion, and therefore the age-adjusted ROC curve should be provided.

6. DISCUSSION

In this paper we explained why it is important to incorporate covariates in the ROC analysis and which effect it has on the curve. We also presented two different ways to take covariates into account, either by working with a conditional ROC curve or with a so-called covariate-adjusted ROC curve. Several estimation procedures were outlined for both approaches. Interested readers can find more details in the provided references.

Although we focused in this review on the estimation of the ROC curve in the presence of covariates, it is clear that apart from the ROC curve itself, interest also lies in summary statistics of the ROC curve, like *e.g.* the AUC, the Youden index and other related indices. Within a parametric or semiparametric framework, some attempts about this topic can be found in Faraggi (2003), in which the induced methodology is employed, and in Dodd and Pepe (2003a,b) and Cai and Dodd (2008), all based on the ROC-regression direct modelling approach.

An interesting extension of the ROC methodology is the extension to functional data. We mention the paper by Inácio *et al.* (2012), who consider the extension to functional covariates. To this end, semiparametric and nonparametric induced ROC-regression estimators are proposed and studied. Also, the extension of the ROC methodology from completely observed data to censored data is a promising field of research. For an overview article on this topic we refer to Pepe *et al.* (2008).

Another interesting point to note is that almost no theory has been done for the nonparametric estimators of the conditional and adjusted ROC curve, except in González-Manteiga *et al.* (2011), who obtain the asymptotic normality of nonparametric estimators of both the conditional ROC curve and the conditional AUC based on induced methodology. Their results are limited to a one-dimensional covariate, but they can be easily extended to multi-dimensional covariates by using Neumeyer and Van Keilegom (2010) in the proofs of the asymptotic results.

A number of issues remain unexplored in the context of ROC curves with covariates. For instance, a lot of work remains to be done to extend the concept of relative distributions to the inclusion of covariates (see Handcock and Morris, 1999, for a textbook on this topic). ROC curves are very much related to relative distributions or relative densities (see *e.g.* Li *et al.*, 1996), but their objective is different. In fact, the ROC curve in a point $0 < p < 1$ equals one minus the relative distribution evaluated in $1 - p$. Since the relative density of one population versus another population equals the uniform density in case both populations have the same distribution, it is clear that deviations from the uniform density give an indication of the way in which the two distributions differ from each other.

Hence, relative densities are more used in the context of comparing the distribution of two populations, whereas ROC curves are used for assessing the discriminatory capacity of a diagnostic test. As far as we are aware of, no formal and detailed study of the concept of relative distribution or relative density in the presence of covariates has been developed so far.

ACKNOWLEDGMENTS

The authors would like to thank the Galician Endocrinology and Nutrition Foundation (*Fundación de Endocrinoloxía e Nutrición Galega—FENGA*) for having supplied the database used in this study. J. C. Pardo-Fernández and M. X. Rodríguez-Álvarez express their gratitude for the support received in the form of the Spanish Ministry of Science and Innovation grants MTM2011-23204 (FEDER support included) and MTM2011-28285-C02-01. I. Van Keilegom acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, from IAP research network P7/06 of the Belgian Government (Belgian Science Policy) and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

REFERENCES

Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques, *Biostatistics*, **3**, 421–432.

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

Cai, T. (2004). Semi-parametric ROC regression analysis with placement values, *Biostatistics*, **5**, 45–60.

Cai, T. and Dodd, L. E. (2008). Regression analysis for the partial area under the ROC curve, *Statistica Sinica*, **18**, 817–836.

Cai, T. and Pepe, M. S. (2002). Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease, *Journal of the American Statistical Association*, **97**, 1099–1107.

Dodd, L. E. and Pepe, M. S. (2003a). Semiparametric regression for the area under the receiver operating characteristic curve, *Journal of the American Statistical Association*, **98**, 409–417.

Dodd, L. E. and Pepe, M. S. (2003b). Partial AUC estimation and regression, *Biometrics*, **59**, 614–623.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall/CRC, Boca Raton.

Faraggi, D. (2003). Adjusting receiver operating characteristic curves and related indices for covariates, *The Statistician*, **52**, 179–192.

González-Manteiga, W.; Pardo-Fernández, J. C. and Van Keilegom, I. (2011). ROC curves in nonparametric location-scale regression models, *Scandinavian Journal of Statistics*, **38**, 169–184.

Handcock, M. S. and Morris, M. (1999). *Relative Distribution Methods in the Social Sciences*, Springer-Verlag, New York.

Hanley, J. A. and Hajian-Tilaki, K. O. (1997). Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: An update, *Academic Radiology*, **4**, 49–58.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman & Hall/CRC, Boca Raton.

Huazhen, L.; Xiao-Hua, Z. and Gang, L. (2012). A direct semiparametric receiver operating characteristic curve regression with unknown link and baseline functions, *Statistica Sinica*, **22**, 1427–1456.

Inácio, V.; González-Manteiga, W.; Febrero-Bande, M.; Gude, F.; Alonzo, T. A. and Cadarso-Suárez, C. (2012). Extending induced ROC methodology to the functional context, *Biostatistics*, **13**, 594–608.

Inácio de Carvalho, V.; Jara, A.; Hanson, T. E. and de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling, *Bayesian Analysis*, **3**, 623–646.

International Diabetes Federation (2006). The IDF consensus worldwide definition of the metabolic syndrome. Accessed January 2014. (`http://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf`).

Janes, H. and Pepe, M. S. (2008). Adjusting for covariate in studies of diagnostic, screening, or prognosis markers: an old concept in a new setting, *American Journal of Epidemiology*, **168**, 89–97.

Janes, H. and Pepe, M. S. (2009a). Adjusting for covariate effects on classification accuracy using the covariate-adjusted ROC curve, *Biometrika*, **96**, 371–382.

Janes, H. and Pepe, M. S. (2009b). Accommodating covariates in ROC analysis, *Stata Journal*, **9**, 17–39.

Li, G.; Tiwari, R. C. and Wells, M. T. (1996). Quantile comparison functions in two sample problems with applications to comparisons of diagnostic markers, *Journal of the American Statistical Association*, **91**, 689–698.

López-de Ullibarri, I.; Cao, R.; Cadarso-Suárez, C. and Lado, M. J. (2008). Nonparametric estimation of conditional ROC curves: Application to discrimination tasks in computerized detection of early breast cancer, *Computational Statistics & Data Analysis*, **52**, 2623–2631.

MacEachern, S. N. (2000). *Dependent Dirichlet processes*, Technical report, Department of Statistics, The Ohio State University.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second Edition, Chapman & Hall/CRC, Boca Raton.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101**, 1067–1078.

Peng, L. and Zhou, X.-H. (2004). Local linear smoothing of receiver operating characteristic (ROC) curves, *Journal of Statistical Planning and Inference*, **118**, 129–143.

Pepe, M. S. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing, *Biometrika*, **84**, 595–608.

Pepe, M. S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results, *Biometrics*, **54**, 124–135.

Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures, *Biometrics*, **56**, 352–359.

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

Pepe, M. S; Zheng, Y.; Jin, Y.; Huang, Y.; Parikh, C. R. and Levy, W. C. (2008). Evaluating the ROC performance of markers for future events, *Lifetime Data Analysis*, **14**, 86–113.

Rodríguez, A. and Martínez, J. C. (2014). Bayesian semiparametric estimation of covariate-dependent ROC curves, *Biostatistics*, **15**, 353–369.

Rodríguez-Álvarez, M. X.; Roca-Pardiñas, J. and Cadarso-Suárez, C. (2011a). ROC curve and covariates: extending induced methodology to the non-parametric framework, *Statistics & Computing*, **21**, 483–499.

Rodríguez-Álvarez, M. X.; Roca-Pardiñas, J. and Cadarso-Suárez, C. (2011b). A new flexible direct ROC regression model: Application to the detection of cardiovascular risk factors by anthropometric measures, *Computational Statistics & Data Analysis*, **55**, 3257–3270.

Rodríguez-Álvarez, M. X.; Tahoces, P. G.; Cadarso-Suárez, C. and Lado, M. J. (2011). Comparative study of ROC regression techniques—Applications for the computer-aided diagnostic system in breast cancer detection, *Computational Statistics & Data Analysis*, **55**, 888–902.

Stone, C. J. (1977). Consistent nonparametric regression, *The Annals of Statistics*, **5**, 595–620.

Tomé, M. A; Botana, M. A.; Cadarso-Suárez, C.; Rego-Iraeta, A; Fernández-Mariño, A; Mato, J. A.; Solache, I. and Pérez-Fernández, R. (2008). Prevalence of metabolic syndrome in Galicia (NW Spain) on four alternative definitions and association with insulin resistance, *Journal of Endocrinological Investigation*, **32**, 505–511.

Yao, F.; Craiu, R. V. and Reiser, B. (2010). Nonparametric covariate adjustment for receiver operating characteristic curves, *The Canadian Journal of Statistics*, **38**, 27–46.

Zheng, Y. and Heagerty, P. J. (2004). Semiparametric estimation of time-dependent ROC curves for longitudinal marker data, *Biostatistics* **4**, 615–632.

Zhou, X. H.; Obuchowski, N. A. and McClish, D. K. (2002). *Statistical Methods in Diagnostic Medicine*, Wiley, New York.