THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Functional Covariate-Adjusted Partial Area under the Specificity-ROC Curve with an Application to Metabolic Syndrome Diagnosis

Citation for published version:
Calhau Fernandes Inacio De Carvalho, V, de Carvalho, M, Alonzo, TA & González-Manteiga, W 2016, 'Functional Covariate-Adjusted Partial Area under the Specificity-ROC Curve with an Application to Metabolic Syndrome Diagnosis' Annals of Applied Statistics, vol. 10, no. 3, pp. 1472-1495. DOI: 10.1214/16-AOAS943

Digital Object Identifier (DOI):
10.1214/16-AOAS943

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Annals of Applied Statistics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 05. Apr. 2019

Inácio de Carvalho, V., de Carvalho, M., Alonzo, T. A., and González-Manteiga, W. (2016), "Functional Partial Area under the Curve Regression with an Application to Metabolic Syndrome Diagnosis," Annals of Applied Statistics, 10, 1472-1495.

Submitted to the Annals of Applied Statistics

FUNCTIONAL COVARIATE-ADJUSTED PARTIAL AREA UNDER THE SPECIFICITY-ROC CURVE WITH AN APPLICATION TO METABOLIC SYNDROME DIAGNOSIS

By Vanda Inácio de Carvalho, Miguel de Carvalho, Todd A. Alonzo and Wenceslao González-Manteiga

*Pontificia Universidad Católica de Chile**, *University of Southern California*† *and Universidade de Santiago de Compostela*‡

Due to recent advances in technology, medical diagnosis data are becoming increasingly complex and, nowadays, applications where measurements are curves or images are ubiquitous. Motivated by the need of modeling a functional covariate on a metabolic syndrome case study, we develop a nonparametric functional regression model for the area under the specificity receiver operating characteristic curve. This partial area is a meaningful summary measure of diagnostic accuracy for cases in which misdiagnosis of diseased subjects may lead to serious clinical consequences, and hence, it is critical to maintain a high sensitivity. Its normalized value can be interpreted as the average specificity over the interval of sensitivities considered, thus summarizing the tradeoff between sensitivity and specificity. Our methods are motivated by, and applied to, a metabolic syndrome study that investigates how restricting the sensitivity of the gamma-glutamyl-transferase, a metabolic syndrome marker, to certain clinical meaningful values, affects its corresponding specificity and how it might change for different curves of arterial oxygen saturation. Application of our methods suggests that oxygen saturation is key to gamma-glutamyl transferase's performance and that some of the different intervals of sensitivities considered offer a good tradeoff between sensitivity and specificity. The simulation study shows that the estimator associated with our model is able to recover successfully the true overall shape of the functional covariate-adjusted partial area under the curve in different complex scenarios.

1. Introduction. Accurate diagnosis of disease is of great importance in public health, clinical practice, and medical research. The major goal of a diagnostic marker is to distinguish diseased from non-diseased subjects, and before a marker is approved for use in practice, its ability to discriminate between these two states must be rigorously assessed through statistical analysis. The accuracy of a dichotomous marker, a marker that yields binary results (e.g., positive or negative), can be summarized by its sensitivity and specificity. The sensitivity, Se, is the marker specific probability of correctly detecting diseased subjects, while the specificity, Sp, is the test specific probability of correctly identifying non-diseased subjects. In turn, the receiver operating characteristic (ROC) curve, which is a plot of $1-\text{Sp}$ against Se, for all cutoff points that can be used to convert continuous marker outcomes into dichotomous outcomes, is the most popular tool to assess the discriminatory ability of a continuous marker.

Keywords and phrases: Arterial oxygen saturation, average specificity, biomarker, functional covariate-adjustment, gamma-glutamyl transferase, kernel regression, metabolic syndrome, partial area under the curve, sensitivity, specificity-receiver operating characteristic curve.

The most commonly used summary index of diagnostic accuracy is the area under the ROC curve (AUC), which can be interpreted as the average sensitivity for all specificity values, or conversely, as the average specificity over all sensitivity values. The AUC can also be interpreted as the probability that a randomly selected diseased individual has a greater marker outcome than that for a randomly selected non-diseased individual. However, in most diagnostic situations, the area under the curve summarizes the marker's performance for sensitivity or specificity values of no clinical interest. For instance, when screening a population for a certain disease for which further testing and/or treatment is invasive or expensive, the region of the ROC curve corresponding to high specificities is of primordial interest, while, on the other hand, when testing for a harmful disease, it is desirable to maintain a high sensitivity. The concept of partial area under the curve, which is a meaningful summary measure of diagnostic accuracy when only certain intervals of sensitivity or specificity are clinically relevant, thus arises naturally in such contexts.

An overall partial area under the curve is useful to summarize the accuracy of a marker over a particular region of sensitivities or specificities in a homogeneous population. However, a marker's ability to discriminate between diseased and non-diseased states may vary substantially over subject-specific characteristics. It is, therefore, important to understand how the performance of a marker evolves over covariates. Although there are numerous articles dedicated to the partial area under the curve in the recent years (see, among others, Hung and Chiang 2011; Wang and Chang 2011; Adimari and Chiogna 2012; Ma et al. 2013; Gigliarano et al. 2014), approaches in the literature adjusting the partial AUC for covariates, to our knowledge, include only Dodd and Pepe (2003) and Cai and Dodd (2008), and both consider only the cases where the covariate is univariate or multivariate and restrict the partial area to a relevant range of specificities. However, nowadays, more often than not, the covariates are curves or images, thus raising the need for new methodology that can properly handle and analyze such data. One of our main contributions is the development of a functional covariate-adjusted estimator for the partial area under the curve, for the case where only particular ranges of sensitivity are clinically meaningful; our proposed estimator can be regarded as a functional covariate-adjusted Mann–Whitney type of statistic.

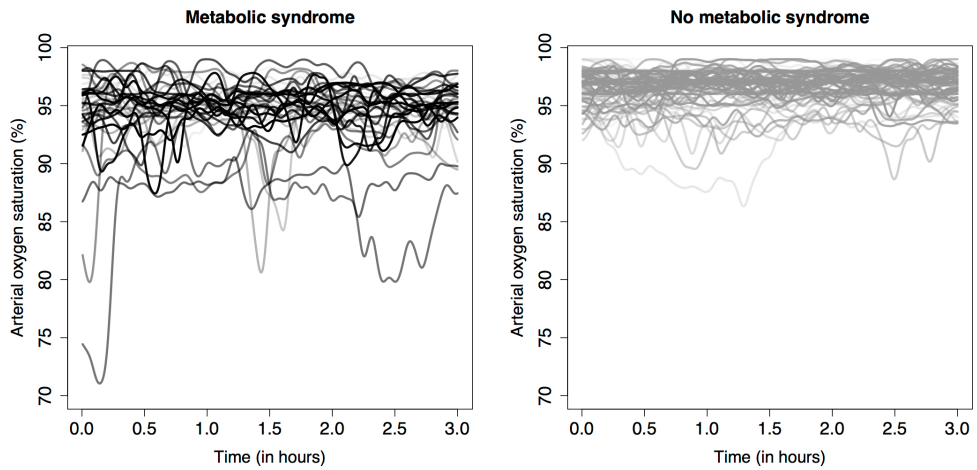
A metabolic syndrome application motivates our methodological developments. Metabolic syndrome is a cluster of risk factors that occur together and increase the risk of, among others, cardiovascular disease, stroke, type-2 diabetes, and atherosclerosis. According to the American Heart Association (*http://www.heart.org*), individuals with metabolic syndrome have a two-fold increase in risk for heart attack or cardiovascular disease, and a five-fold increase risk for developing diabetes when compared with individuals who do not have metabolic syndrome. According to the same source, almost 35% of American adults are affected by this condition and it has also been acknowledged that the prevalence of this syndrome is also increasing worldwide (Eckel, Grundy, and Zimmet, 2005). The identification of diagnostic markers for metabolic syndrome is thus of crucial importance. Serum gamma-glutamyl transferase (GGT), a well-known marker of alcohol consumption and liver disfunction, is also associated with components of the metabolic syndrome. In fact, elevated GGT is an indicator of the presence of metabolic syndrome (Lee et al., 2007). GGT also has the nice feature of being a low-cost and frequently used laboratory marker. Given that metabolic syndrome is a serious condition that places individuals at higher risk for cardiovascular disease, stroke, and diabetes, it is critical to maintain a high sensitivity when screening for this condition; thus avoiding misdiagnosing subjects with metabolic syndrome, so that intervention can be initiated. However, since re-

stricting the sensitivity above a preselected value corresponds to operating with lower cutoff values in practice, it is expected that restricting the sensitivity to high values will result in a loss of specificity. Thus, for each interval that might make clinical sense to restrict the sensitivity, it is mandatory to ascertain the tradeoffs between sensitivity and specificity. Recent studies suggest a strong association between GGT levels in serum and nocturnal hypoxemia, which is characterized by a decrease in arterial oxygen saturation of hemoglobin (Gude et al., 2009). To this end, the arterial oxygen saturation was measured densely over patient's sleep, leading to a curve of oxygen measurements per patient (see more details in Section 2).

Another contribution of this work rests on applying the developed methods for approaching the question: "How does the restriction of GGT sensitivity above a preselected value affect its corresponding specificity and how does such tradeoff change for different curves of arterial oxygen saturation?". We thus extend the work of Inácio et al. (2012), who introduced the functional covariate-adjusted ROC curve and studied how the discriminatory ability of the GGT, as a marker for metabolic syndrome, was affected by different curves of oxygen saturation when neither the sensitivity nor the specificity were restricted.

In Section 2, we describe our motivating metabolic syndrome data. Preliminary concepts, our modeling framework for the estimation of the nonparametric functional covariate-adjusted partial area restricted to a meaningful sensitivity interval and its practical implementation are presented in Section 3. In Section 4, we apply our methods to the motivating metabolic syndrome case study. In Section 5, we assess the finite sample performance of our methods by simulation. Concluding remarks are given in Section 6.

2. Metabolic syndrome data. With the aim of evaluating the hypothesized relationship between GGT levels and nocturnal hypoxemia, a study was conducted using a sample of 220 subjects randomly selected from the Galician (NW Spain) adult population; further details on this study can be found in Gude et al. (2009). Metabolic syndrome diagnosis was conducted through the Adult Treatment Panel III criterion, which is based on the following five items: (a) abdominal obesity, (b) hypertriglyceridaemia, (c) low HDL-cholesterol levels, (d) increased blood pressure, and (e) hyperglycemia. Subjects who met at least three of these criteria were classified as metabolic syndrome patients. Arterial oxygen saturation was recorded at the patient's home using a pulse oximeter, which is a noninvasive monitoring technique used to estimate the percentage of hemoglobin saturated with oxygen at the time of measurement. Measurements were made every 20 seconds during the patient's sleep, thus leading to genuine functional data. As it is known that the nocturnal arterial oxygen saturation has different patterns during the several sleep phases, for all subjects we skipped the first two hours of measurements and focused on the following three hours. Hence, at the end, we had a total of 540 measurements per subject. Since GGT values are elevated among regular drinkers, we restricted the analysis to 115 women who reported no alcohol consumption, so that possible higher values are not due to differences in alcohol consumption and gender. In short, the data analyzed consist of GGT values plus three hours of oxygen saturation measurements for 35 women with metabolic syndrome and 80 women without metabolic syndrome. Figure 1 of the supplemental article (Inácio de Carvalho et al., 2016) shows the histogram and variable-width boxplot of the GGT (both in the original and in the log scale) for the two group of women, while Figure 1 shows the curves of arterial oxygen saturation for women with metabolic syndrome (a) and women without metabolic syndrome (b).



(a) (b)

FIG 1. *Levels of arterial oxygen saturation for women with metabolic syndrome (a) and for women without metabolic syndrome (b).*

3. Functional covariate-adjusted partial area under the specificity-ROC curve.

3.1. *Preliminaries.* Let Y_D and $Y_{\bar{D}}$ be two independent random variables denoting the marker outcomes in the diseased and non-diseased populations, with cumulative distribution functions F_D and $F_{\bar{D}}$, respectively. Further, let c be a cutoff value for defining a positive marker result and, without loss of generality, we proceed with the assumption that a subject is classified as diseased when the marker outcome is equal or greater than c, and as non-diseased when it is below c. Then, for each cutoff value c, the sensitivity associated with such decision criterion is $\mathrm{Se}(c) = \Pr(Y_D \geqslant c) = 1 - F_D(c)$, while the specificity is $\mathrm{Sp}(c) = \Pr(Y_{\bar{D}} < c) = F_{\bar{D}}(c)$. The ROC curve represents the plot $\{(1 - F_{\bar{D}}(c), 1 - F_D(c)) : c \in \mathbb{R}\}$ and provides a visual description of the tradeoffs between the sensitivity and specificity as the cutoff c changes. For $0 \leqslant p = 1 - F_{\bar{D}}(c) \leqslant 1$, the ROC curve can be equivalently written as $\mathrm{ROC}(p) = 1 - F_D\{F_{\bar{D}}^{-1}(1-p)\}$. The AUC is given by $\mathrm{AUC} = \int_0^1 \mathrm{ROC}(p)\mathrm{d}p$.

As shown by Dodd and Pepe (2003), interpretations of the partial area corresponding to sensitivities in a specified interval are more easily obtained by performing a 270° rotation to the ROC curve, so to obtain the graph

$$\{(1 - F_D(c), F_{\bar{D}}(c)) : c \in \mathbb{R}\} = \{(\mathrm{Se}(c), \mathrm{Sp}(c)) : c \in \mathbb{R}\}. \tag{3.1}$$

The curve in (3.1) is referred to as the specificity-ROC curve, $\mathrm{ROC}_{\mathrm{Sp}}$, since specificity is plotted on the y-axis (Dodd and Pepe, 2003); see Figure 2. For $p = 1 - F_D(c)$, the $\mathrm{ROC}_{\mathrm{Sp}}$ curve can be expressed as

$$\mathrm{ROC}_{\mathrm{Sp}}(p) = F_{\bar{D}}\{F_D^{-1}(1-p)\}, \quad 0 \leqslant p \leqslant 1.$$

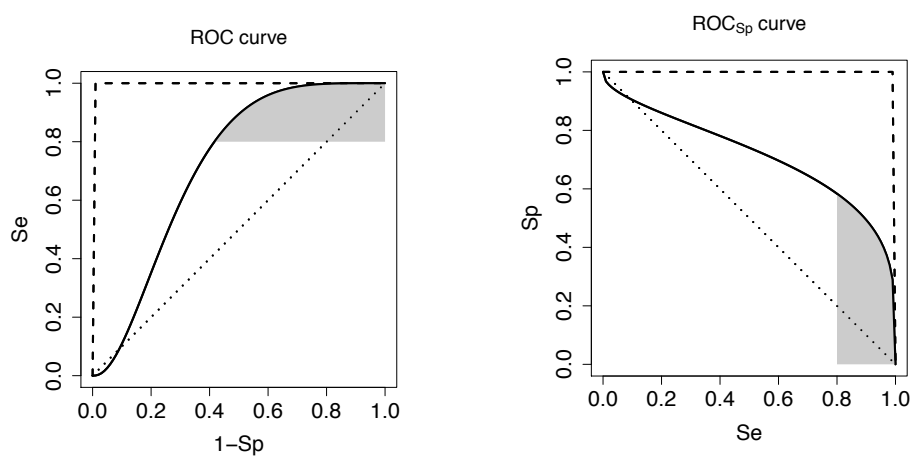
The partial AUC over the range of sensitivities $(u, 1)$ is defined as

$$\mathrm{pAUC}_{\mathrm{Se}}(u) = \int_u^1 \mathrm{ROC}_{\mathrm{Sp}}(p)\mathrm{d}p.$$

Although we consider the interval $(u, 1)$, note that the $\mathrm{pAUC}_{\mathrm{Se}}$ between u_1 and u_2, with $0 \leqslant u_1 < u_2 \leqslant 1$, can be obtained by $\mathrm{pAUC}_{\mathrm{Se}}(u_1) - \mathrm{pAUC}_{\mathrm{Se}}(u_2)$. In Figure 2, we plot the $\mathrm{ROC}_{\mathrm{Sp}}$ and ROC curves corresponding to a perfect marker, a useless marker, and a marker whose underlying distributions are $Y_D \sim \mathrm{N}(2, 0.8^2)$ and $Y_{\bar{D}} \sim \mathrm{N}(1, 1.55^2)$. For an useless marker, $\mathrm{Sp}(c) = 1 - \mathrm{Se}(c)$ for all c, and $\mathrm{pAUC}_{\mathrm{Se}} = (1-u)^2/2$ while, on the other hand, a curve that reaches the upper right corner, with $\mathrm{Se}(c) = \mathrm{Sp}(c) = 1$ for some c, corresponds to a perfect marker and $\mathrm{pAUC}_{\mathrm{Se}} = 1 - u$.

In Figure 2, we also plot the regions of the ROC and $\mathrm{ROC}_{\mathrm{Sp}}$ curves, whose underlying distributions are the same as those referred to above, corresponding to sensitivities in the interval $(0.8, 1)$. As can be observed, computations of the partial area of interest are possible in both cases, but they are much more straightforward when working with the $\mathrm{ROC}_{\mathrm{Sp}}$ curve. The area of the region of the ROC curve plotted in Figure 2 corresponding to sensitivities over $(0.8, 1)$ can be computed by $\int_{\mathrm{ROC}^{-1}(0.8)}^{1} \mathrm{ROC}(p)\mathrm{d}p - \{1 - \mathrm{ROC}^{-1}(0.8)\} \times 0.8$, while the area on the corresponding region of the $\mathrm{ROC}_{\mathrm{Sp}}$ curve is simply the area under this curve, i.e., $\int_{0.8}^{1} \mathrm{ROC}_{\mathrm{Sp}}(p)\mathrm{d}p$.

As pointed out by Jiang, Metz, and Nishikawa (1996), the normalized $\mathrm{pAUC}_{\mathrm{Se}}$, $\mathrm{pAUC}_{\mathrm{Se}}(u)/(1-u)$, offers several advantages. First, it can be interpreted as the average specificity over the interval $(u, 1)$, denoted by Average $\mathrm{Sp}(u)$, and thus summarizes the trade-off between gains in sensitivity and losses in specificity. Second, the normalization allows us to express the values of the partial area on a numerical scale similar to that of AUC, since the normalized area ranges from $(1-u)/2$ to 1. Hereby, all reported results will be based on the average specificity.



(a) (b)

FIG 2. *ROC curve (a) and ROC_{Sp} curve (b) corresponding to a perfect marker (dashed line), a useless marker (dotted line), and a marker whose underlying distributions are $Y_D \sim N(2, 0.8^2)$ and $Y_{\bar{D}} \sim N(1, 1.55^2)$ (solid line); the shaded regions correspond to the partial AUC over the interval of sensitivities* $(0.8, 1)$*, represented on the ROC curve (a) and ROC_{Sp} curve (b).*

3.2. *Modeling framework.* The key object of our model framework is the functional covariate-adjusted $\mathrm{ROC}_{\mathrm{Sp}}$ curve, which consists of a 270° rotation of the functional covariate-adjusted ROC curve proposed by Inácio et al. (2012). For a random curve X, whose realizations are defined on a certain functional space $\mathbb{X}$, the functional covariate-adjusted $\mathrm{ROC}_{\mathrm{Sp}}$ curve is

defined as

$$\mathrm{ROC}_{\mathrm{Sp}}(p \mid X) = F_{\bar{D}}\{F_D^{-1}(1-p \mid X) \mid X\}, \quad 0 \leqslant p \leqslant 1, \tag{3.2}$$

where F_D and $F_{\bar{D}}$ are the conditional distribution functions of Y_D and $Y_{\bar{D}}$ given a curve X, i.e.

$$F_D(y \mid X) = \Pr(Y_D \leqslant y \mid X), \quad F_{\bar{D}}(y \mid X) = \Pr(Y_{\bar{D}} \leqslant y \mid X).$$

For each curve X of interest, we possibly obtain a different $\mathrm{ROC}_{\mathrm{Sp}}$ curve. We incorporate functional covariate information on the $\mathrm{ROC}_{\mathrm{Sp}}$ curve in an alternative way, through the specification of functional regression models for Y_D and $Y_{\bar{D}}$. More specifically, we assume that the relationship between Y_D and $Y_{\bar{D}}$ and a curve X, can be expressed using location–scale regression models

$$Y_D = \mu_D(X) + \sigma_D(X)\varepsilon_D, \quad Y_{\bar{D}} = \mu_{\bar{D}}(X) + \sigma_{\bar{D}}(X)\varepsilon_{\bar{D}}, \tag{3.3}$$

where $\mu_s(X) = \mathrm{E}(Y_s \mid X)$ and $\sigma_s^2(X) = \mathrm{var}(Y_s \mid X)$ are the conditional mean and conditional variance functions, for $s \in \{D, \bar{D}\}$, and the errors ε_D and $\varepsilon_{\bar{D}}$ are independent of X. The specification in (3.3) allows us to rewrite the functional covariate-adjusted $\mathrm{ROC}_{\mathrm{Sp}}$ curve in (3.2) as

$$\mathrm{ROC}_{\mathrm{Sp}}(p \mid X) = F_{\varepsilon_{\bar{D}}}\left\{\frac{\mu_D(X) - \mu_{\bar{D}}(X)}{\sigma_{\bar{D}}(X)} + \frac{\sigma_D(X)}{\sigma_{\bar{D}}(X)} F_{\varepsilon_D}^{-1}(1-p)\right\}, \quad 0 \leqslant p \leqslant 1, \tag{3.4}$$

where F_{ε_D} and $F_{\varepsilon_{\bar{D}}}$ are the distribution functions of the regression errors ε_D and $\varepsilon_{\bar{D}}$, respectively. An advantage of this formulation is that the distribution and quantile functions of the regression errors are not conditional, thus reducing the computational burden (Pardo-Fernandéz, Rodríguez-Álvarez, and Van Keilegom, 2014, p. 31).

For a given lower limit of sensitivity, $u \in (0, 1)$, the functional covariate-adjusted $\mathrm{pAUC}_{\mathrm{Se}}$ is defined as

$$\mathrm{pAUC}_{\mathrm{Se}}(u \mid X) = \int_u^1 \mathrm{ROC}_{\mathrm{Sp}}(p \mid X)\mathrm{d}p.$$

Integrating by parts yields a simple, yet important, result which will be the basis for the construction of our estimator,

$$\mathrm{pAUC}_{\mathrm{Se}}(u \mid X) = -u\mathrm{ROC}_{\mathrm{Sp}}(u \mid X) - \int_u^1 p \; \mathrm{dROC}_{\mathrm{Sp}}(p \mid X) = \mathrm{E}[\max\{u, Z_X\} \mid X] - u. \tag{3.5}$$

Here, $Z_X = S_D(Y_{\bar{D}} \mid X) = 1 - F_D(Y_{\bar{D}} \mid X)$ and $\mathrm{E}[\max\{u, Z_X\} \mid X]$ is computed with respect to Z_X. Note that the survival function of Z_X is $\mathrm{ROC}_{\mathrm{Sp}}(p \mid X)$; see Section 1.1 of the supplemental article (Inácio de Carvalho et al., 2016) for further details. The average specificity over the interval $(u, 1)$ for a curve X is thus given by

$$\text{Average Sp}(u \mid X) = \mathrm{pAUC}_{\mathrm{Se}}(u \mid X)/(1-u). \tag{3.6}$$

3.3. *Proposed estimator and its implementation.* Let $\{(X_{Dj}, Y_{Dj})\}_{j=1}^{n_D}$ and $\{(X_{\bar{D}i}, Y_{\bar{D}i})\}_{i=1}^{n_{\bar{D}}}$, be random samples from the diseased and non-diseased groups, respectively, where X_{Dj} is the covariate curve for the jth subject and Y_{Dj} is the test outcome for the jth subject; $X_{\bar{D}i}$ and $Y_{\bar{D}i}$ are analogously defined.

Suppose for now that we have estimates of the conditional mean and conditional variance functions, and respectively denote these by $\widehat{\mu}_s(X)$ and $\widehat{\sigma}_s^2(X)$, for $s \in \{D, \bar{D}\}$; later we discuss how they can be estimated. The starting point for the construction of our estimator is the estimation of the standardized residuals in each group

$$\widehat{\varepsilon}_{Dj} = \frac{Y_{Dj} - \widehat{\mu}_D(X_{Dj})}{\widehat{\sigma}_D(X_{Dj})}, \quad \widehat{\varepsilon}_{\bar{D}i} = \frac{Y_{\bar{D}i} - \widehat{\mu}_{\bar{D}}(X_{\bar{D}i})}{\widehat{\sigma}_{\bar{D}}(X_{\bar{D}i})},$$

and using these estimated standardized residuals we can construct the so-called working samples (Yao, Craiu, and Reiser, 2010), $\{\widehat{Y}_{Dj|X}\}_{j=1}^{n_D}$ and $\{\widehat{Y}_{\bar{D}i|X}\}_{i=1}^{n_{\bar{D}}}$, as if they were all observed at a curve X,

$$\widehat{Y}_{Dj|X} = \widehat{\mu}_D(X) + \widehat{\sigma}_D(X)\widehat{\varepsilon}_{Dj}, \quad \widehat{Y}_{\bar{D}i|X} = \widehat{\mu}_{\bar{D}}(X) + \widehat{\sigma}_{\bar{D}}(X)\widehat{\varepsilon}_{\bar{D}i}.$$

An empirical version of (3.5) leads us to the following functional covariate-adjusted estimator for $\mathrm{pAUC}_{\mathrm{Se}}$,

$$\begin{aligned}\widehat{\mathrm{pAUC}}_{\mathrm{Se}}(u \mid X) &= \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}}\left[\max\left\{u, \frac{1}{n_D}\sum_{j=1}^{n_D} I(\widehat{Y}_{Dj|X} \geqslant \widehat{Y}_{\bar{D}i|X})\right\}\right] - u \\ &= \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}}\left[\max\left\{u, \frac{1}{n_D}\sum_{j=1}^{n_D} I(\widehat{\mu}_D(X) + \widehat{\sigma}_D(X)\widehat{\varepsilon}_{Dj} \geqslant \widehat{\mu}_{\bar{D}}(X) + \widehat{\sigma}_{\bar{D}}(X)\widehat{\varepsilon}_{\bar{D}i})\right\}\right] - u,\end{aligned} \tag{3.7}$$

where $I(\cdot)$ denotes the indicator function and, hence

$$\widehat{\text{Average}}\ \mathrm{Sp}(u \mid X) = \widehat{\mathrm{pAUC}}_{\mathrm{Se}}(u \mid X)/(1-u). \tag{3.8}$$

The estimator in (3.7) can be regarded as a functional covariate-adjusted Mann–Whitney type of statistic. It should be noticed that we are often interested in estimating Average $\mathrm{Sp}(u \mid X)$ even for curves X which were not measured in either group or both. We remark that when $u = 0$, (3.7) is an extension to the functional covariate case of the estimator of Yao, Craiu, and Reiser (2010). In the article (Inácio de Carvalho et al., 2016) we provide a calculus-based construction of (3.7). An estimate of the $\mathrm{pAUC}_{\mathrm{Se}}(u \mid X)$, similar to that obtained by using (3.7), could be obtained by performing a 270° rotation of the estimated functional covariate-adjusted ROC curve as presented in Inácio et al. (2012) and then integrating over the interval $(u, 1)$. However the estimator in (3.7) has the nice feature of having a closed form expression, hence not requiring rotation and numerical integration, which should be appealing for practitioners.

To use (3.8) in practice, we need to estimate only the mean and variance functions, which we estimate through an extension of the Nadaraya–Watson estimator to the functional context (Ferraty and Vieu, 2002). The main reason for this choice is that with functional data it is difficult to know which parametric model would best fit the data, so nonparametric regression comes naturally in this context. Given a random curve X, the estimates of

$\boldsymbol{\mu}(X) = (\mu_D(X), \mu_{\bar{D}}(X))$ and $\boldsymbol{\sigma}^2(X) = (\sigma^2_D(X), \sigma^2_{\bar{D}}(X))$ are, respectively, given by

$$\begin{aligned} \widehat{\boldsymbol{\mu}}(X) &= \left(\frac{\sum_{j=1}^{n_D} K\{h_{D\mu}^{-1} d(X, X_{Dj})\} Y_{Dj}}{\sum_{j=1}^{n_D} K\{h_{D\mu}^{-1} d(X, X_{Dj})\}}, \frac{\sum_{i=1}^{n_{\bar{D}}} K\{h_{\bar{D}\mu}^{-1} d(X, X_{\bar{D}i})\} Y_{\bar{D}i}}{\sum_{i=1}^{n_{\bar{D}}} K\{h_{\bar{D}\mu}^{-1} d(X, X_{\bar{D}i})\}} \right), \\ \widehat{\boldsymbol{\sigma}}^2(X) &= \left(\frac{\sum_{j=1}^{n_D} K\{h_{D\sigma}^{-1} d(X, X_{Dj})\} \{Y_{Dj} - \widehat{\mu}_D(X_{Dj})\}^2}{\sum_{j=1}^{n_D} K\{h_{D\sigma}^{-1} d(X, X_{Dj})\}}, \frac{\sum_{i=1}^{n_{\bar{D}}} K\{h_{\bar{D}\sigma}^{-1} d(X, X_{\bar{D}i})\} \{Y_{\bar{D}i} - \widehat{\mu}_{\bar{D}}(X_{\bar{D}i})\}^2}{\sum_{i=1}^{n_{\bar{D}}} K\{h_{\bar{D}\sigma}^{-1} d(X, X_{\bar{D}i})\}}, \right), \end{aligned} \tag{3.9}$$

where K is a kernel function, $h_{D\mu}, h_{\bar{D}\mu}, h_{D\sigma}$, and $h_{\bar{D}\sigma}$ are positive smoothing parameters or bandwidths, and $d : \mathbb{X} \times \mathbb{X} \mapsto [0, \infty)$ is a semimetric (van der Vaart, 1998, p. 255).

Once we have estimates of the regression and variance functions, we can compute our estimator by substituting (3.9) into (3.7). The practical implementation of our estimator requires three choices: kernel, bandwidths, and semimetric. It is well known that the choice of the kernel has little impact on the estimates and so, following the most commonly made choice, throughout this work we have used the asymmetric Gaussian kernel

$$K(w) = \frac{2}{\sqrt{2\pi}} \exp(-w^2/2) I(w \geqslant 0). \tag{3.10}$$

In turn, the bandwidth plays a key role on the performance of the estimator, and its choice entails a bias–variance tradeoff. We have chosen the four bandwidths in a data-driven way using generalized cross validation (GCV). The criterion for selecting, for instance, $h_{D\mu}$, is to choose the bandwidth which minimizes the following GCV objective function

$$\mathrm{GCV}(\mathrm{h}_{D\mu}) = \frac{1}{n_D} \sum_{j=1}^{n_D} \left(\frac{Y_{Dj} - \mu_D(X_{Dj})}{1 - n_D \sum_{j=1}^{n_D} S_{Djj}(h_{D\mu})} \right)^2, \quad S_{Djj}(h_{D\mu}) = \frac{K\{h_{D\mu}^{-1} d(X_{Dj}, X_{Dj})\}}{\sum_{l=1}^{n_D} K\{h_{D\mu}^{-1} d(X_{Dj}, X_{Dl})\}}. \tag{3.11}$$

The estimator of $\sigma^2_D(X)$ ($\sigma^2_{\bar{D}}(X)$) depends on both bandwidths, $h_{D\mu}$ and $h_{D\sigma}$ ($h_{\bar{D}\mu}$ and $h_{\bar{D}\sigma}$), which are selected sequentially. Finally, the semimetric d, which measures the proximity between the curves in the functional space $\mathbb{X}$, must be related to the particular features of the data at hand. Specifically, when the curves are smooth, (Ferraty and Vieu, 2006, pp. 28–32) suggest using the L_2-norm of the qth derivative of the curves and thus, in what follows, we use the $L_2[a, b]$-norm, i.e.,

$$d(X, X^*) = \left[\int_a^b \{X(t) - X^*(t)\}^2 \mathrm{d}t \right]^{1/2},$$

for any two curves X and X^*.

3.4. *Bootstrap-based inference.* Confidence intervals for the covariate-adjusted average specificity can be obtained through the bootstrap (Davison and Hinkley, 1997). As it is advised in the nonparametric kernel regression literature (Härdle and Marron, 1991, p. 781), we use a bootstrap of the residuals to resample the regression models and then the percentile method to obtain pointwise bootstrap intervals for the functional covariate-adjusted average specificity; a related bootstrap scheme can be found in Ferraty, Van Keilegom and Vieu (2010). More specifically, the bootstrap confidence interval for $\widehat{\text{Average}}\,\mathrm{Sp}(u \mid X)$ is obtained with the following resampling algorithm.

Bootstrap algorithm
Consider a fixed $X \in \mathbb{X}$.
`for` $(b = 1, \ldots, B)$ `do`:

Step 1. Sample with replacement from the estimated standardized residuals $\{\widehat{\varepsilon}_{\bar{D}i}\}_{i=1}^{n_{\bar{D}}}$ and $\{\widehat{\varepsilon}_{Dj}\}_{j=1}^{n_D}$ to form bootstrap sets $\{\widehat{\varepsilon}_{\bar{D}i}^{(b)}\}_{i=1}^{n_{\bar{D}}}$ and $\{\widehat{\varepsilon}_{Dj}^{(b)}\}_{j=1}^{n_D}$.

Step 2. Use the estimated mean and variance functions from the observed data to construct bootstrap samples at curve X,

$$\widehat{Y}_{\bar{D}i|X}^{(b)} = \widehat{\mu}_{\bar{D}}(X) + \widehat{\sigma}_{\bar{D}}(X)\widehat{\varepsilon}_{\bar{D}i}^{(b)}, \quad \widehat{Y}_{Dj|X}^{(b)} = \widehat{\mu}_{D}(X) + \widehat{\sigma}_{D}(X)\widehat{\varepsilon}_{Dj}^{(b)}.$$

Step 3. Estimate Average $\mathrm{Sp}(u \mid X)$ using (3.8), i.e., compute

$$\widehat{\text{Average}}\ \mathrm{Sp}^{(b)}(u \mid X) = \left(\frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} \max\left\{ u, \frac{1}{n_D} \sum_{j=1}^{n_D} I(\widehat{Y}_{Dj|X}^{(b)} \geqslant \widehat{Y}_{\bar{D}i|X}^{(b)}) \right\} - u \right) / (1-u).$$

Once this process is completed, and according to the percentile method, a bootstrap confidence interval for Average $\mathrm{Sp}(u \mid X)$ of confidence level $1-\alpha$ is given by

$$\left(\widehat{\text{Average}}\ \mathrm{Sp}(u \mid X)^{\alpha/2}, \widehat{\text{Average}}\ \mathrm{Sp}(u \mid X)^{1-\alpha/2}\right),$$

where $\widehat{\text{Average}}\ \mathrm{Sp}^{(b)}(u \mid X)$ represents the τth percentile of the estimated $\widehat{\text{Average}}\ \mathrm{Sp}^{(b)}(u \mid X)$, for $b = 1, \ldots, B$.
It is worth mentioning that this bootstrap scheme tends to produce intervals that show some undercoverage

4. Metabolic syndrome data revisited.

4.1. *Exploratory analysis.* Our data consist of GGT values (in international units per milliliter, IU/ml) plus three hours of oxygen saturation for 35 women with metabolic syndrome and 80 women without metabolic syndrome. We use the same preprocessing step as in Inácio et al. (2012), and thus the arterial oxygen saturation curves were smoothed with a Gaussian kernel (Febrero-Bande and Oviedo de la Fuente 2012, Section 2.2); description of additional details of the data can be found in Section 2. Women with metabolic syndrome tend to have larger GGT values (IQR $= 26 - 15$, IU/ml) than women without metabolic syndrome (IQR $= 18.00 - 10.75$, IU/ml). Figure 1 (top) of the supplemental article (Inácio de Carvalho et al., 2016) shows the histogram along with the variable-width boxplot of GGT for each group of women, while Figure 1 (bottom) of the supplemental article (Inácio de Carvalho et al., 2016) displays the same information but for the log transformed GGT. Since the log transformation helped to symmetrize the GGT data in both groups, hereby we proceed with the log transformed data. In Figure 1 we present the arterial oxygen saturation curves, and as it can be observed there is a clear difference between arterial oxygen saturation curves of women with and without metabolic syndrome, with women suffering from metabolic syndrome tending to have lower levels of oxygen saturation and higher variance. This is in line with what has been reported in the literature (Gude et al., 2009). Results from an exploratory functional principal component analysis reported in the supplemental article (Inácio de Carvalho et al., 2016) further reveal that the estimated scores associated with the

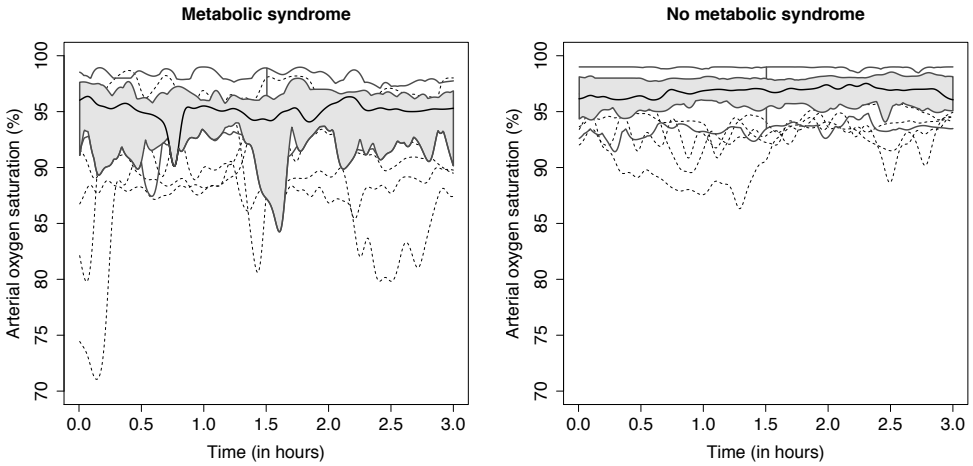
first principal component of the arterial oxygen saturation curves, by themselves, have already a quite good ability to discriminate between women with and without metabolic syndrome [AUC $= 0.811$ $(0.727, 0.887)$]. In Figure 3 we present the functional boxplot (Sun and Genton, 2011) of each group of curves; once more, the different type of pattern of arterial oxygen saturation in each group of women is evident. Some curves are also identified as atypical.

Additionally, we have also investigated in the case where the oxygen saturation is not taken into account, how restricting the sensitivity of the GGT above a preselected value affects its corresponding specificity. To our knowledge there is no current standard of clinically meaningful intervals for the sensitivity when testing for metabolic syndrome, and hence we have evaluated different intervals, namely we have considered $u = 0, 0.6, 0.7, 0.8, 0.9,$ and 0.95. The no covariate-adjusted estimator of the average specificity is computed using $\widehat{\mathrm{pAUC}}_{\mathrm{Se}}(u)/(1-u)$, where

$$\widehat{\mathrm{pAUC}}_{\mathrm{Se}}(u) = \frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} \left[\max \left\{ u, \frac{1}{n_D} \sum_{j=1}^{n_D} \left(I(Y_{Dj} > Y_{\bar{D}i}) + \frac{1}{2} I(Y_{Dj} = Y_{\bar{D}i}) \right) \right\} \right] - u,$$

where Y_{Dj} and $Y_{\bar{D}i}$, respectively, denote GGT levels for women with and without metabolic syndrome, for $j = 1, \ldots, 35$ and $i = 1, \ldots, 80$. Although GGT is measured in a continuous scale, in practice ties can occur, and so the extra term $(1/2) \times I(Y_{Dj} = Y_{\bar{D}i})$ corrects for ties. However, we point out that with our oxygen saturation-adjusted $\widehat{\mathrm{pAUC}}_{\mathrm{Se}}$ estimator (3.7) ties do not occur. Table 1 of the supplemental article (Inácio de Carvalho et al., 2016) presents the resulting estimates along with 95% bootstrap confidence intervals ($B = 1000$). As can be observed, the average specificity decreases from 0.790 $(0.709, 0.868)$ when the sensitivity is not restricted to 0.459 $(0.330, 0.630)$, when $u = 0.95$, i.e., when the sensitivity belongs to the interval $(0.95, 1)$. Between these two extreme cases, the intervals corresponding to sensitivities in $(0.7, 1)$ and $(0.8, 1)$ seem to provide a good balance between gains in sensitivity and losses in specificity.

4.2. *Arterial oxygen saturation-adjusted analyses.* Our goal is to assess how the tradeoffs between sensitivity and specificity of the GGT, when restricting the sensitivity to some prespecified intervals, might change for different curves of arterial oxygen saturation. We thus go one step ahead of the analysis presented in Inácio et al. (2012) who assessed the discriminatory ability of the GGT in the absence of any restriction. We consider the same sensitivity intervals that were considered for the no oxygen saturation-adjusted analyses. In the results reported below we have used the asymmetric Gaussian kernel (3.10), the bandwidths were chosen using the generalized cross validation criterion (3.11), and the semimetric d has been set to be the usual $L^2[0, 3]$-norm. For the curves presented in Figure 4, we have estimated the $\mathrm{ROC}_{\mathrm{Sp}}$ curves (Figure 2 of the supplemental article Inácio de Carvalho et al. 2016) and the corresponding average specificity, which are presented in Table 1. Roughly, the results suggest that the discriminatory ability is better for curves with higher levels of oxygen saturation, across all intervals. It is interesting to note (see also Table 2 in the supplemental article Inácio de Carvalho et al. 2016) that for curves that present substantial variation, there is a marked decreasing in average specificity from $u = 0$ to $u = 0.6$. For oxygen saturation curves corresponding to high values, the intervals $(0.7, 1)$ and $(0.8, 1)$, as in the no oxygen saturation-adjusted analyses, seem to provide a good sensitivity/specificity balance. While it would be interesting to smooth the $\mathrm{ROC}_{\mathrm{Sp}}$ curves, the corresponding estimates of the covariate-adjusted



(a) (b)

FIG 3. *Functional boxplot of arterial oxygen saturation curves for women with metabolic syndrome (a) and for women without metabolic syndrome (b). Outlier curves are represented by dashed lines, while the solid black line is the median curve. The solid dark gray curves denote envelopes.*

	$\widehat{\text{Average}}\ \text{Sp}(u \mid X)$ (95% bootstrap CI)			
	X_{D4}	X_{D12}	$X_{\bar{D}51}$	$X_{\bar{D}65}$
$u = 0$	0.631 $(0.493, 0.863)$	0.552 $(0.406, 0.857)$	0.877 $(0.782, 0.955)$	0.772 $(0.700, 0.910)$
$u = 0.6$	0.357 $(0.160, 0.735)$	0.225 $(0.088, 0.711)$	0.775 $(0.631, 0.907)$	0.595 $(0.467, 0.820)$
$u = 0.7$	0.316 $(0.131, 0.702)$	0.195 $(0.068, 0.678)$	0.758 $(0.593, 0.894)$	0.559 $(0.402, 0.796)$
$u = 0.8$	0.263 $(0.105, 0.659)$	0.157 $(0.054, 0.638)$	0.727 $(0.536, 0.882)$	0.500 $(0.318, 0.768)$
$u = 0.9$	0.191 $(0.075, 0.600)$	0.113 $(0.036, 0.573)$	0.663 $(0.448, 0.852)$	0.382 $(0.210, 0.721)$
$u = 0.95$	0.173 $(0.055, 0.560)$	0.093 $(0.025, 0.500)$	0.608 $(0.386, 0.830)$	0.313 $(0.163, 0.682)$

TABLE 1

Estimated average specificity and 95% bootstrap confidence interval for the four different curves of arterial oxygen saturation presented in Figure 4.

average specificity, which are the main object of interest, tend to be smooth (Figure 6), and thus we prefer to avoid adding another bandwidth into the analysis.

As in the univariate setting (González-Manteiga, Pardo-Fernández, and Van Keilegom, 2011; Inácio de Carvalho et al., 2013; Pardo-Fernandéz, Rodríguez-Álvarez, and Van Keilegom, 2014), it is helpful to graphically represent how the average specificity changes for different curves of arterial oxygen saturation, so that the practitioner can have an idea of what is going on. Below we construct a data-based set of curves conditionally on which we predict the functional covariate-adjusted average specificity. The main ideas underlying this construction are as follows. We are interested in predicting Average $\text{Sp}(u \mid X)$ over a grid of arterial oxygen saturation curves X of interest. Such curves should reflect the observed curves, and in some way they should have an order. In an univariate context, we would plot on the x-axis a grid over the domain of the covariate and on the y-axis the corresponding predicted average specificity values, but in the functional context this is impossible since we cannot plot curves on the x-axis. To overcome this difficulty, and with the aim of creating a grid of curves that resemble the structure and range of variation of observed arterial oxygen curves, we follow

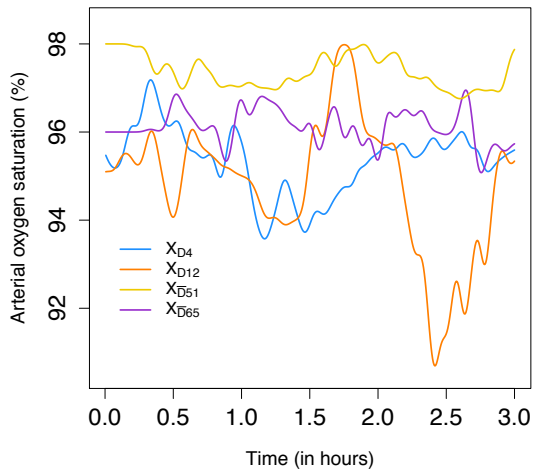


FIG 4. *Four observed curves of arterial oxygen saturation.*

Inácio et al. (2012) and consider curves of the form

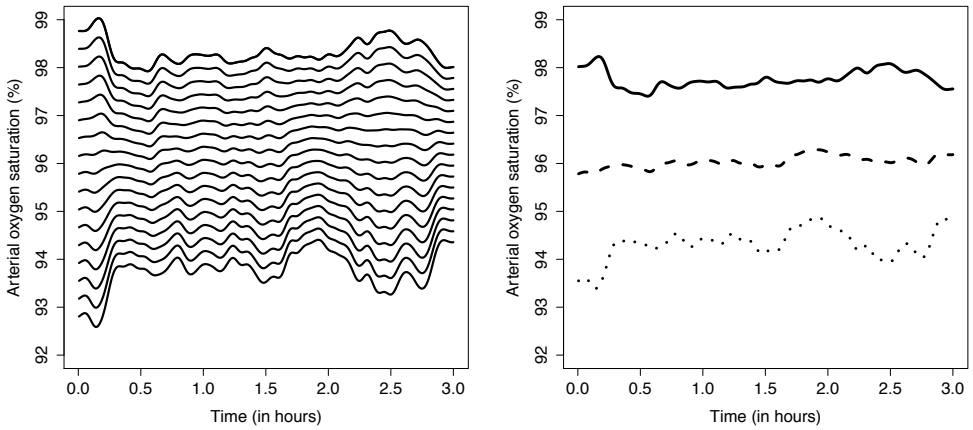
$$X_z(t) = \bar{X}(t) + z\widehat{v}_1(t), \quad t \in [0,3]. \tag{4.1}$$

Here $\bar{X}$ is the mean function of the pooled data, z is a weight parameter that lies on the range of the estimated first principal component scores, and $\widehat{v}_1(t)$ is the estimated eigenfunction associated with the first principal component. As we vary z, we obtain data-based curves, which act as a grid of curves conditionally on which we predict the functional-adjusted average specificity; to mimic the range of levels of arterial saturation in our data, we have chosen z to lie on the interval $[-4,4]$. The generated grid of oxygen curves, which are depicted in Figure 5 (a) vary, roughly, from an oxygen saturation of 93% to one of 99%, and formally consist of

$$\{(t, X_z(t)) : X_z(t) = \bar{X}(t) + z\widehat{v}_1(t); t \in [0,3]\}_{z\in\mathcal{Z}}, \quad \mathcal{Z} = \{-4+0.5k : k \in \{0,1,\ldots,16\}\}. \tag{4.2}$$

Note that $|\mathcal{Z}| = 17$, where $|\cdot|$ denotes the cardinal operator, and thus below we will be considering a grid of 17 curves; of course $\mathcal{Z}$ could consist of a finer grid, but the one in (4.2) suffices for our purposes. An oxygen saturation of 93% corresponds to a moderately low value, with 90% being already a worrying value from a clinical perspective. The standard desirable values of oxygen saturation are between 95% and 99%. It would be interesting to study the performance of GGT even for oxygen saturation values below 90%, since such values may still occur in practice, but as it can be seen from Figure 1 we have few women for which most part of the curve lies below 90%, thus ruling out the possibility of producing reliable inferences for such low levels.

There is a direct and unique correspondence between z and each curve of the generated grid (4.2). Higher values of z correspond to lower values of arterial oxygen saturation and vice-versa (due to a programming error Inácio et al. (2012) concluded the opposite). In Figure 5 (b) are shown arterial oxygen saturation curves corresponding to z values of -3, 0, and 3. This relationship depends on the signal of the estimated first eigenfunction, $\widehat{v}_1(t)$. If the signal



(a) (b)

FIG 5. *(a) Grid of artificial curves of arterial oxygen saturation constructed according to* (4.2) *and conditionally on which we will predict* $pAUC_{TP}$. *(b) Curves corresponding to* $z = -3$ *(solid line),* $z = 0$ *(dashed curve), and to* $z = 3$ *(dotted curve).*

of $\widehat{v}_1(t)$ is positive, then higher values of z will correspond to curves with higher values. On the other hand, if the signal of $\widehat{v}_1(t)$ is negative, then higher values of z will correspond to curves with lower values. In short, this relationship is data dependent. Using this trick, we can graphically represent $\widehat{\text{Average}}\ \text{Sp}(u \mid X_z)$ against z, and so that we can easily assess how the predicted values of $\widehat{\text{Average}}\ \text{Sp}(u \mid X_z)$ change for different curves of arterial oxygen saturation levels X_z. Note that this is not a data reduction step, since we use all the available data. This is just a way to graphically represent the average specificity over the sequence of data-driven curves in (4.2).

For the 17 curves presented in Figure 5 (a), we have estimated the oxygen saturation-adjusted average specificity using (3.8). The results are shown in Figure 6, where at each panel we present the estimated oxygen saturation-adjusted average specificity, along with the 95% bootstrap confidence bands ($B = 1000$), and the corresponding estimate of the average specificity that ignores the effect of oxygen saturation on GGT (and its 95% bootstrap confidence interval). We can thus have an idea of how the average specificity evolves over different curves of arterial oxygen saturation, which can possibly reveal nonlinearities or other features that would be difficult to assess just by inspecting a few curves. Overall, regardless of the interval of sensitivities considered, GGT appears to have a good performance for high values of oxygen saturation (low z); while, conversely, for moderately low values (large z) its discriminatory ability is not so good; a slight nonlinearity is also observed (although with a large variance associated). Specifically, for $z > 2$ which corresponds, roughly, to arterial oxygen saturations below 94%-95%, the discriminatory ability is quite poor. Note further that, for all values of u considered, by ignoring the oxygen saturation effect, we would be underestimating the accuracy of GGT for women with high oxygen saturation values, and overestimating its accuracy for those who have moderately low levels of oxygen saturation. Additionally, for some levels of the covariate curves, the whole confidence interval of the no covariate-adjusted average specificity is almost completely outside of the confidence bands of

the oxygen saturation-adjusted average specificity. Similarly, to the results presented before, it can be observed that as u increases, the specificity decreases, and the decreasing is much more marked for lower levels of oxygen saturation. For instance, when $u = 0$, i.e., when the sensitivity is not constrained, the average sensitivity ranges from 0.896 $(0.776, 0.972)$ ($z = -4$, high values of oxygen saturation), to 0.518 $(0.327, 0.833)$ ($z = 2.5$, i.e., oxygen saturations around 94.5%). On the other hand, when $u = 0.95$, the average specificity varies from 0.675 $(0.359, 0.886)$ ($z = 4$) to 0.088 $(0.013, 0.459)$ ($z = 2.5$). In particular, and as noted earlier, the intervals $(0.7, 1)$ and $(0.8, 1)$, for curves of oxygen saturation around 95% or above, exhibit a good balance between gains in sensitivity and losses in specificity. We recognize that the choice of u to be used in practice is complex and our aim with this analysis is to provide an insight on the performance of the GGT, adjusted for oxygen saturation, over different intervals of sensitivity. Of course, we also recognize that our sample size is reduced and thus further studies are needed.

A sensitivity analysis has been performed to evaluate the impact of the choices of the kernel, the semimetric, and the penalizing functions in the GCV criterion; the results, not shown, do not reveal any significant differences. We have also conducted an additional analysis, by removing the outlier curves identified by the functional box plot in Figure 3; the results obtained do not show substantial changes and, hence, are not reported.

We end remarking that in the supplemental article (Inácio de Carvalho et al., 2016) we show how our method compares with simpler ones, namely, with average specificity estimators constructed using univariate kernel regression methods (González-Manteiga, Pardo-Fernández, and Van Keilegom, 2011) based on the mean and minimum arterial oxygen saturation.

5. Simulation study. We have conducted two different simulation studies. Here, we describe and report a simulation study that, to a certain extent, mimics the metabolic syndrome data. In the supplementary material, we report another simulation study, whose main purpose is to evaluate the performance of the estimator in a general setup.

5.1. *Data-generating scenarios.* We start by describing how we simulate the functional covariates, and we then present the data-generating scenarios over which we assess the finite sample performance of our methods. The simulated covariate curves mimic, to a certain extent, the covariate curves from the metabolic syndrome study in Section 4, where the distribution of the covariate curves vary by disease status. To this end we have considered

$$\mathrm{X}_{\bar{D}}(t) = \min\left\{100, \bar{X}_{\bar{D}}(t) + \textstyle\sum_{l=1}^{3} \gamma_{\bar{D}l}, v_{\bar{D}l}(t)\right\}, \quad X_D(t) = \min\left\{100, \bar{X}_D(t) + \textstyle\sum_{l=1}^{3} \gamma_{Dl} v_{Dl}(t)\right\}, \tag{5.1}$$

where $\bar{X}_D$ and $\bar{X}_{\bar{D}}$ are, respectively, the mean function from the group of women with and without metabolic syndrome [Figure 5 (a) in the supplemental article Inácio de Carvalho et al. 2016]. Additionally, v_{Dl} and $v_{\bar{D}l}$, for $l = 1, 2, 3$, are eigenfunctions from the group of women with and without metabolic syndrome and are shown in Figure 5 (b) and (c) of the supplemental article (Inácio de Carvalho et al., 2016), respectively. Lastly, $\gamma_{\bar{D}}$ and γ_D were set to

$$\gamma_{D1} \sim \mathrm{N}(0, 3^2), \quad \gamma_{D2} \sim \mathrm{N}(0, 2.25^2), \quad \gamma_{D3} \sim \mathrm{N}(0, 1.75^2),$$
$$\gamma_{\bar{D}1} \sim \mathrm{N}(0, 3^2), \quad \gamma_{\bar{D}2} \sim \mathrm{N}(0, 1.85^2), \quad \gamma_{\bar{D}3} \sim \mathrm{N}(0, 1.35^2).$$

Note that the minimum between 100 and the generated curves using the aforementioned representation was taken to reflect the fact that, in practice, oxygen saturations above 100%

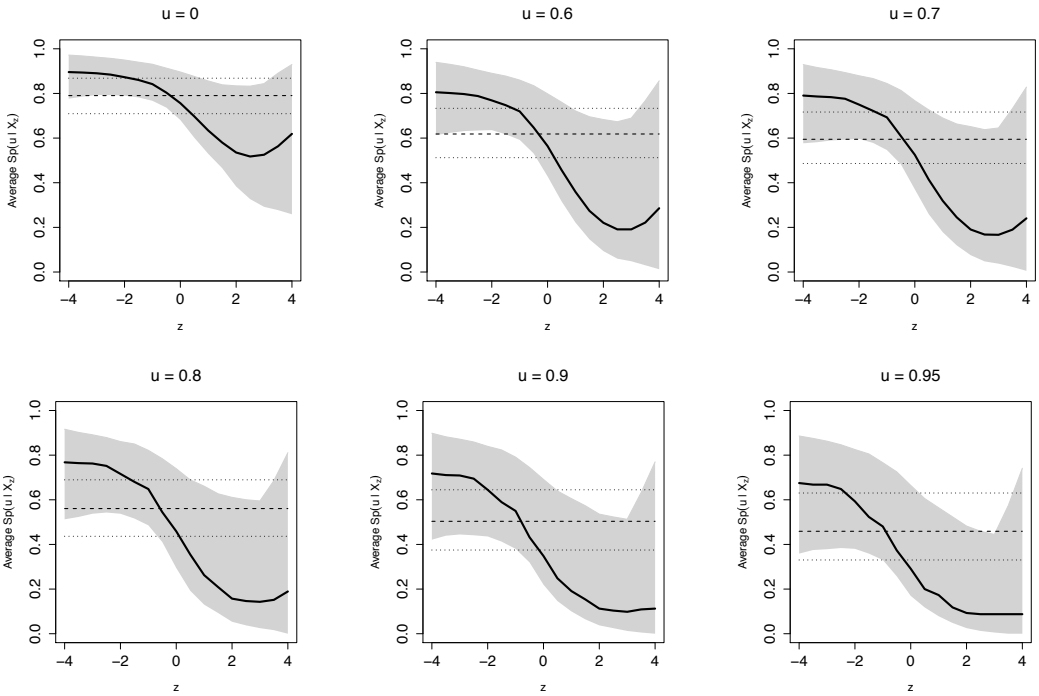


FIG 6. *Predicted average specificity (solid line), along with* 95% *bootstrap confidence bands (gray area) for GGT as a diagnostic test to detect women with metabolic syndrome, adjusted for oxygen saturation, over different intervals of TPFs, namely the intervals:* $(0, 1)$, $(0.6, 1)$, $(0.7, 1)$ *(top),* $(0.8, 1)$, $(0.9, 1)$, *and* $(0.95, 1)$ *(bottom). The no covariate-adjusted average specificity and its* 95% *bootstrap confidence interval are also represented (dashed line and dotted lines, respectively).*

cannot occur. Figure 6 of the supplemental article (Inácio de Carvalho et al., 2016) shows 100 curves of each group obtained in one simulation run.

We consider two different data-generating configurations (Scenarios A and B) and for each one, $M = 1000$ datasets were generated for each of three different sample sizes: $(n_D, n_{\bar{D}}) = (50, 100)$, $(n_D, n_{\bar{D}}) = (100, 100)$, and $(n_D, n_{\bar{D}}) = (200, 200)$. Specifically, we assume the following regression models for the marker outcome in the diseased and non-diseased groups

$$\begin{cases} Y_{\bar{D}} = 3 + 2\langle \beta, X_{\bar{D}} - 94\rangle + 2\varepsilon_{\bar{D}}, \quad Y_D = 3 + 4\langle \beta, X_D - 94\rangle + 3\varepsilon_D, & \text{(Scenario A)} \\ Y_{\bar{D}} = 1.5 + 1.5\langle \beta, \sin(X_{\bar{D}} + 1.75)\rangle + 2\varepsilon_{\bar{D}}, \quad Y_D = 3 + 2.5\langle \beta, \sin(X_D + 1.25)\rangle + 2.5\varepsilon_D. & \text{(Scenario B)} \end{cases}$$

In both cases, $\langle \beta, X_s\rangle = \int_0^1 \beta(t)X_s(t)\mathrm{d}t$, for $s \in \{D, \bar{D}\}$, $\beta(t) = t/5$, for $t \in [0, 3]$, and $\varepsilon_{\bar{D}}$ and ε_D follow the standard normal distribution. These scenarios lead to a linear (Scenario A) and to a nonlinear (Scenario B) average specificity curve. Note that

$$\begin{aligned} \mathrm{pAUC}_{\mathrm{Se}}(u \mid X) &= \int_u^1 \mathrm{ROC}_{\mathrm{Sp}}(p \mid X)\ \mathrm{d}p \\ &= \int_u^1 F_{\varepsilon_{\bar{D}}}\left(\frac{\mu_D(X) - \mu_{\bar{D}}(X)}{\sigma_{\bar{D}}(X)} + \frac{\sigma_D(X)}{\sigma_{\bar{D}}(X)} F^{-1}_{\varepsilon_D}(1-p)\right) \mathrm{d}p, \end{aligned}$$

and thus the true average specificity for Scenarios A and B are, respectively

$$\begin{cases} \text{Average Sp}(u \mid X) &= \{\int_u^1 \Phi\left(\langle \beta, X - 94\rangle + \frac{3}{2}\Phi^{-1}(1-p)\right) \mathrm{d}p\}/(1-u), \\ \text{Average Sp}(u \mid X) &= \left\{\int_u^1 \Phi\left(\frac{1.5 + 2.5\langle \beta, \sin(X+1.25)\rangle - 1.5\langle \beta, \sin(X+1.75)\rangle}{2} + \frac{2.5}{2}\Phi^{-1}(1-p)\right) \mathrm{d}p\right\}/(1-u). \end{cases} \tag{5.2}$$

Below we consider four different values of u: $0, 0.6, 0.8$, and 0.95. In addition, we consider the same grid of curves that was used in Section 4 to predict the average specificity values; see Figure 5 (a).

5.2. *Simulation results.* Our estimator was implemented using the asymmetric Gaussian kernel (3.10), the bandwidths were selected using the generalized cross validation criterion (3.11), and the semimetric d is the $L^2[0, 3]$-norm.
In Figure 7 we report the estimated Monte Carlo average of the functional covariate-adjusted average specificity from the $M = 1000$ datasets generated, against the true functional covariate-adjusted average specificity, evaluated at the grid of curves defined by (4.2). Specifically, what is represented in Figure 7 is $\{M^{-1}\sum_{m=1}^M \widehat{\text{Average}}\ \text{Sp}_m(u \mid X_z(t)) : t \in [0, 3]\}_{z\in\mathcal{Z}}$; against $\{\text{Average Sp}(u \mid X_z(t)) : t \in [0, 3]\}_{z\in\mathcal{Z}}$, where $\widehat{\text{Average}}\ \text{Sp}_m$ denotes the mth Monte Carlo estimate; for each data-generating scenario the true average specificity in (5.2) is evaluated through numerical integration. The 2.5% and 97.5% simulation quantiles are also presented. As can be seen from Figure 7 our estimator is able to recover the overall true shape of the average specificity for the two different scenarios. Here we only present the results for the sample sizes $(n_D, n_{\bar{D}}) = (50, 100)$, which are close to the metabolic syndrome data sample sizes, but the results for the remaining cases can be found in the supplemental article (Inácio de Carvalho et al., 2016). Specifically, Figures 8 and 10 of the supplemental article (Inácio de Carvalho et al., 2016) show for Scenario A and B, respectively, the true values of the functional covariate-adjusted average specificity versus the Monte Carlo averages of the estimated functional covariate-adjusted average specificities, along with the 2.5% and 97.5% simulation quantiles, for the different values of u and sample sizes considered. As can be seen from these

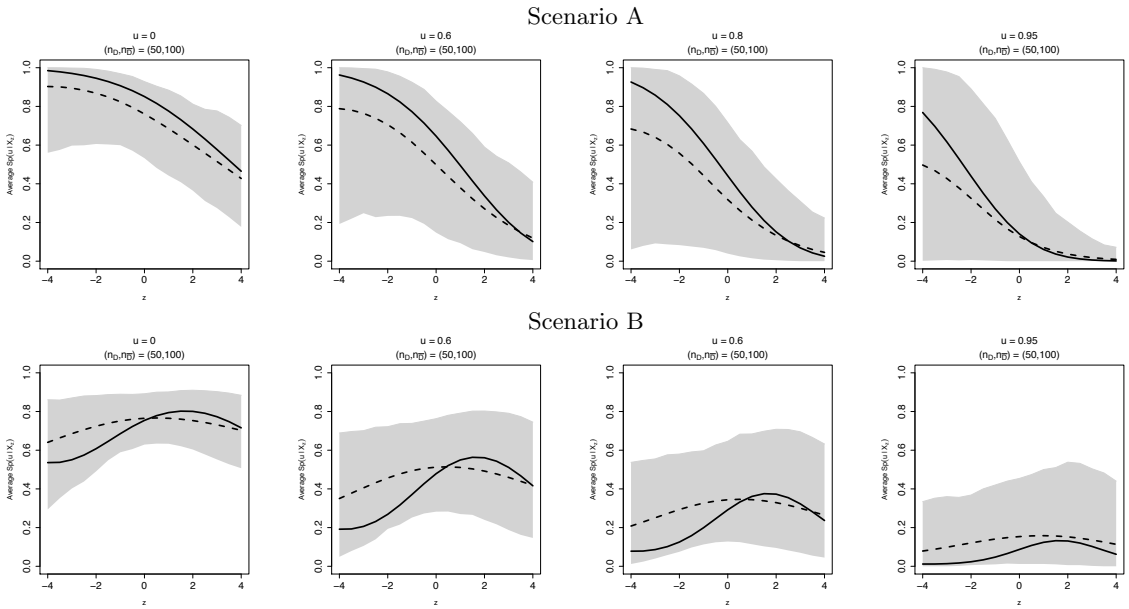


FIG 7. *True functional covariate-adjusted average specificity (solid line) versus the mean of Monte Carlo estimates (dashed line) along with* 2.5% *and* 97.5% *simulation quantiles (gray area) for* $u = 0, 0.6, 0.8,$ *and* 0.95. *For all scenarios* $(n_D, n_{\bar{D}}) = (200, 200)$.

figures, our estimator is able to recover the true functional form of the average specificity over all the different scenarios, values of u, and sample sizes considered. As expected, the variability of the estimates decreases as the sample size increases. It can also be observed that as the sample size increases and more of the curve is integrated out (i.e., as u decreases), the bias decreases; Dodd and Pepe (2003) have concluded the same, but in the case where no covariates are considered. In Figures 7 and 9 of the supplemental article (Inácio de Carvalho et al., 2016) we show the true $\mathrm{ROC}_{\mathrm{Sp}}$ curves, for Scenario A and B, along with the estimated Monte Carlo average and the 2.5% and 97.5% simulation quantiles corresponding to covariate curves associated with z values equal to -3, 0, and 3. From these figures it can be observed that our estimator is able to successfully recover the true shape of the different $\mathrm{ROC}_{\mathrm{Sp}}$ curves.

A sensitivity analysis has been performed to assess the influence of the choices of kernel and semimetric, and results, not shown, do not provide evidence of significant changes from those reported here and in the supplementary material. In addition, we have also considered different types of penalizing functions (Härdle, 1990) in the GCV criterion and results, not shown, also do not reveal significant changes.

We point out that methods are not computationally time consuming and a simple implementation can be made with the aid of routines from the R package (R Development Core Team, 2011) `fda.usc` (Febrero-Bande and Oviedo de la Fuente, 2012). In the supplemental article (Inácio de Carvalho et al., 2016) we provide R code to implement our estimator.

6. Conclusions and discussion. Motivated by a metabolic syndrome application, where we aimed to quantify how restricting the sensitivity of the GGT, a marker of metabolic syndrome, affected its corresponding specificity and how this might change for different curves of arterial oxygen saturation, we have developed an estimator for the functional covariate-

adjusted partial area under the specificity-ROC curve and, consequently, for its normalized value, the average specificity. We learned that the intervals $(0.7, 1)$ and $(0.8, 1)$ for curves of high levels of oxygen saturation (say, oxygen saturations levels above 95%) offer a good balance between sensitivity and specificity.

Simulation studies showed a good performance of the proposed estimator in recovering the true functional form of the average specificity. The variability of the estimates decreased with sample size increasing and the bias also decreased for larger sample sizes and as more of the $\mathrm{ROC}_{\mathrm{Sp}}$ curve was integrated out.

A possible extension of the work developed in this article is to the case where the average specificity is adjusted for both a functional and a scalar covariate. In this case, given a covariate curve X and a scalar covariate W, the regression models between the marker outcomes and the covariates X and W, could for instance, take the form

$$Y_D = \mu_D(X) + \beta_D W + \sigma_D(X, W)\varepsilon_D, \quad Y_{\bar{D}} = \mu_{\bar{D}}(X) + \beta_{\bar{D}} W + \sigma_{\bar{D}}(X, W)\varepsilon_{\bar{D}}, \tag{6.1}$$

and the partial AUC over the interval $(u, 1)$ of sensitivities would be given by

$$\mathrm{pAUC}_{\mathrm{Se}}(u; X, W) = \int_u^1 F_{\varepsilon_D}\left\{\frac{\mu_D(X) - \mu_{\bar{D}}(X) - W(\beta_D - \beta_{\bar{D}})}{\sigma_{\bar{D}}(X, W)} + \frac{\sigma_D(X, W)}{\sigma_{\bar{D}}(X, W)} F^{-1}_{\varepsilon_D}(1-p)\right\} \mathrm{d}p.$$

Estimation can be accomplished by

$$\begin{aligned}\widehat{\mathrm{pAUC}}_{\mathrm{Se}}(u; X, W) &= \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}}\left[\max\left\{u, \frac{1}{n_D}\sum_{j=1}^{n_D} I(\widehat{Y}_{Dj|X,W} > \widehat{Y}_{Di|X,W})\right\}\right] - u \\ &= \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}}\left[\max\left\{u, \frac{1}{n_D}\sum_{j=1}^{n_D} I(\widehat{\mu}_D(X) + \widehat{\beta}_D W + \widehat{\sigma}_D(X, W)\widehat{\varepsilon}_{Dj} > \widehat{\mu}_{\bar{D}}(X) + \beta_{\bar{D}} W + \widehat{\sigma}_{\bar{D}}(X, W)\widehat{\varepsilon}_{\bar{D}i})\right\}\right] - u.\end{aligned}$$

Details on how to estimate each quantity in (6.1) are given in Aneiros-Perez and Vieu (2005).

The proposed methods can be easily adapted to the case where the interest is to restrict the specificity to a relevant clinical interval $(1-u, 1)$, so that $1 - \mathrm{Sp}$ lies in $(0, u)$. In such case, interest lies on

$$\mathrm{pAUC}(\mathrm{u}|\ X) = \int_0^u \mathrm{ROC}(p \mid X)\mathrm{d}p = \int_0^u \left[1 - F_{\varepsilon_D}\left\{\frac{\mu_{\bar{D}}(X) - \mu_D(X)}{\sigma_D(X)} + \frac{\sigma_{\bar{D}}(X)}{\sigma_D(X)} F^{-1}_{\varepsilon_{\bar{D}}}(1-p)\right\}\right] \mathrm{d}p,$$

which can be estimated using

$$\begin{aligned}\widehat{\mathrm{pAUC}}(u \mid X) &= u - \frac{1}{n_D}\sum_{j=1}^{n_D}\min\left\{u, \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}} I(\widehat{Y}_{\bar{D}i|X} > \widehat{Y}_{Dj|X})\right\} \\ &= u - \frac{1}{n_D}\sum_{j=1}^{n_D}\min\left\{u, \frac{1}{n_{\bar{D}}}\sum_{i=1}^{n_{\bar{D}}} I(\widehat{\mu}_{\bar{D}}(X) + \widehat{\sigma}_{\bar{D}}(X)\widehat{\varepsilon}_{\bar{D}i} > \widehat{\mu}_D(X) + \widehat{\sigma}_D(X)\widehat{\varepsilon}_{Dj})\right\}.\end{aligned}$$

The normalized value, $\mathrm{pAUC}(u \mid X)/u$, can be interpreted as the average sensitivity over the interval of specificities $(1-u, 1)$. The estimators of Cai and Dodd (2008) and Wang and Chang (2011) are particular cases of this estimator when no covariates are considered. Also, when $u = 1$, this estimator corresponds to an extension to the functional covariate case of the estimator proposed by Yao, Craiu, and Reiser (2010).

Finally, we remark that the approach used in (4.1) to construct the grid of curves to graphically represent the results obtained is not unique; for instance, grids based on depth measures (López-Pintado and Romo, 2009) could be a possible alternative.

Acknowledgements. We thank the Co-Editor, the AE and the two referees, whose comments and suggestions led to an improved version of the manuscript. We also thank Francisco Gude, M.D., for making the metabolic syndrome data available and for useful recommendations and comments on the case study, and to Adam Branscum, PhD, for helpful comments. The research of the first and second authors is partially supported by the Chilean NSF through Fondecyt grants 11130541 and 11121186, respectively, and the fourth author's research is partially supported by the Spanish Ministry of Science and Innovation through project MTM2008-03010.

SUPPLEMENTARY MATERIAL

Technical details and supplementary empirical reports. The supplement consists of three parts. The first part provides auxiliary results on the construction of our estimator. The second contains supplemental empirical analysis of the metabolic syndrome data and a comparison with simpler approaches. Finally, the third part contains an additional simulation study and R code to implement our methods.

REFERENCES

Adimari, G. and Chiogna, M. (2012). Jackknife empirical likelihood based confidence intervals for partial areas under ROC curves. *Statist. Sinica* **22** 1457–1477.

Aneiros-Perez, G. and Vieu, P. (2005). Semi-functional partial linear regression. *Statist. Prob. Lett.* **76** 1102–1110.

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

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge UK: Cambridge University Press.

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

Eckel, R. H., Grundy, S. M. and Zimmet, P. Z. (2005). The metabolic syndrome. *The Lancet* **365** 1415–1428.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: The R Package `fda.usc`. *J. Statist. Soft.* **51** 1–28.

Ferraty, F., Van Keilegom, I. and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scand. J. Statist.* **37** 286–306.

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17** 545–564.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis.* New York: Springer.

Gigliarano, C., Figini, S. and Muliere, P. (2014). Making classifier performance comparisons when ROC curves intersect. *Comput. Statist. Data Anal.* **77** 300–312.

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

Gude, F., Rey-Garcia, J., Fernandez-Merino, C., Meijide, L., García-Ortiz, L., Zamarron, C. and Gonzalez-Quintela, A. (2009). Serum levels of gamma-glutamyl transferase are associated with markers of nocturnal hypoxemia in general adult population. *Clinica Chimica Acta* **407** 67–71.

Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.

Härdle, W. (1990). *Smoothing Techniques: With Implementation in S.* New York: Springer.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications.* New York: Springer.

Hung, H. and Chiang, C. T. (2011). Nonparametric methodology for the time-dependent partial area under the ROC curve. *J. Statist. Plann. Infer.* **141** 3829–3838.

Inácio, V., González-Manteiga, W., Febrero-Bande, M., Gude, F., Alonzo, T. A. and Cadarso-Suaréz, 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 Anal.* **8** 623–646.

Inácio de Carvalho, V., de Carvalho, M., Alonzo, T. A. and González-Manteiga, W. (2016). Supplement to "Functional covariate-adjusted partial area under the specificity-ROC curve with an application to metabolic syndrome diagnosis." *Ann. Appl Statist.*.

Jiang, Y., Metz, C. E. and Nishikawa, R. M. (1996). A receiver operating characteristic partial index for highly sensitive diagnostic tests. *Radiology* **201** 745–750.

Lee, D. S., Ewans, J. C., Robins, S. J., Wilson, P. W., Albano, I., Fox, C. S., Wang, T. J., Benjamin, E. J. and Vasan, R. S. (2007). Gamma glutamyl transferase and metabolic syndrome, cardiovascular disease, and mortality risk: the Framingham heart study. *Arteriosclorosis, Trombosis, and Vascular Biology* **27** 127–133.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734.

Ma, H., Bandos, A. I., Rockette, H. E. and Gur, D. (2013). On use of partial area under the ROC curve for evaluation of diagnostic performance. *Statist. Med.* **32** 344–3458.

Pardo-Fernandéz, J. C., Rodríguez-Álvarez, M. X. and Van Keilegom, I. (2014). A review on ROC curves in the presence of covariates. *Revstat—Statist. J.* **12** 21–41.

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

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* New York: Springer.

R Development Core Team (2011). R: *A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *J. Comput. Graph. Statist.* **20** 316–334.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge UK: Cambridge University Press.

Wang, Z. and Chang, Y.-C. I. (2011). Marker selection via maximizing the partial area under the ROC curve of linear risk scores. *Biostatistics* **12** 369–385.

Yao, F., Craiu, R. V. and Reiser, B. (2010). Nonparametric covariate adjustment for receiver operating characteristic curves. *Canad. J. Statist.* **38** 27–46.

Vanda Inácio de Carvalho, Miguel de Carvalho
Faculty of Mathematics
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna, 4860, Santiago, Chile
E-mail: vanda.kinets@gmail.com
E-mail: mmbbcarvalho@gmail.com

Todd Alonzo
Department of Preventive Medicine
University of Southern California
440 E. Huntington Dr. Arcadia, CA 91006, USA
E-mail: talonzo@childrensoncologygroup.org

Wenceslao González-Manteiga
Department of Statistics and Operations Research
University de Santiago de Compostela
15782 Santiago de Compostela, Spain
E-mail: wenceslao.gonzalez@usc.es