Universidade de Vigo

ROC curve and covariates: extending induced methodology to the non-parametric framework.

María X. Rodríguez-Álvarez, Javier Roca-Pardiñas and Carmen Cadarso-Suárez

Report 09/05

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://eioweb.uvigo.es/
E-mail: depc05@uvigo.es

Universidade de Vigo

ROC curve and covariates: extending induced methodology to the non-parametric framework.

María X. Rodríguez-Álvarez, Javier Roca-Pardiñas and Carmen Cadarso-Suárez

Report 09/05

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita: 

Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

ROC curve and covariates: extending induced methodology to the non-parametric framework.

María X. Rodríguez-Álvarez[1], Javier Roca-Pardiñas[2], Carmen Cadarso-Suárez[1]

[1] Unit of Biostatistics, Dept. of Statistics and OR. University of Santiago de Compostela.
[2] Dept. of Statistics and OR. University of Vigo.

Abstract

Continuous diagnostic tests are often used to discriminate between diseased and healthy populations. The receiver operating characteristic (ROC) curve is a widely used tool that provides a graphical visualisation of the effectiveness of such tests. The potential performance of the tests in terms of distinguishing diseased from healthy people may be strongly influenced by covariates, and a variety of regression methods for adjusting ROC curves has been developed. Until now, these methodologies have assumed that covariate effects have parametric forms, but in this paper we extend the induced methodology by allowing for arbitrary non-parametric effects of a continuous covariate. To this end, local polynomial kernel smoothers are used in the estimation procedure. Our method allows for covariate effect not only on the mean, but also on the variance of the diagnostic test. We also present a bootstrap-based method for testing for a significant covariate effect on the ROC curve. To illustrate the method, endocrine data were analysed with the aim of assessing the performance of anthropometry for predicting clusters of cardiovascular risk factors in an adult population in Galicia (NW Spain), duly adjusted for age. The proposed methodology has proved useful for providing age-specific thresholds for anthropometric measures in the Galician community.
Keywords ROC curve; non-parametric regression; bootstrap; cardiovascular risk factors; anthropometric measures.

1 Introduction

Appropriate statistical techniques for evaluating the accuracy of a given (continuous) diagnostic test in distinguishing between diseased and healthy subjects are based on ROC curve analysis (Metz 1978, Swets 1982, Zhou 2002b). Based on the concept of using a threshold to classify subjects as healthy or diseased, the ROC curve is a plot of the sensitivity or true-positive rate (TPR, the probability that a diseased subject has a positive test) versus 1-the specificity or false-positive rate (FPR, the probability that a healthy subject has a positive test), across all possible threshold values.

It is known, however, that diagnostic test performance in many biomedical applications may be strongly influenced by covariates. Hence, the ROC curve (or other summary indices, such as the area under the curve, AUC) may be of little value if important covariates are neglected. Characterising the impact of covariates on diagnostic test accuracy is the first step towards determining the optimal populations or conditions for the test, or alternatively, those where the test is not useful.

Furthermore, different thresholds for defining test positivity can be chosen to vary with covariates values.

To assess possible covariate effects on the ROC curve, two different regression approaches have been suggested in the statistical literature. 'Induced' methodology (Faraggi 2003, Pepe 1998, Tosteson and Begg 1988, Zheng and Heagerty 2004) is based on using separate regression models for the result of the diagnostic test in healthy and diseased populations respectively. Covariate effects on the associated ROC curve can then be computed by deriving the induced form of the ROC curve. Instead of targeting the diagnostic test, 'direct' methodology (Pepe 2003, Alonzo and Pepe 2002, Cai and Pepe 2002, Cai 2004) assumes a regression model for the ROC curve, with the effect of the covariates thus being directly evaluated on the ROC curve. Roughly speaking, both these approaches use a generalised linear model (GLM) framework to characterise covariate effects on the ROC curve.

A crucial point when applying such methodologies is how to model the continuous effect of continuous covariates on the test result or directly on the ROC curve, parametrically. Misleading conclusions may be reached if the effects are incorrectly specified. Zheng and Heagerty (2004) have proposed a semi-parametric estimator for the time-dependent ROC curve, based on induced methodology. In this approach the effect of the covariates on the result of the diagnostic test is modelled parametrically, though the authors incorporate flexibility by using regression splines (de Boor 2001). However, the use of regression splines is not altogether satisfactory, due to the need to choose the number and location of knots. This problem can be overcome by using more general flexible regression techniques, such as non-parametric regression. This technique has the advantage of not assuming a parametric form for the effects of continuous covariates, and avoids the need to impose functional assumptions. The only assumption required is that the effect of the continuous covariate follows a smooth function.

Although the literature is replete with non-parametric approaches to the estimation of the ROC curve without covariates (see e.g. the papers by Hsieh and Turnbull 1996, Lloyd 1998, Peng and Zhou 2004, Ren et al. 2004, Zhou 2002a, Zou et al. 1997), in the case of the conditional ROC curve very few contributions have been made so far in the non-parametric framework. Recently, López-de-Ullibarri et al. (2008) studied the effect of covariates on the ROC curve, using a non-parametric approach based on local linear estimation of conditional survival functions in healthy and diseased subjects. In the induced regression context, a technical report by González-Manteiga et al. (2008) proposed a new non-parametric estimator of the conditional ROC curve based on kernel-type regression techniques. In these papers, however, the authors do no propose inference procedures to examine the effects of covariates on the ROC curve.

In this paper, we extend the induced methodology to the non-parametric framework in a similar fashion of that proposed by González-Manteiga et al. (2008). Our methodology allows for: (a) arbitrary, non-parametric effects of a continuous covariate on the mean and variance of the diagnostic test result and thus, indirectly, on the ROC curve; and (b) non-distributional assumptions about the diagnostic test. The aims of this study are twofold. Firstly, we present an estimator for the covariate-specific ROC curve, based on the use of local polynomial kernel smoothers (Fan and Gijbels 1996, Wand and Jones 1995). Among the advantages of using such smoothers is the possible use of binning-type acceleration techniques (Fan and Marron 1994) to reduce computational time, thereby rendering the methodology presented in this paper feasible for practical use. Secondly, in the context of ROC regression it is important to be able to assess the presence of significant covariate effects. Accordingly, we propose a bootstrap method to test for the possible effect of the continuous covariate on the ROC curve. The test is based on the comparison of the

estimators of the covariate-specific ROC curve and the covariate-adjusted ROC curve, as defined by Janes and Pepe (2009).

The layout of this paper is as follows. Section 2 introduces the induced covariate-specific ROC curve in the non-parametric framework, with subsection 2.1 being devoted to the estimation procedure. The bootstrap procedure to test for the possible effect of a continuous covariate on the covariate-specific ROC curve is presented in Section 3. In Section 4, the performance of the estimation procedure and the bootstrap method for detecting covariate effect are evaluated by means of simulations. We illustrate our method in Section 5, using data from the endocrine field, and conclude with a discussion in Section 6. Some technical details have been added by way of an appendix.

2 Non-parametric ROC regression analysis

Induced ROC methodology is based on specifying a model for the test result as a function of covariates, in both healthy and diseased populations. From these models, the induced covariate-specific ROC curve is then computed.

In this study, a non-parametric regression model is assumed for the test result Y, in healthy ($\bar{D}$) and disease (D), such that

$$Y_{\bar{D}} = \mu_{\bar{D}}(X) + \sigma_{\bar{D}}(X)\varepsilon_{\bar{D}}, \tag{1}$$

$$Y_D = \mu_D(X) + \sigma_D(X)\varepsilon_D, \tag{2}$$

where X is a continuous covariate, $\mu_{\bar{D}}$ and μ_D are the regression functions, and $\sigma^2_{\bar{D}}$ and σ^2_D are the variance functions. The errors $\varepsilon_{\bar{D}}$ and ε_D are assumed to be independent of the covariate X, with zero mean, unit variance, and survival functions $S_{\bar{D}}$ and S_D respectively. Within this context, it is easy to show that the induced covariate-specific ROC curve can be expressed as

$$ROC_X(t) = S_D\left(\frac{\mu_{\bar{D}}(X) - \mu_D(X)}{\sigma_D(X)} + \frac{\sigma_{\bar{D}}(X)}{\sigma_D(X)} S^{-1}_{\bar{D}}(t)\right), t \in (0,1), \tag{3}$$

where $S^{-1}_{\bar{D}}(t)$=sup $\{z; S_{\bar{D}}(z) \geq t\}$.

By way of particular cases, this formulation includes other models previously addressed in the literature. Faraggi (2003) assumes a parametric regression model for Y, with homoscedastic variance and normal error, in both healthy and diseased populations. Pepe (1998) relaxes the distributional assumptions, by not assuming a known probability distribution for the test result, though the same distribution is considered for both populations. The advantages of using the ROC formulation based on the model given by equations (1) and (2), include the possibility of incorporating: (a) a non-parametric effect of the continuous covariate; (b) a possible interaction between the continuous covariate and the false-positive fraction t, achieved through heteroscedasticity; and (c) different probability distributions for the test result in healthy and diseased populations.

2.1 Estimation procedure

In this section, we present the estimation procedure for the covariate-specific ROC curve given in (3) above. Our method is based on firstly estimating the regression and variance functions of (1) and (2), using local polynomial kernel smoothers, and then estimating the survival functions $S_{\bar{D}}$ and S_D by the empirical survival distribution of the standardised residuals. More specifically, let

$\left\{\left(x_i^{\bar{D}}, y_i^{\bar{D}}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^D, y_j^D\right)\right\}_{j=1}^{n_D}$ be two independent random samples drawn from the healthy and diseased populations respectively. The proposed procedure for estimating $ROC_X(t)$ is as follows:

Step 1. Estimate the regression functions $\mu_{\bar{D}}$ and μ_D as

$$\hat{\mu}_{\bar{D}}(x) = \hat{\psi}\left(x, \left\{\left(x_i^{\bar{D}}, y_i^{\bar{D}}\right)\right\}_{i=1}^{n_{\bar{D}}}, h_{\bar{D}}, p_{\bar{D}}\right), \hat{\mu}_D(x) \quad = \hat{\psi}\left(x, \left\{\left(x_j^D, y_j^D\right)\right\}_{j=1}^{n_D}, h_D, p_D\right), \quad (4)$$

where $\hat{\psi}$ is the local polynomial kernel estimator (see Appendix for details), $h_{\bar{D}}$ and h_D are the smoothing parameters or bandwidths, and $p_{\bar{D}}$ and p_D are the order of the polynomials, in the healthy and disease populations respectively.

Step 2. Estimate the variance functions $\sigma_{\bar{D}}^2$ and σ_D^2 in a similar fashion

$$\hat{\sigma}_{\bar{D}}^2(x) = \hat{\psi}\left(x, \left\{\left(x_i^{\bar{D}}, z_i^{\bar{D}}\right)\right\}_{i=1}^{n_{\bar{D}}}, g_{\bar{D}}, q_{\bar{D}}\right), \hat{\sigma}_D^2(x) \quad = \hat{\psi}\left(x, \left\{\left(x_j^D, z_j^D\right)\right\}_{j=1}^{n_D}, g_D, q_D\right), \quad (5)$$

where $z_i^{\bar{D}} = \left(y_i^{\bar{D}} - \hat{\mu}_{\bar{D}}\left(x_i^{\bar{D}}\right)\right)^2$, $z_j^D = \left(y_j^D - \hat{\mu}_D\left(x_j^D\right)\right)^2$, $g_{\bar{D}}$, and g_D are the bandwidths and $q_{\bar{D}}$ and q_D are the order of the polynomials.

Step 3. Estimate the survival functions $S_{\bar{D}}$ and S_D by the empirical distribution of the standardised residuals

$$\hat{S}_{\bar{D}}(z) = \frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} I\left[\frac{y_i^{\bar{D}} - \hat{\mu}_{\bar{D}}(x_i^{\bar{D}})}{\hat{\sigma}_{\bar{D}}(x_i^{\bar{D}})} \geq z\right], \quad (6)$$

$$\hat{S}_D(z) = \frac{1}{n_D} \sum_{j=1}^{n_D} I\left[\frac{y_j^D - \hat{\mu}_D(x_j^D)}{\hat{\sigma}_D(x_j^D)} \geq z\right]. \quad (7)$$

Step 4. Finally, compute the covariate-specific ROC curve (3) as follows

$$\widehat{ROC}_{X=x}(t) = \hat{S}_D\left(\frac{\hat{\mu}_{\bar{D}}(x) - \hat{\mu}_D(x)}{\hat{\sigma}_D(x)} + \frac{\hat{\sigma}_{\bar{D}}(x)}{\hat{\sigma}_D(x)} \hat{S}_{\bar{D}}^{-1}(t)\right),$$

where $\hat{S}_{\bar{D}}^{-1}(t) = \sup\left\{z; \hat{S}_{\bar{D}}(z) \geq t\right\}$.

To implement the above estimation procedure, the order of the polynomials in Steps 1 and 2 must be chosen. In the context of polynomial kernel smoothers, theoretical results (Wand and Jones 1995) have shown that odd-order polynomials possess attractive bias and boundary properties. Moreover, the asymptotic performance of the resulting estimates improves as the order increases. However, it is also well known (Wand and Jones 1995) that, in practice, the benefits of using higher degree polynomials are not clear because the variance of the estimator increases as p increases, and, in many situations, a very large sample is required for there to be a substantial improvement in practical performance. In this study, we therefore used $p_D = p_{\bar{D}} = 1$ to estimate the regression functions. Nevertheless, when estimating the variance functions it is necessary to ensure that the resulting estimates are positive. Accordingly, our proposal for the variance estimation is to use $q_D = q_{\bar{D}} = 0$. Note that, in this case, the local polynomial kernel estimator becomes the Nadaraya-Watson estimator (Nadaraya 1964, Watson 1964).

In the context of kernel estimation, it is well known that the estimates obtained depend heavily on the smoothing parameter or bandwidth. The bandwidth controls the trade-off between the bias and the variance of the resulting estimates. In this study, cross-validation was used for the automatic choice of bandwidths (see Appendix). It should be noted here that the optimal bandwidths obtained for the regression functions in Step 1 were used for the purpose of estimating the variance functions in Step 2. Cross-validation implies a high computational cost, since it is necessary to repeat the estimation procedure several times in order to select the optimal bandwidth. To speed up this process, we used binning-type acceleration techniques (Fan and Marron 1994) to obtain the binning approximations of $\hat{\mu}_{\bar{D}}$, $\hat{\mu}_D$, $\hat{\sigma}^2_{\bar{D}}$ and $\hat{\sigma}^2_D$ (see Appendix for details).

3 Testing for continuous covariate effect

In this section, we introduce a bootstrap procedure (Efron 1993) to test for the possible effect of the continuous covariate on the covariate-specific ROC curve shown in (3). For a given continuous covariate X, our interest is focused on the null hypothesis

$$H_0 : ROC_X(t) = ROC(t),$$

namely, that the covariate-specific ROC curve is not affected by X. It is important to note that this does not necessarily imply that the diagnostic test Y is not affected by covariate X. For instance, under our ROC model (3), the null hypothesis H_0 holds if the following conditions apply

$$\frac{\mu_{\bar{D}}(X) - \mu_D(X)}{\sigma_D(X)} = \alpha \text{ and } \frac{\sigma_D(X)}{\sigma_{\bar{D}}(X)} = \beta, \tag{8}$$

for some constants α and β. Hence, under H_0, the common covariate-specific ROC curve $ROC(t)$, does not necessarily coincide with the ROC curve obtained on pooling healthy and diseased individuals, regardless of the value of X, and the statistic to test for the null hypothesis must be based on other quantities.

Before defining the test, some definitions must be introduced. Let F_{X_D} the cumulative distribution function of X on the disease population and $S_{\bar{D}X} = P[Y_{\bar{D}} > y|X]$ the survival function of Y in the healthy population, conditional on X. The covariate-adjusted ROC curve (Janes and Pepe 2009), denoted by $AROC$, is defined as

$$\begin{aligned} AROC(t) &= P\left[Y_D > S^{-1}_{\bar{D}X_D}(t)\right] \\ &= \int ROC_X(t) dF_{X_D}(X), t \in (0,1). \end{aligned} \tag{9}$$

Thus, the $AROC$ can be seen as a weighted average of covariate-specific ROC curves. Assuming that the regression model (1) holds, the $AROC$ (9) can be estimated by (Janes and Pepe 2009)

$$\widehat{AROC}(t) = \frac{1}{n_D} \sum_{j=1}^{n_D} I\left[\frac{y^D_j - \hat{\mu}_{\bar{D}}(x^D_j)}{\hat{\sigma}_{\bar{D}}(x^D_j)} > \hat{S}^{-1}_{\bar{D}}(t)\right]. \tag{10}$$

For convenience, we will denote the global sample as $\{(x_k, y_k, d_k)\}_{k=1}^{n_{\bar{D}}+n_D}$, were d is a binary indicator, taking the value 1 for disased individuals and 0 for healthy. Under the null hypothesis,

$H_0 : ROC_X(t) = ROC(t)$, the $AROC$ is the common covariate-specific ROC cuve $ROC(t)$. As a result, to test the null hypothesis

$$ROC_X(t) = ROC(t) = S_D\left(a + bS_{\bar{D}}^{-1}(t)\right),$$

versus the general hypothesis

$$ROC_X(t) = S_D\left(\frac{\mu_{\bar{D}}(X) - \mu_D(X)}{\sigma_D(X)} + \frac{\sigma_{\bar{D}}(X)}{\sigma_D(X)} S_{\bar{D}}^{-1}(t)\right),$$

the following test statistic is considered

$$T = \frac{1}{n_{\bar{D}} + n_D} \sum_{k=1}^{n_{\bar{D}}+n_D} \int |\widehat{ROC}_{X=x_k}(t) - \widehat{AROC}(t)| dt,$$

where $\widehat{ROC}_X(t)$ is the estimation of $ROC_X(t)$ obtained in the Step 4 of the estimation procedure (see Section 2 above), and $\widehat{AROC}(t)$ is the estimation of $AROC(t)$ given in (10).

Note that if the null hypothesis is verified, then T should be close to zero but will be positive. Thus, the test rule for checking $H = H_0$ with a significance level of α is that the null hypothesis is rejected if T is larger than its upper α-percentile. The bootstrap-based testing procedure consists of the following steps:

Step 1. Estimate (1) and (2) under the null hypothesis and the corresponding empirical survival functions of the estandardised residuals. Let $\hat{\mu}_{\bar{D}}^0$, $\hat{\mu}_D^0$, $\hat{\sigma}_{\bar{D}}^0$, $\hat{\sigma}_D^0$, $\hat{S}_{\bar{D}}^0$ and $\hat{S}_D^0$ be these estimates.

For $b = 1, \ldots, B$

Step 2. Generate the bootstrap resamples $\left\{\left(x_i^{\bar{D}}, y_{i,b}^{\bar{D}*}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^D, y_{j,b}^{D*}\right)\right\}_{j=1}^{n_D}$ where

$$y_{i,b}^{\bar{D}*} = \hat{\mu}_{\bar{D}}^0\left(x_i^{\bar{D}}\right) + \hat{\sigma}_{\bar{D}}^0\left(x_i^{\bar{D}}\right)\varepsilon_{i,b}^{\bar{D}*},$$

$$y_{j,b}^{D*} = \hat{\mu}_D^0\left(x_j^D\right) + \hat{\sigma}_D^0\left(x_j^D\right)\varepsilon_{j,b}^{D*},$$

and $\left\{\varepsilon_{i,b}^{\bar{D}*}\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\varepsilon_{j,b}^{D*}\right\}_{j=1}^{n_D}$ are two independently and identically distributed (i.i.d) samples from the distributions $\hat{S}_{\bar{D}}^0$ and $\hat{S}_D^0$ respectively.

Step 3. From $\left\{\left(x_i^{\bar{D}}, y_{i,b}^{\bar{D}*}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^D, y_{j,b}^{D*}\right)\right\}_{j=1}^{n_D}$ obtain T_b.

Since the the bootstrap resamples are constructed under the null hypothesis, this procedure approximates the distribution of T under H_0. Consequently, the test rule based on T consists of rejecting the null hypothesis if $T > T_\alpha$ where T_α is the empirical (1-α)-percentile of the values of $T_1, \ldots, T_B$ obtained in Step 3.

In Step 1 of the above testing procedure, (1) and (2) must be estimated under the null hypothesis. As seen in the introduction to this section, the conditions (8) must be met if the null hypothesis H_0 is true. For practical reasons, these conditions can be regarded as

$$\sigma_D(X) = \beta\sigma_{\bar{D}}(X) \text{ and } \mu_{\bar{D}}(X) - \mu_D(X) = \alpha\sigma_D(X).$$

Thus, in order to obtain $\hat{\mu}_{\bar{D}}^0$, $\hat{\mu}_D^0$, $\hat{\sigma}_{\bar{D}}^0$, $\hat{\sigma}_D^0$, $\hat{S}_{\bar{D}}^0$ and $\hat{S}_D^0$, we propose the following procedure

1. estimate $\mu_{\bar{D}}$, μ_D, $\sigma_{\bar{D}}$, σ_D from $\left\{\left(x_i^{\bar{D}}, y_i^{\bar{D}}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^D, y_j^D\right)\right\}_{j=1}^{n_D}$ as explained in Section 2 above;

2. obtain the null variances estimates $\hat{\sigma}_{\bar{D}}^0(X) = \hat{\sigma}_{\bar{D}}(X)$ and $\hat{\sigma}_D^0(X) = \hat{\beta}\hat{\sigma}_{\bar{D}}(X)$, where $\hat{\beta}$ is obtained by using ordinary least squares to fit a linear model (without the intercept) of $\hat{\sigma}_{\bar{D}}(x_j^D)$ on $\hat{\sigma}_D(x_j^D)$;

3. obtain the null regression functions $\hat{\mu}_{\bar{D}}^0(X) = \hat{\mu}_{\bar{D}}(X)$ and $\hat{\mu}_D^0(X) = \hat{\mu}_{\bar{D}}(X) + \hat{\alpha}\hat{\sigma}_D^0(X)$, where $\hat{\alpha}$ is obtained by using ordinary least squares to fit a linear model (without the intercept) of $\hat{\sigma}_D^0\left(x_j^D\right)$ on $\hat{\mu}_{\bar{D}}\left(x_j^D\right) - \hat{\mu}_D\left(x_j^D\right)$; and,

4. finally, estimate the survival functions $S_{\bar{D}}^0$ and S_D^0 by using the empirical distribution of the standardised residuals

$$\hat{S}_{\bar{D}}^0(z) = \frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} I\left[\frac{y_i^{\bar{D}} - \hat{\mu}_{\bar{D}}^0(x_i^{\bar{D}})}{\hat{\sigma}_{\bar{D}}^0(x_i^{\bar{D}})} \geq z\right],$$

$$\hat{S}_D^0(z) = \frac{1}{n_D} \sum_{j=1}^{n_D} I\left[\frac{y_j^D - \hat{\mu}_D^0(x_j^D)}{\hat{\sigma}_D^0(x_j^D)} \geq z\right].$$

4 Simulation study

This section reports the results of a simulation study conducted to evaluate the practical performance of the proposed method. We began by studying the performance of: (a) the estimation procedure described in Section 2 above; and, (b) the bootstrap method for detecting covariate effect set out in Section 3 above. Since the proposed model does not require a parametric form to be specified for the effects of the covariates, this model can be expected to be more robust than other parametric or semi-parametric approaches previously addressed in the literature. It is also true that, in some circumstances, the flexibility afforded by our model may not be necessary, on, say, the effects of the covariates proving to be linear. In order to examine both aspects, this section ends with a simulation study in which the proposed model is compared to the induced semi-parametric approach proposed by Pepe (Pepe 1998) and Cai's direct semi-parametric approach (Cai 2004). In view of the fact that the comparative study has been conducted with other models (Faraggi 2003, Alonzo and Pepe 2002) and yielded similar results, it is not shown in this study. In the case of Pepe's model (1998), different distributions for the errors of the regression models in healthy and diseased populations were considered.

To study the practical performance of the proposed method, the following models were considered for the test result in healthy and diseased populations:

$$Y_{\bar{D}} = \sin(\pi X) - a0.3X^3 + \sqrt{0.2 + 0.5\exp(X)}\varepsilon_{\bar{D}},$$
$$Y_D = \sin(\pi X) + \sqrt{0.2 + 0.5\exp(X)} + \sqrt{0.2 + 0.5\exp(X)}\varepsilon_D,$$

where a is a real constant, X has uniform distribution on $[-1, 1]$, and $\varepsilon_{\bar{D}}$ and ε_D have the standard normal distribution. With this configuration, the corresponding covariate-specific $ROC_X(t)$ is

$$ROC_X(t) = \Phi\left(\frac{a0.3X^3 + \sqrt{0.2 + 0.5\exp(X)}}{\sqrt{0.2 + 0.5\exp(X)}} + \Phi^{-1}(t)\right),$$

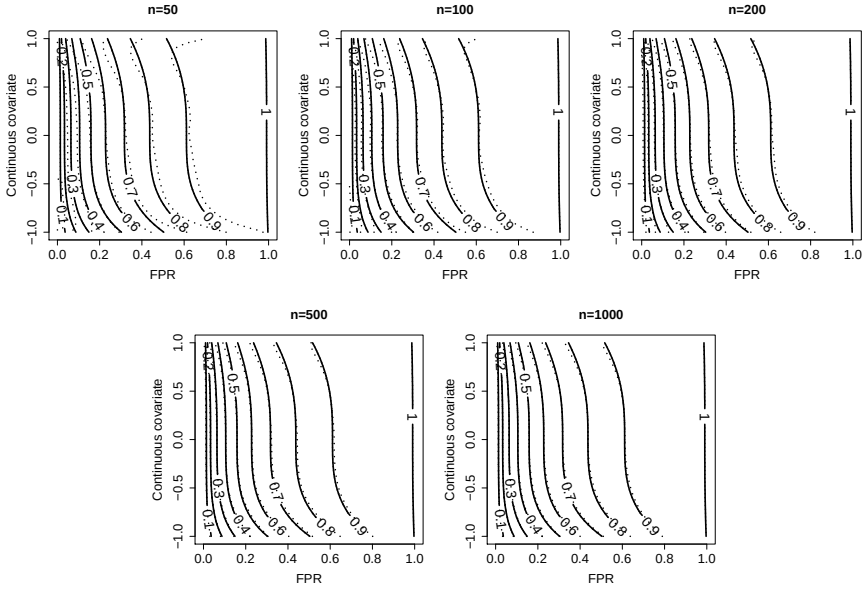

Figure 1: Contour plot for the true ROC (solid line) versus the average of simulated ROCs (dashed line) for different samples sizes ($n = n_D = n_{\bar{D}}$), based on 1000 estimates.

where Φ is the CDF of a standard normal variable.

To study the size and power of the proposed test, different values were considered for a, ranging from 0 to 2. It should be noted that the value $a = 0$ corresponds to the null hypothesis, i.e., the covariate-specific ROC curve, $ROC_X(t) = \Phi\left(1 + \Phi^{-1}(t)\right)$, is not affected by X, and the more the constant a shifts towards zero, the greater the effect of the covariate on the ROC curve.

With respect to the study of the behaviour of the estimation process, this section presents the results for the value $a = 1$, based on 1000 repetitions. In all cases, the same sample size was considered in healthy and diseased subjects, with $n = n_D = n_{\bar{D}} = 50, 100, 200, 500, 1000$. Figure 1 depicts the true ROC curve (solid line) and the corresponding average of the estimated ROCs (dashed line) for the different sample sizes. The estimated standard deviation is shown in Figure 2. As can be seen, the estimator behaves well for all sample sizes, with the variance decreasing as sample size increases. The proposed estimator displays a worse behaviour at the boundaries of the continuous covariate, probably inherited from the well-known poor performance of kernel-type estimators at the boundaries. All the above can be clearly seen in Figure 3, where the average of the estimated AUC is shown, along with 2.5 and 97.5 simulation quantiles, for the different sample sizes.

Insofar as the issue of testing was concerned, we considered the null hypothesis $H_0 : ROC_X(t) = ROC(t)$ (or equivalently $a = 0$). The bootstrap procedure described in Section 3 was applied, specifically using $B = 400$ bootstrap samples to calculate type I error and $B = 200$ bootstrap samples to calculate the power under the alternative. Both type I error and power were calculated on the basis of 1000 simulation runs.

The type I error registered by the test for different significance levels and sample sizes is shown

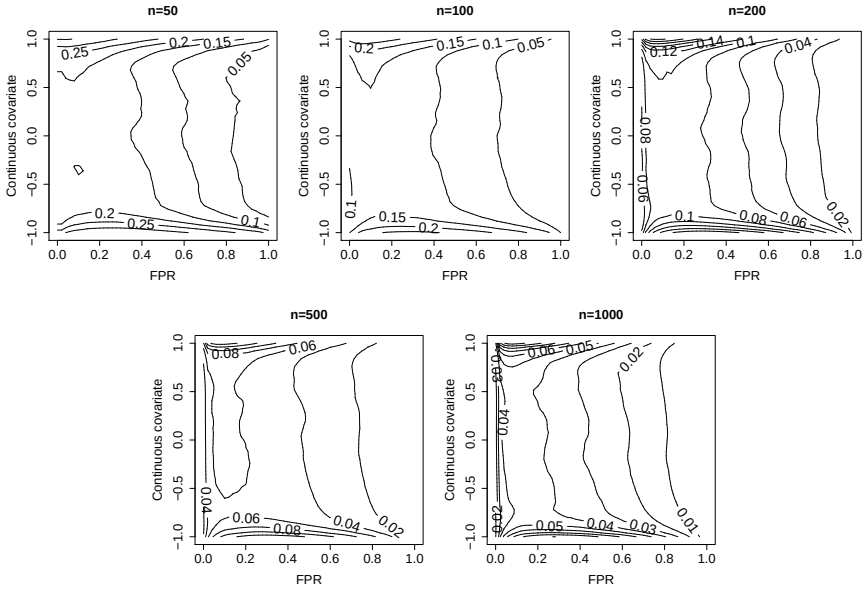

Figure 2: Contour plot for standard deviation of simulated ROCs for different samples sizes ($n = n_D = n_{\bar{D}}$), based on 1000 estimates

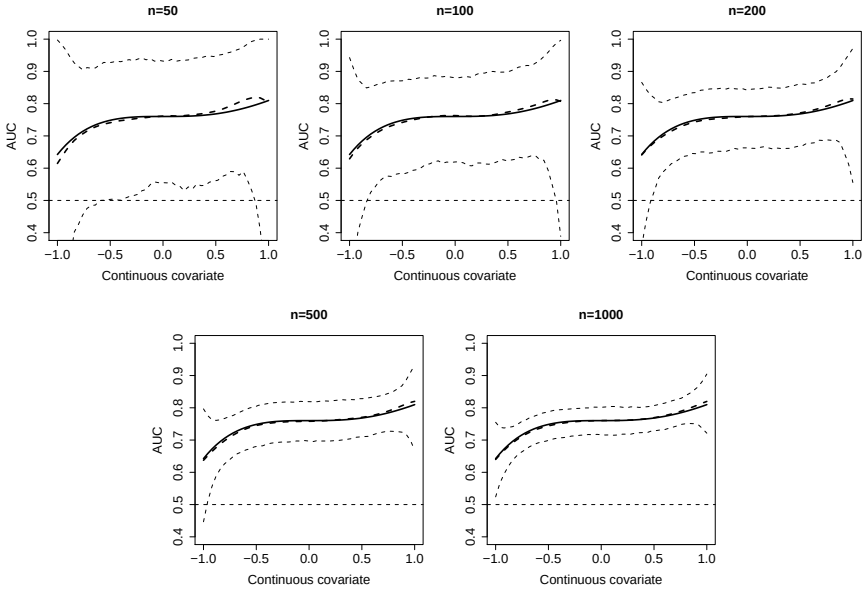

Figure 3: True AUC (solid line) versus the average of simulated AUCs (dashed line), along with 2.5 and 97.5 simulation quantiles for different samples sizes ($n = n_D = n_{\bar{D}}$), based on 1000 estimates.

Table 1: Type I error of the test under the null hypothesis

	Level				
$n_D = n_{\bar{D}}$	**0.01**	**0.05**	**0.10**	**0.15**	**0.20**
50	0.006	0.034	0.084	0.120	0.174
100	0.012	0.055	0.100	0.147	0.197
200	0.007	0.047	0.101	0.162	0.206
500	0.011	0.038	0.094	0.147	0.201
1000	0.012	0.047	0.104	0.155	0.201

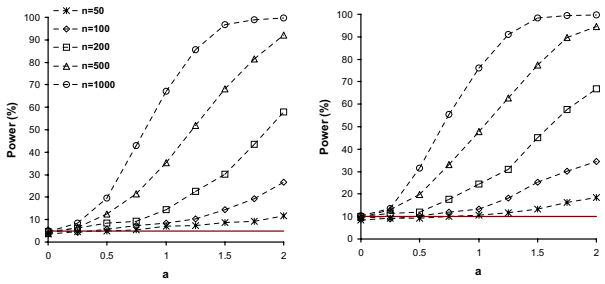

Figure 4: Rejection probability of the proposed test as a function of the parameter a at significance levels 0.05 and 0.1, for different sample sizes ($n = n_D = n_{\bar{D}}$).

in Table 1. As can be seen, the test performed well in general, with type I errors proving to be relatively close to nominal errors. Some test performance results in terms of power are shown in Table 2 and Figure 4. As expected, the probability of rejection rose with the increase in the constant a and sample size.

We now show the results of the simulation study comparing the proposed model and the semi-parametric approaches by Pepe (1998) and Cai (2004). Two different simulation scenarios were considered, namely: (a) a linear scenario; and (b) a non-linear scenario, which, from an applied standpoint, appears plausible in the light of the results obtained in the data analysis (see section 5). The scenarios considered were as follows:

- Scenario I

$$Y_D = 2 + 4X + 2\varepsilon_D,$$
$$Y_{\bar{D}} = 0.5 + X + 1.5\varepsilon_{\bar{D}}.$$

- Scenario II

$$Y_D = 0.5 + X^2 + \varepsilon_D,$$
$$Y_{\bar{D}} = \sin(\pi(X+1)) + 0.5\varepsilon_{\bar{D}}.$$

Table 2: Power of the test

a	$n_D = n_{\bar{D}}$	Level **0.01**	**0.05**	**0.10**	**0.15**	**0.20**
0.5	**50**	0.017	0.057	0.110	0.179	0.236
	100	0.025	0.071	0.123	0.180	0.233
	200	0.027	0.078	0.141	0.217	0.289
	500	0.059	0.157	0.247	0.316	0.375
	1000	0.092	0.249	0.354	0.453	0.518
1.0	**50**	0.019	0.074	0.142	0.215	0.264
	100	0.044	0.110	0.177	0.252	0.309
	200	0.069	0.182	0.280	0.355	0.414
	500	0.238	0.435	0.536	0.613	0.668
	1000	0.521	0.724	0.818	0.879	0.909
1.5	**50**	0.032	0.110	0.176	0.257	0.319
	100	0.077	0.183	0.282	0.374	0.450
	200	0.164	0.348	0.458	0.557	0.630
	500	0.557	0.744	0.822	0.877	0.906
	1000	0.928	0.979	0.992	0.996	0.998
2.0	**50**	0.059	0.160	0.238	0.315	0.384
	100	0.141	0.307	0.413	0.530	0.589
	200	0.344	0.566	0.680	0.758	0.808
	500	0.851	0.946	0.968	0.986	0.989
	1000	0.998	1.000	1.000	1.000	1.000

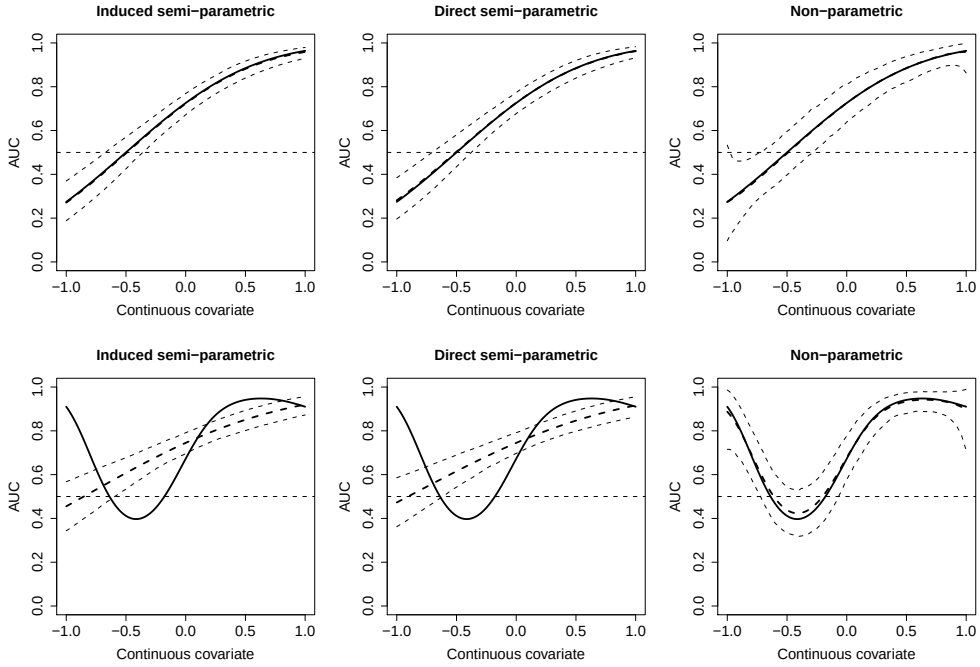

Figure 5: True AUC (solid line) versus the average of simulated AUCs (dashed line), along with 2.5 and 97.5 simulation quantiles, for $n = 200$, based on 1000 estimates. From left to right: Pepe's induced semi-parametric approach, Cai's direct semi-parametric approach, and proposed non-parametric approach. Top row: Scenario I. Botton row: Scenario II.

With these configurations, the corresponding covariate-specific ROC curves are

$$ROC_X(t) = \Phi\left(\frac{1}{2}\left(1.5 + 3X + 1.5\Phi^{-1}(t)\right)\right),$$
$$ROC_X(t) = \Phi\left(0.5 + X^2 - \sin\left(\pi(X+1)\right) + 0.5\Phi^{-1}(t)\right),$$

for Scenario I and II respectively. In both scenarios, X has uniform distribution on $[-1, 1]$, and $\varepsilon_{\bar{D}}$ and ε_D have the standard normal distribution.

As in the previous simulation study, the results shown are based on 1000 repetitions. The same sample size was considered for healthy and diseased subjects alike, with $n = n_D = n_{\bar{D}} = 200$. Figure 5 depicts the average of the estimated AUC, along with 2.5 and 97.5 simulation quantiles.

As would be expected under Scenario I, the models that displayed the lowest variance were the semi-parametric approaches, though our method also performed satisfactorily (see Figure 5). For Scenario II, however, the effect of the covariate on the test result was far from linear, and the estimates obtained by the semi-parametric models were thus not suitable, as can be seen from Figure 5. In contrast, the good performance of the proposed model is evident, with it recovering the functional form of the corresponding true AUC very successfully.

5 Detecting cardiovascular risk factors by means of anthropometry

Atherosclerotic diseases are the leading cause of death in adults world-wide (Peden et al. 2002). The prevalence of these diseases is directly associated with the existence of different cardiovascular disease (CVD) risk factors, most of which tend to occur in clusters. In such clustering (e.g., metabolic syndrome) obesity is a common feature (Karelis et al. 2004) and is acknowledged as being the sixth leading risk factor contributing to the world's overall disease burden (Haslam et al. 2005). Obesity contributes to hypertension, high serum cholesterol, low HDL-c and hyperglycaemia, and is independently associated with higher CVD risk (Hu et al. 2004, Zimmet et al. 2001, Carey et al. 1997).

We applied the ROC methodology proposed in this paper to an endocrine study (Tomé et al. 2009), with the aim of using anthropometric measures (as body mass index, BMI; waist circumference, WC; and waist to hip ratio, WHR) to detect patients having a higher risk of cardiovascular problems, and ascertaining the possible effect of age on the accuracy of these measures. It is well established that anthropometric measures perform differently according to gender, and so separate analyses were conducted on men and women respectively.

To date, biomedical studies that link anthropometric indices to cardiovascular risk factors have tended to incorporate age categorically (Mirmiran et al. 2004). Though easy to interpret, this option is not entirely appropriate: categorising age, not only results in a possible loss of statistical power, but also raises the problem of how many categories should be used and where cut-off points should be located. Moreover, using this procedure to detect fine changes in the accuracy of anthropometric measures along age is a difficult task. In this paper, the effect of age is incorporated continuously, meaning that the researcher avoids having to select the cut-off values of the categories, as well as the above-mentioned problems.

Apart from the need for a study to evaluate the accuracy of a given anthropometric measure in predicting CVD risk factors, it is also important to study whether age-specific thresholds should be provided in order to enable effective screening. It should be pointed out here that, when a diagnostic test's discriminatory capacity is not affected by a covariate: (a) the common covariate-specific ROC curve, or equivalently the covariate-adjusted ROC curve, $AROC$, should be used to choose the optimal sensitivity and specificity pairing; and, (b) this does not necessarily mean that the threshold value for which optimal operational characteristics are attained will not vary with covariate. In all cases where a covariate affects the test, even though it might not affect its discriminatory capacity, covariate-specific threshold values should be provided (for a more detailed review of this topic, readers are urged to consult reference Janes and Pepe 2009).

5.1 Data source

The study was conducted using a random sample, representative of the Galician adult population (2850 subjects, age range 18-85 years). Diseased subjects were defined as those presenting with two or more CVD risk factors (raised triglycerides, reduced HDL-cholesterol, raised blood pressure and raised fasting plasma glucose) as per the International Diabetes Federation criteria [24]. Direct anthropometric indices were measured, including weight (in kilograms), height (in metres), WC (in centimeters) and hip circumference (in centimeters), deriving the WHR. BMI was calculated as weight divided by height squared. Of the total of 2850 subjects, 46.2% were men (899 healthy and 418 diseased) and 53.8% women (1250 healthy and 273 diseased). A detailed description of this

dataset can be found in (Tomé et al. 2009).

For the sake of illustration, in the following we will show only the statistical analysis conducted with BMI.

5.2 Statistical analysis

By taking BMI as diagnostic test, separate covariate-specific ROC curves were fitted for men and women respectively, assuming the following non-parametric models for healthy and diseased populations:

$$BMI_{\bar{D}} = \mu_{\bar{D}}\left(AGE\right) + \sigma_{\bar{D}}\left(AGE\right)\epsilon_{\bar{D}},$$
$$BMI_{D} = \mu_{D}\left(AGE\right) + \sigma_{D}\left(AGE\right)\epsilon_{D},$$

In addition to the estimated covariate-specific ROC curves, another summary measure of accuracy, the AUC, was obtained. Bootstrap-based methods were used for constructing confidence intervals for this measure (Efron 1993). Given a point x in the range of X, the steps for the construction of the confidence interval, based on the samples $\left\{\left(x_i^{\bar{D}}, y_i^{\bar{D}}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^{D}, y_j^{D}\right)\right\}_{j=1}^{n_D}$, were as follows:

1. for $b = 1, \ldots, 500$, generate the bootstrap resamples $\left\{\left(x_i^{\bar{D}}, y_{i,b}^{\bar{D}*}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^{D}, y_{j,b}^{D*}\right)\right\}_{j=1}^{n_D}$ where
$$y_{i,b}^{\bar{D}*} = \hat{\mu}_{\bar{D}}\left(x_i^{\bar{D}}\right) + \hat{\sigma}_{\bar{D}}\left(x_i^{\bar{D}}\right)\varepsilon_{i,b}^{\bar{D}*},$$
$$y_{j,b}^{D*} = \hat{\mu}_{D}\left(x_j^{D}\right) + \hat{\sigma}_{D}\left(x_j^{D}\right)\varepsilon_{j,b}^{D*},$$
and $\left\{\varepsilon_{i,b}^{\bar{D}*}\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\varepsilon_{j,b}^{D*}\right\}_{j=1}^{n_D}$ are two i.i.d samples from the distributions $\hat{S}_{\bar{D}}$ and $\hat{S}_D$ respectively; and,
2. from $\left\{\left(x_i^{\bar{D}}, y_{i,b}^{\bar{D}*}\right)\right\}_{i=1}^{n_{\bar{D}}}$ and $\left\{\left(x_j^{D}, y_{j,b}^{D*}\right)\right\}_{j=1}^{n_D}$ obtain $\widehat{ROC}^{*}_{X=x,b}(t)$ and compute
$$\widehat{AUC}^{*}_{X=x,b} = \int_0^1 \widehat{ROC}^{*}_{X=x,b}(t)dt.$$

Once the above process has been completed, the 100 per cent x $(1-\alpha)$ limits for the confidence interval for the true $AUC_{X=x}$ are given by

$$\left(\widehat{AUC}^{\alpha/2}_{X=x}, \widehat{AUC}^{1-\alpha/2}_{X=x}\right)$$

where $\widehat{AUC}^{p}_{X=x}$ represents the p-percentile of the estimated $\widehat{AUC}^{*}_{X=x,b}$ $(b = 1, \ldots, B)$.

5.3 Results

Figure 6 shows the non-parametric estimates of the regression functions and the standard deviation of the adjusted regression models for each gender, along with the pointwise bootstrap 95% confidence intervals (contructed using the methodology described above for the AUC). Healthy men and women behaved as expected, with BMI displaying an age-dependent increase, except in the case of subjects

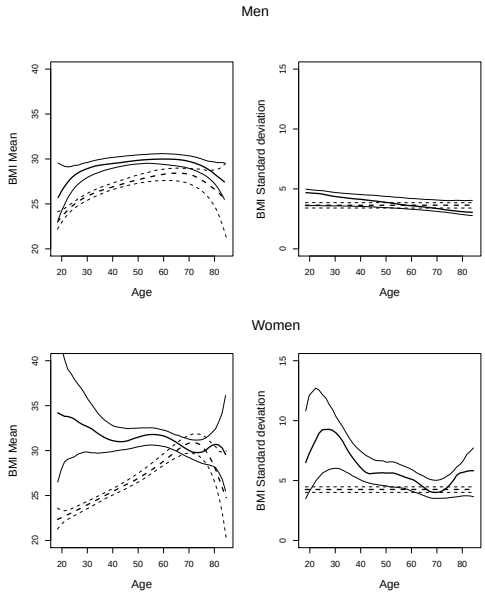

Figure 6: Non-parametric estimates of BMI by age (years) in male and female populations, along with 95% pointwise bootstrap confidence interval. Solid line: diseased population. Dashed line: healthy population. Left: Regression functions. Right: Variance functions.

over the age of 70 years, women in particular. Insofar as variability was concerned, no change was in evidence across age. Focusing on diseased subjects, men seemed to register a similar behaviour to that of their healthy counterparts, albeit with systematically higher BMI values, as can be seen in Figure 6, where the average curves in healthy and diseased subjects were approximately parallel. No major changes were observed in BMI across age in the case of diseased women, with these displaying a clearly different behaviour from that of healthy women. From ages 40 to 50 years, an inflection point was observable, probably related to climateric/perimenopausal years. As was to be expected, variability in diseased subjects changed across age, with young individuals registering the highest values. This change was particularly important among women.

The induced covariate-specific ROC curves and the AUC curves together with the corresponding 95% bootstrap confidence intervals are shown in Figure 7.

To evaluate the effect of age on the accuracy of BMI in both males and females, we performed the test presented in Section 3 above. Whereas the p-value obtained in the case of men was 0.8, in the case of women it was < 0.001. P-values were based on 200 bootstrap replications. Binning acceleration made it possible for the test to be concluded in under 28 seconds (using a 1.6GHz Intel Core2, 1GB RAM).

Finally, age-specific thresholds for BMI, based on the criterion of equal sensitivity and specificity, are shown in Figure 8 for men and women respectively. In the case of women, the threshold values were obtained on the basis of an age-specific ROC curve; in the case of men, however, the covariate-adjusted ROC curve was used because age does not affect 's discriminatory capacity. It should be noted that most BMI thresholds remain in the classic overweight range ($25 \leq \text{BMI} < 30$) for all

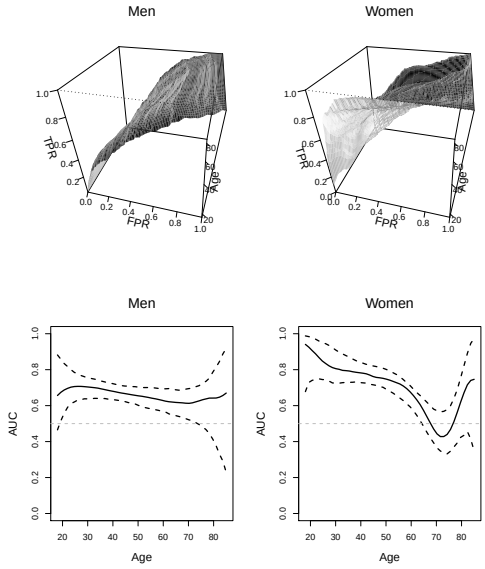

Figure 7: Estimated ROC surfaces and AUC curves with 95% pointwise bootstrap confidence interval for men and women.

ages. Our analysis suggests that age- and gender-specific BMI thresholds must be established by taking age as a continuous covariate. This is an expected consequence of the well-known change in body composition that takes place with ageing (Chumlea et al. 1992), yet none of the current classifications for the establishment of pathological values for BMI takes this into account.

6 Discussion

This study presents a non-parametric estimator for the ROC curve in the presence of a continuous covariate, along with a testing technique to check whether the covariate affects diagnostic test performance.

The proposed estimation procedure is an extension to the non-parametric case of induced methodology. Local polynomial estimators are used to estimate regression and variance functions in healthy and diseased subjects, and survival functions are empirically estimated on the basis of standardised residuals. Use is made of the cross-validation criterion to choose optimal smoothing parameters, and binning techniques to speed up the computation time.

Insofar as the testing procedure is concerned, the technique is based on the distance between the non-parametric estimator of the covariate-specific ROC curve and the covariate-adjusted ROC curve. Under the null hypothesis of non-effect of the continuous covariate on the ROC curve, both curves coincide. The simulation study performed shows that our estimator performs well and that, as was to be expected, its variance decreases as sample size increases. The proposed test for assessing covariate effect yields type I errors relatively close to nominal errors, regardless of sample size. In terms of power, this grows: (a) as one moves further away from the null hypothesis; and,

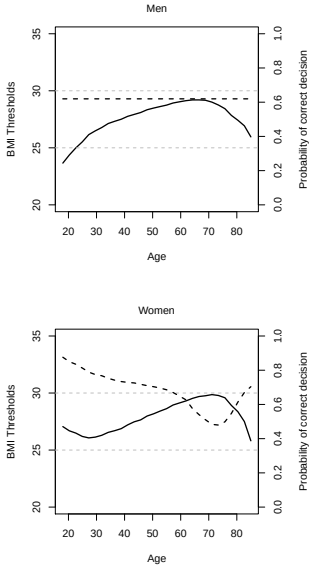

Figure 8: Estimated threshold curves for BMI, along with the sensitivity and specificity linked to these values. Solid line: threshold values. Black dashed line: sensitivity and specificity. Grey dashed lines: range for overweight. The threshold values shown are those for which sensitivity and specificity coincide.

(b) as sample size increases. The test's behaviour can be considered satisfactory.

Finally, this methodology was used to analyse an endocrine dataset, in order to evaluate the effect of age on BMI's discriminatory capacity when it came to detecting the presence of two or more cardiovascular risk factors. The analysis was performed separately by gender. The test proposed in this paper for evaluating the effect of a continuous covariate on the ROC curve 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.

It should be pointed out that, in practical situations, the new methodology represents a flexible exploratory tool for identifying non-linear covariates effects on the ROC curve. In many instances, the proposed methodology can also be used in a diagnostic mode, as an aid for choosing (where necessary) parametric transformations for the covariates. Once reparametrizations have been discovered, subsequent fitting can then be undertaken by using parametric regression models.

Our methodology suffers from the limitation of solely being able to address a single continuous covariate. Despite the fact that it would be difficult from a practical standpoint (where the goal sought is the possible choice of different threshold values) to find situations where the effect on the ROC of more than one continuous covariate and a categorical covariate might have to be assessed, further work is nevertheless needed to extend the proposed methodology to the case of multiple covariates.

Even though local polynomial estimators were used in this paper to estimate regression and variance functions, the algorithm proposed in Section 2 accepts any smoother, e.g., P-splines (Eilers and Marx 1996). A comparative study between different approaches is an area for further research.

When several diagnostic markers are available, an optimal linear combination of them can be computed such that the AUC of the combination is maximized (MacIntosh and Pepe 2002, Su and Liu 1993). Schisterman et al. (2004) discuss covariate effects on this linear combination assuming that the multiple markers, possibly transformed, follow a multivariate normal distribution. These ideas can be used to extend the results presented in this paper to a multiple response setup.

Acknowledgements

The authors would like to express their gratitude for the support received in the form of the Spanish MEC Grant MTM2008-01603 and Galician Regional Authority (Xunta de Galicia) projects INCITE08PXIB208113PR and PGIDIT07PXIB300191PR, and would like to thank the Galician Endocrinology & Nutrition Foundation (*Fundacion de Endocrinoloxia e Nutricion Galega - FENGA*) for having supplied the database used in this study.

Appendix: Univariate local polynomial kernel estimator. Computational aspects.

Let $\{(x_i, y_i)\}_{i=1}^n$ be an independent random sample of (X, Y). The local polynomial kernel estimator $\hat{\psi}(x) = \hat{\psi}(x, \{(x_i, y_i)\}_{i=1}^n, h, p)$ at a location x is defined as

$$\hat{\psi}(x) = \hat{\beta}_0, \tag{11}$$

where $\hat{\beta} = \left(\hat{\beta}_0, \ldots, \hat{\beta}_p\right)$ is the minimiser of

$$\sum_{i=1}^{n} \left\{ y_i - \sum_{j=0}^{p} \beta_j \left(x_i - x\right)^p \right\}^2 K_h \left(x_i - x\right),$$

where $K_h(.) = K(./h)/h$, $K(.)$ denotes a kernel function (a symmetric density), $h > 0$ is the smoothing parameter and p is the order of the polynomial. It should be noted that for $p = 0$, $\hat{\psi}\left(x\right)$ becomes the Nadaraya-Watson estimator; for $p = 1$ the local linear smoother is obtained and so on.

In this paper, the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}\exp\left(-u^2/2\right)}$ was used. The smoothing parameter, h, was selected automatically by minimising the following cross-validation error criterion

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{\psi}^{(-i)}\left(x_i\right) \right)^2, \tag{12}$$

where $\hat{\psi}^{(-i)}(x_i)$ indicates the estimate at x_i leaving out the i^{th} element of the sample.

To speed up the estimation and smoothing parameter selection processes, binning-type acceleration techniques were used (Fan and Marron 1994). In so-called linear binning, the idea is to create a grid of N equidistant points along the range of X, and assign to each grid point a weight equal to the number of observations in its bin.

Specifically, let $\tilde{X}_1 < \tilde{X}_2 < \ldots < \tilde{X}_N$ be a grid of N equidistant points along the range of X, with ζ being the distance between consecutive grid points. The weight of the i^{th} observation is assigned to the closest grid points according to

$$\tilde{W}_i^r = \left(1 - \left|x_i - \tilde{X}_r\right| / \zeta\right)_+, r = 1, \ldots, N.$$

In this way, the binning responses $\tilde{Y}_r$ and the binning weights $\tilde{W}_r$ for $r = 1, \ldots, N$ are constructed as follows:

$$\tilde{W}_r = \sum_{i=1}^{n} \tilde{W}_i^r \text{ and } \tilde{Y}_r = \frac{1}{\tilde{W}_r} \sum_{i=1}^{n} \tilde{W}_i^r y_i,$$

and the binning approximation of the estimator $\hat{\psi}$ in (11) is obtained by minimising

$$\sum_{r=1}^{N} \left\{ \tilde{Y}_r - \sum_{j=0}^{p} \beta_j \left(\tilde{X}_r - x\right)^p \right\}^2 K_h \left(\tilde{X}_r - x\right) \tilde{W}_r,$$

w.r.t $\beta = (\beta_0, \ldots, \beta_p)$.

As in the estimation procedure, in the case of the binning technique the cross-validation error (12) can be approximated by

$$CV \approx \sum_{r=1}^{N} \tilde{W}_r \left(\tilde{Y}_r - \hat{\psi}^{(-r)}\left(\tilde{X}_r\right) \right)^2.$$

The choice of the number of grid points is a compromise between approximation error and computational speed: the finer the grid of points selected, the better the binning approximations. In this paper, we used $N = 30$ grid points along the range of X. In practice, depending on the sample size, n, and the distribution of the covariate, a larger amount of grid points might be more appropriate.

References

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

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

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.

Cai, T. and Zheng, Y. (2007). Model Checking for ROC Regression Analysis. *Biometrics* **63**, 152–163.

Carey, V.J., Walters, E.E., Colditz, G.A. et al. (1997). Body fat distribution and risk of noninsulin-dependent diabetes in women: the Nurses Health Study. *Am. J. Epidemiol.* **145**, 614–19.

Chumlea, W.C., Baumgartner, R.N., Garry, P.J., Rhyne, R.L., Nicholson, C. and Wayne, S. (1992). Fat distribution and blood lipids in a sample of healthy elderly people. *International Journal of Obesity* **16**, 125–133.

Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall, New York.

de Boor, C.A. (2001). *A practical guide to splines.* Revised Edition. Springer-Verlag, New York.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, CRC.

Fan, J. and Marron, J.S. (1994). Fast implementation of non-parametric curve estimators. *Journal of Computational and Graphical Statistics* **3**, 35–56.

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. (2008). ROC curves in non-parametric location-sacle regression models. *Tecnical Report 08-01, reports in Statistics and Operations Research.* Universidade de Santiago de Compostela.

Haslam, D.W. and James, W.P.T. (2005). Obesity. *Lancet* **366**, 1197–1209.

Hsieh, F. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics* **24**, 24–40.

Hu, G., Qiao, Q., Tuomilehto, J. et al. (2004). Blasma insulin and cardiovascular mortality in non-diabetic European men and women: a meta-analysis of data from eleven prospective studies. The DECODE Insulin Study Group. *Diabetologia* **47**, 1245–56.

International Diabetes Federation (2006). The IDF consensus worldwide definition of the metabolic syndrome. http://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf. Accessed 7 September 2009.

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

Karelis, A.D., St-Pierre, D.H., Conus. F., Rabasa-Lhoret, R. and Poehlman, E.T. (2004). Metabolic and body composition factors in subgroups of obesity: What do we know?. *Journal of Clinical Endocrinology and Metabolism* **89**, 2569–2575.

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.

Lloyd, C.J.(1998). Using smooth receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association* **93**, 1356–1364.

MacIntosh, M.W. and Pepe, M.S. (2002). Combining Several Screening Test: Optimality of the Risk Score. *Biometrics* **58**, 657–664.

Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**, 183–298.

Mirmiran, P., Esmaillzadeh, A. and Azizi, F. (2004). Detection of cardiovascular risk factors by anthropometric measures in Tehranian adults: receiver operating characteristic (ROC) curve analysis. *European Journal of Clinical Nutrition* **58**, 1110–1118.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications* **9**, 141–142.

Peden, M., McGee, K. and Krug, E. (Eds) (2002). Injury: A Leading Cause of the Global Burden of Disease, 2000. World Health Organization: Geneva, Switzerland.

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. (1998). Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results. *Biometrics* **54**, 124–135.

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

Ren, H., Zhou, X.H. and Liang, H. (2004). A Flexible Method for Estimating the ROC curve. *Journal of Applied Statistics* **31**, 773–784.

Schisterman, E.F., Faraggi, D. and Reiser, B. (2004). Adjusting the generalized ROC curve for covariates. *Statistics in Medicine* **23**, 3319-3331.

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

Swets, J.A. and Pickett, R.M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.

Tomé, M.A., Botana, M.A., Cadarso-Suárez, C., Rego-Iraeta, A., Fernández-Mario, A., Mato, J.A., Solache, I. and Perez-Fernandez, R. (2009). Prevalence of metabolic syndrome in Galicia (NW Spain) on four alternative definitions and association with insulin resistance. J*ournal of Endocrinological Investigation*. To appear.

Tosteson, A.N. and Begg, C.B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **8**, 204–215.

Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. Chapman & Hall, London.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya Series A* **26**, 359–372.

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

Zimmet, P., Alberti, K.G.M.M. and Shaw, J. (2001). Global and societal implications of the diabetes epidemic. *Nature* **414**, 782–7.

Zhou, X.H. and Harezlak, J. (2002). Comparison of bandwidthselection methods for kernel smoothing of ROc curves. *Statistics in Medicine* **21**, 2045–2055.

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

Zou, K.H., Hall, W.J. and Shapiro, D.E. (1997). Smooth nonparametric receiver operating characteristic (ROC) curves for continuous diagnostic test. *Statistics in Medicine* **16**, 2143–2165.

Zweig, M.H. and Campbell, G. (1993). Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**, 561–577.