Adjusting Covariates in CRIB Score Index Using ROC Regression Analysis

Maria Filipa Mourão[1(✉)], Ana C. Braga[2], Alexandra Almeida[3], Gabriela Mimoso[4], and Pedro Nuno Oliveira[5]

[1] School of Technology and Management, Polytechnic Institute of Viana do Castelo, 4900-348 Viana do Castelo, Portugal
fmourao@estg.ipvc.pt
[2] ALGORITMI Centre, University of Minho, 4710-057 Braga, Portugal
acb@dps.uminho.pt
[3] Serviço de Neonatologia e Cuidados Intensivos Pediátricos, Centro Materno Infantil do Norte, Centro Hospitalar do Porto, 4099-001 Porto, Portugal
maria.alexandra.almeida@gmail.com
[4] Serviço de Neonatologia B da Maternidade Bissaya Barreto, Centro Hospitalar e Universitário de Coimbra, 3000-075 Coimbra, Portugal
gabriela.mimoso@gmail.com
[5] Biomedical Sciences Abel Salazar Institute, University of Porto, 4050-313 Porto, Portugal
pnoliveira@icbas.up.pt

Abstract. In medical studies, the receiver operating characteristic (ROC) curve is a tool of extensive use to analyze the discrimination capability of a diagnostic variable. In certain situations, the presence of related covariate, continuous or categorical, to the diagnostic variable can increase the discriminating power of the ROC curve [3].

The Clinical Risk Index for Babies (CRIB) scale, appeared in 1993 to predict the mortality of babies with very low birthweight (VLBW) and/or less than 32 weeks of gestation [2]. Braga and Oliveira [1] concluded that this index performs well in computing the risk of death for VLBW infants (< 1500 g).

In previous works, the authors studied the effect of the baby's sex [17] and the mother's age [18] on CRIB scale, using results of an intensive care unit of a Portuguese hospital.

In the present work, we propose to analyze the discriminative power of CRIB scale, using ROC regression analysis with GLM (Generalized Linear Models), in the classification of babies with and without the presence of covariates (newborn gender and mothers age).

This study is carried out using a random sample obtained from data collected during the period from $2010 - 2012$. The data source was the "Portuguese VLBW infants network" that encompasses all newborns with less than 1500 g or 32 weeks of gestational age born in Portugal.

Keywords: Conditional ROC curve and CRIB · Nonparametric regression model · Covariates

O. Gervasi et al. (Eds.): ICCSA 2015, Part II, LNCS 9156, pp. 157–171, 2015.
DOI: 10.1007/978-3-319-21407-8_12

1 Introduction

1.1 CRIB Index

Assessment of newborn risk in neonatal intensive care units tends to rely on the risk of mortality adjusted for birthweight and gestational age. This is especially true when we refer to VLBW infants or newborns with a gestational age less than 32 weeks. In the 90s, neonatal scoring systems were developed taking into account other factors to predict newborn mortality. The use of scoring scales also allows us to standardize the patients severity, making hospital benchmarking easier.

Among the many existing scales to assess and classify the clinical status of very premature babies, the CRIB and SNAPPE II (Score for Neonatal Acute Physiology Perinatal Extension) are the most commonly used scales in Portugal.

The CRIB scale uses six variables (birth weight, gestational age, congenital malformations and three physiological measures), collected in the first 12 hours of life, to evaluate the clinical severity of newborns, which makes it less susceptible to the effect of treatments [2]. The result of this scale is based on a weighted sum of these six variables and their possible final value is between 0 and 23. The higher the results in the scale, the higher the probability of death.

1.2 ROC Curve

The ROC curve is a graph representation of true positive rates on the vertical axis and false positive rates on the horizontal axis for different values of a classification threshold. It can be interpreted as a curve that summarizes the information of the cumulative distribution functions of the scores of the two classes considered.

We take as a starting point the existence of two populations, a positive/abnormal population which we denote by P and a negative/normal population denoted by N, together with a classification rule for allocating the individuals by each of these populations. We assume this classification rule to be in some continuous function $t(x)$ of the random vector X of variables measured on each individual, conventionally arranged so that large values of the function are more indicative of population P and small ones more indicative of population N. Thus if x is the observed value of X for a particular individual and $t(x)$ is the function score for this individual, then he is allocated to population P or population N in case $t(x)$ exceeds or does not exceed some threshold c, respectively. Supposing that c is the value of the threshold in a particular classification rule, so as that an individual is allocated to population P if its classification score t exceeds c and to population N in case it doesn't exceed. In order to assess the efficacy of this classifier we need to calculate the probability of making an incorrect allocation. More specifically, we can define four probabilities and their associated rates for the classifier:

- the probability that an individual from P is correctly classified, i.e., the true positive fraction, $TPF = Prob(t > c|P)$;
- the probability that an individual from N is misclassified, i.e., the false positive fraction $FPF = Prob(t > c|N)$;

– the probability that an individual from N is correctly classified, i.e., the true negative fraction $TNF = Prob(t \leq c|N)$; and
– the probability that an individual from P is misclassified, i.e., the false negative fraction $FNF = Prob(t \leq c|P)$.

Given probability densities $Prob(t|P)$, $Prob(t|N)$, and the value c, numerical values lying between 0 and 1 can be obtained readily for these four rates and this gives a full description of the performance of the classifier.

Clearly, for a good performance, we require high true and low false fractions. However, this only appens for a particular choice of threshold c, and the best choice of this threshold is not generally known in advance but must be determined as part of the classifier construction. Varying c and evaluating the four probabilities mentioned above, will give full information on which this decision will be based and hence to assess the performance of the classifier. The ROC curve provides a much more easily accessible summary. It is the curve obtained on varying c, but when using just the true and false positive fractions and plotting (FPF, TPF) as points against orthogonal axes. Here FPF is the value on the horizontal axis (abscissa) and TPF is the value on the vertical axis (ordinate).

The purpose of the ROC curve is to provide an assessment of the classifier over the whole range of potential threshold values rather than at just a single one. Clearly, the value of a classifier can be judged by the extent to which the two distributions of its scores $Prob(t|P)$ and $Prob(t|N)$ differ. The more they differ, the lesser will there be any overlap between them and so the less likely will the incorrect allocations be made. Hence, the more successful will the classifier be in the making of correct decisions. Conversely, the more the two distributions are alike, the more overlap there is between them and so the more likely for incorrect allocations to be made.

In practice, the ROC curve is a continuous curve which is between $(0, 0)$ and $(1, 1)$, situated at the upper triangle that defines the chart. The closer it is to the upper left corner of the graph, the closer we are to a situation of complete separation between the populations and therefore, the better the performance of the classifier.

When the score t is a continuous variable, it will possess density and distribution functions in each of the two populations. Let us denote its density and distribution functions in population N by f_N, F_N respectively, and in population P by f_P, F_P respectively. Thus,

$$\begin{aligned} Prob(t|N) &= f_N(t) & \qquad Prob(t|P) &= f_P(t) \\ x_N(c) &= 1 - F_N(c) & \qquad x_P(c) &= 1 - F_P(c) \end{aligned}$$

Eliminating c from this last pair of equations, using standard mathematics, yields the form of the ROC curve

$$ROC(x) = 1 - F_P[F_N^{-1}(1 - x)] \quad \text{for } x \in (0, 1) \tag{1}$$

and the expression (1) will be the most convenient equation of the ROC curve to use with continuous scores.

We have seen above that the ROC is a convenient summary of the full set of information that would be needed for a comprehensive description of the performance of a classifier over all its possible threshold values. However, even such a summary may be too complicated in some circumstances, for instance, if a plot is difficult to produce or if very many different classifiers need to be compared, the interest therefore relies on obtaining deriving simpler summaries. Particular attention has been focused on single scalar values that might capture the essential features of a ROC curve, like the Area Under the ROC Curve (AUC). This is the most widely used summary index studied by Green and Swets [20], Bamber [21], Hanley and McNeil [12], [22] among others.

Simple geometry establishes the upper and lower bounds of AUC: for the case of perfect separation of P and N distributions, AUC is the area under the upper borders of the ROC (i.e., the area of a square of side 1). So the upper bound is 1.0, while for the case of random allocation AUC is the area under the chance diagonal (i.e., the area of a triangle whose base and height are both equal to 1) so the lower bound is 0.5. For all other cases, the formal definition is

$$AUC = \int_0^1 ROC(x)dx \tag{2}$$

One immediate interpretation of AUC follows from this definition, elementary calculus and probability theory, plus the fact that the total area of the ROC domain is 1.0: AUC is the average of true positive fractions, taken uniformly over all possible false positive fractions in the range $(0, 1)$. Another interpretation is as a linear transformation of the average misclassification rate, weighed by the mixture distribution of the true P and N classes [19]. We can assume from the above definition that, if A and B are two classifiers for witch the ROC curve for A is not inferior the ROC curve for B, then AUC for A must be greater than or equal to AUC for B. Unfortunately, the reverse implication is not true because of the possibility that the two curves can cross each other.

1.3 Estimating ROC Curve

For continuous scores t, as we have seen, the ROC curve can be expressed by equation (3).

The estimation problem is thus reduced to the estimation of this curve from the given data. To obtain the empirical estimator, we simply apply the relevant definitions above to the observed data. Thus, if n_P and n_N are the numbers of individuals in the samples from populations P and N respectively, and if $n_A(c)$ denotes the number of individuals in the sample from population A (where A is either N or P) whose classification scores are greater than c, then the empirical estimators of the true positive fraction $TPF = p(t > c|P)$ and false positive fraction $FPF = p(t > c|N)$ at the classifier threshold c are given by

$$\widehat{TPF} = \frac{n_P(c)}{n_P}$$

$$\widehat{FPF} = \frac{n_N(c)}{n_N}$$

Thus, plotting the set of values $(1 - \widehat{FPF})$ against c yields the empirical distribution function $\hat{F}_N(c)$ and doing the same for values $(1 - \widehat{TPF})$ yields the empirical distribution function $\hat{F}_P(c)$. The empirical ROC curve is then simply given by plotting the points $(\widehat{FPF}, \widehat{TPF})$ obtained on varying c with equation given by (3).

$$\widehat{ROC}(x) = 1 - \hat{F}_P[\hat{F}_N^{-1}(1-x)] \quad \text{for } x \in (0,1) \tag{3}$$

The expression (3) is called the empirical ROC curve.

2 ROC Curve and Covariates

Once a classifier t has been constructed from the vector X of principal variables and is in use for allocating individuals to one or the other of the populations N and P, confounding occurs in evaluating classification accuracy when there is a covariate which is associated with both the classifier and the binary outcome. For maximum benefit, such additional variable should be incorporated into any analysis involving the classifier. Let us denote a set of z covariates by the vector Z, recognizing that in many practical applications we may have just one covariate Z. In such cases, it is necessary to adjust the ROC curve and summaries derived therefrom before drawing any inferences. In particular, it will often be relevant to compute the ROC curve and allied summaries at particular values of the covariates Z, in order to relate these covariate-specific curves and summaries to sample members that have covariate values. Ignoring covariate values leads to the calculation of a single pooled ROC curve (the pooling being over all the possible covariate values).This traditional pooled ROC curve, which combines all case and control observations, regardless of the covariate value, is biased. Pepe [16] provides several important results linking the covariate-specific and the pooled ROC curves.

The ROC curve, in presence of covariates, can be considered for each value z of the covariate. Changes that occur in the curve, due to these values, might mean that the covariate has an effect on the discrimination power of the diagnostic test. The conditional ROC curve is defined as

$$ROC_z(x) = F_{PZ}[F_{NZ}^{-1}(x)] \quad \text{for } x \in (0,1) \tag{4}$$

where F_{PZ} and F_{NZ} are the conditional survival functions associated with subjects of populations P and N, respectively, and are estimated by

$$\hat{F}_{PZ}(c) = \frac{1}{n_P} \sum_{i=1}^{n_P} I(X_{Pi} \geq c) \tag{5}$$

$$\hat{F}_{NZ}(c) = \frac{1}{n_N} \sum_{i=1}^{n_N} I(X_{Ni} \geq c) \tag{6}$$

So, the covariate-adjusted ROC curve, is a measure of covariate-adjusted classification accuracy.

Adjusting ROC curves derived from continuous classification scores has been considered by several authors like Smith and Thompson [5], Pepe [6], [7], [8], Faraggi [9], Janes and Pepe [10], [11], Pepe and Cai [23] and Janes et al. [24]. Adjustment of summary values derived from the ROC curve has further been studied by Faraggi [9] and Dodd and Pepe [25], [26].

To obtain de covariate-adjusted ROC curve, two distinct approaches can be followed. The *induced adjustment*, proposed by Tolsteson and Begg [4], Zheng and Heagerty [13], Faraggi [9], in which the covariates have an effect on the diagnostic test is modeled in the two populations (P and N) separately and the ROC curve is then derived from the modified distributions; and *direct adjustment*, proposed by Pepe [8], Alonzo and Pepe [15], Cai [14], in which the effect of the covariates is modeled on the ROC curve itself. The authors, in [18], provide details about these two methodologies.

There are covariates that affect the classifier distribution among negatives. For example, center effects in multicenter studies may affect classifier observations. Other covariates may affect the inherent discriminatory accuracy of this (i.e., the ROC curve). As an example, disease severity can often affect the classifier accuracy, thus less severe positives can be more difficult to distinguish from negatives.

A separate ROC curve should be estimated for each covariate group that affects the discriminatory accuracy of the classifier. Covariate adjustment is often a necessary first step in estimating covariate-specific ROC curves in order to adjust the effects of the covariate on classifier observations among negatives.

In this work, we will use the STATA software to obtain the covariate-adjusted ROC curve.

3 ROC Regression with STATA Software

STATA software uses the ROC regression as methodology to model the classifier ROC curve as a function of covariates [8], [15]. Implementation proceeds in two steps:

1^{st}**:** Modeling the distribution of the classifier among negatives, as a function of covariates, and calculating the case Percentile Values or specificity;

2^{nd}**:** Modeling their cumulative distribution function (i.e., the ROC curve) as a function of covariates.

The result is an estimate of the ROC curve for the classifier as a function of covariates, or a covariate-specific ROC curve.

In STATA, the `rocreg` command is used to perform receiver operating characteristic (ROC) analyses with rating and discrete classification data under the presence of covariates.

This function can fit three models: a nonparametric model, a parametric probit model that uses the bootstrap for inference, and a parametric probit model fit using maximum likelihood.

The syntax to perform this analysis is given by:

`rocreg` *refvar classvar*[*classvars*] [if] [*in*] [, $np_{options}$]

The two variables *refvar* and *classvar* must be numeric. The reference variable indicates the true state of the observation − such as diseased (abnormal, P) and nondiseased (normal, N) − and must be coded as 1 and 0, respectively. The *refvar* coded as 0 can also be called the control population, while the *refvar* coded as 1 comprises the case population. The rating or outcome of the diagnostic test or test modality is recorded in *classvar*, which must be ordinal, with higher values indicating higher risk.

The covariate-adjusted ROC curve [24] at a given false-positive rate x is equivalent to the expected value of the covariate-specific ROC at x over all covariate combinations. When the covariates in question do not affect the case distribution of the classifier, the covariate-specific ROC will have the same value at each covariate combination. So here the covariate-adjusted ROC is equivalent to the covariate-specific ROC, regardless of covariate values.

4 Applications and Results

In this work, we intend to analyze and evaluate the discriminatory power of CRIB scale in the classification of risk of death for VLBW infants in Portugal. Available data refers to the period between 2010 and 2012, and we decided to select a random sample (approximately 50%) of original data collected by the RNMBP − Registo Nacional de Muito Baixo Peso (Portuguese VLBW infants network) for this period. We analyzed the performance of this scale without the inclusion of covariates and then considered their inclusion. We first perform ROC analysis for the classifier while adjusting for babies gender, then for maternal age and finally combining the information from both. This is done by specifying these variables in the *ctrlcov()* option. We adjust the covariates using a linear regression rule, by specifying *ctrlmodel(linear)*. This means that when a user of the diagnostic test chooses a threshold conditional on covariates, he assumes that the diagnostic test classifier has some linear dependence and equal variance, since their levels vary. Our cluster adjustment is made by specifying the *cluster()* option.

STATA software computes the percentile, or specificity, values empirically, and thus so do false-positive rates, $(1 - specificity)$. Also, the ROC curve values are empirically defined by the true-positive rates.

4.1 CRIB Index Without Covariates

In this section we verify the performance of CRIB scale in the classification of babies as "deceased" and "alive". To achieve its purpose, the study focused on the results presented in Table 1, which characterizes the grade given to newborns.

Table 1. Mortality rate

	Cases number	Percent
Deceased	166	10.45%
Alive	1423	89.55%
Total	1589	100.00%

From the results collected, 10.45% of the babies were classified as "deceased". The distribution of CRIB by result can be illustrated by the graph in Figure 1.

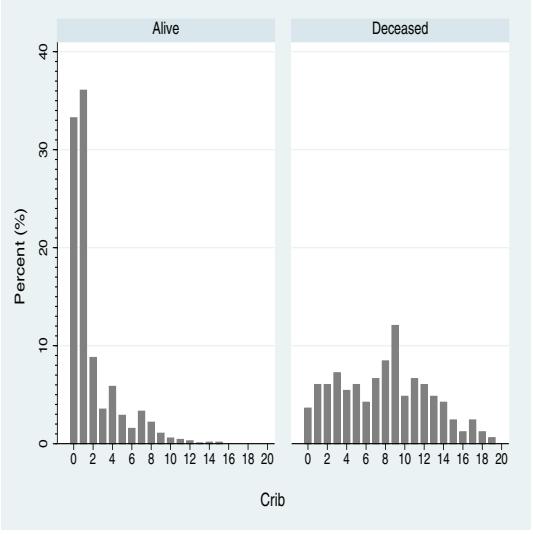

Fig. 1. CRIB classification by result

For these results, the empirical ROC curve was estimated, as is shown in Figure 2. Respective AUC, variability and confidence intervals were also obtained (Table 2).

From the estimated AUC value for the global empirical ROC curve (Figure 2) it is possible to observe that, without considering the effect of any the covariates, the CRIB index discriminates well between babies "alive" and "dead" in 87.9% of the cases.

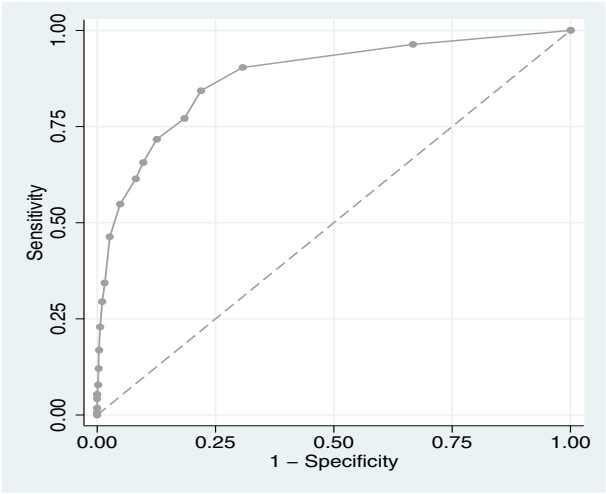

Fig. 2. Empirical ROC curve for CRIB index (output STATA)

Table 2. AUC's and Standard Errors results

	AUC	SE	95% LBCI	95% UBCI
CRIB	0.879	0.015	0.849	0.909

4.2 CRIB Index with Covariates

To verify if maternal age and the babie's sex contribute to the discrimination power of CRIB index, we propose to combine the CRIB information with additional information provided by these covariates. We use Nonparametric ROC estimation provided by STATA software to answer the research question:

"Does the discriminatory power of CRIB index increase with the inclusion of these covariates?"

We start by analyzing the contribution that each of the covariates provides in discriminatory power of the CRIB index, assuming that there is a linear relationship between the CRIB index and the covariate in question.

The conditioned ROC curve for the babie's sex is shown in Figure 3 and the associated results are in table 3. The AUC, SE and confidence intervals associated with ROC curve are shown in table 4.

When we include the sex of babies in the analysis, the AUC for conditioned ROC curve discriminate in 86.19% of the babies; less than the discriminates power without gender and higher standard error. Our covariate adjustment model shows that the newborn's sex has a negative effect on CRIB index.

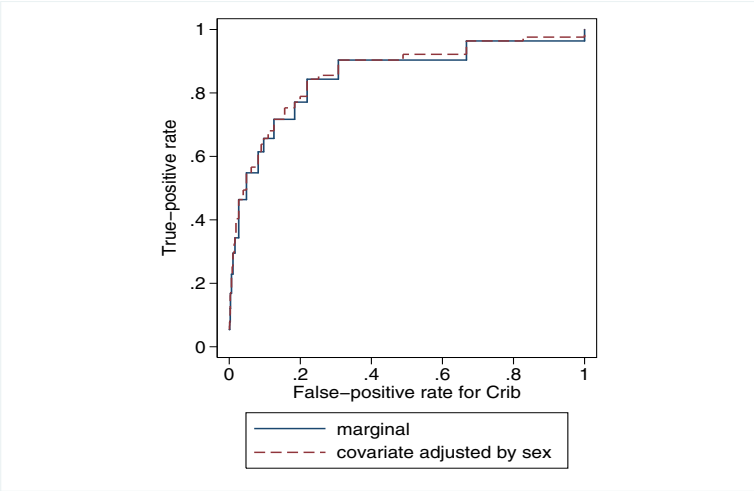

Fig. 3. Specific sex ROC curve (output STATA)

Table 3. Linear Regression results - CRIB conditioned by newborn sex

CRIB	Coefficient	SE	t-value	$P > \|t\|$	95% LBCI	95% UBCI
Sex	-0.0643	0.1306	-0.49	0.622	-0.3206	0.1919
Const	1.8934	0.2044	9.26	0.000	1.4923	2.2945

Table 4. AUC, SE and CI - CRIB conditioned by newborn sex

AUC	Bias	SE	95% LBCI	95% UBCI
0.8619	0.0022	0.1801	0.8266	0.8972
			0.8287	0.8977
			0.8234	0.8934

Let us now analyze the mother's age impact on CRIB index, assuming that there is a linear relationship between the CRIB index and this covariate. The specific CRIB ROC curve is shown in figure 4. The results for this linear regression are shown in table 5 and the correspondent AUC in table 6. When we include the mother's age in the analysis, the AUC for conditioned ROC curve discriminates in 87.98% of the babies, close to the AUC for CRIB without covariates. Despite the negative coefficient associated with this covariate in the regression model, it was no statistical significance ($p - value = 0.239 > 0.05$). It seems that the covariate has no impact in the discrimination power of the CRIB, in the towards of closeness of this AUC and the AUC of the CRIB ROC curve.

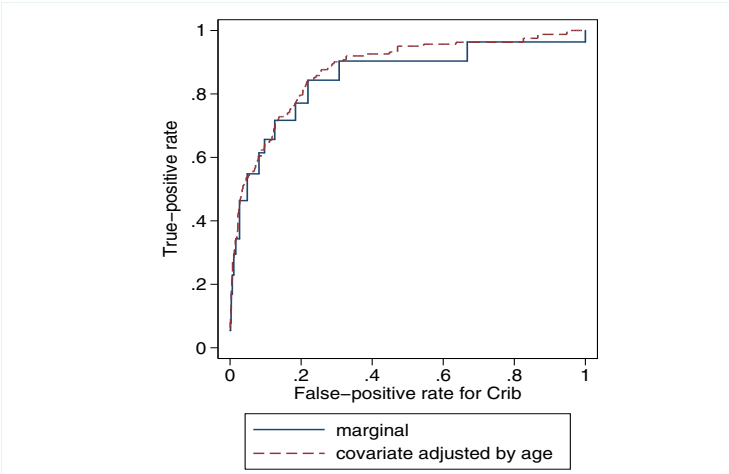

Fig. 4. Specific maternal age ROC curve (output STATA)

Table 5. Linear Regression results - CRIB conditioned by mother's age

CRIB	Coefficient	SE	$t-value$	$P>\|t\|$	95% LBCI	95% UBCI
Maternal age	-0.0100	0.0085	-1.18	0.239	-0.0269	0.0067
Const	2.1034	0.2689	7.82	0.000	1.5759	2.6309

Table 6. AUC, SE and CI - CRIB conditioned by mother's age

AUC	Bias	SE	95% LBCI	95% UBCI
0.8798	-0.0002	0.0152	0.8500	0.9097
			0.8487	0.9074
			0.8476	0.9062

Let us now combine the CRIB information with the information provided by these two covariates. We assume, once again, that there is a linear regression between the CRIB index and the two covariates. In figure 5, we show the specific CRIB ROC curve conditioned by the two covariates. Table 7 shows the results of the linear regression performed using STATA when we introduced the additional information of the two covariates. Table 8 shows the correspondent AUC, SE and confidence intervals.

Analyzing the results presented in Table 7, we can see that the coefficients for the two covariates are not different from zero ($p-value>0.05$). So, we can say that there is no statistical influence of these covariates in the performance

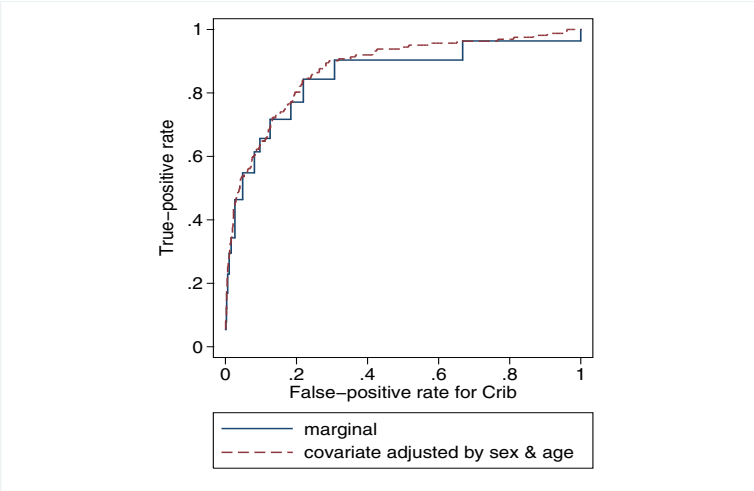

Fig. 5. Specific sex and maternal age ROC curve after ROC regression (output STATA)

Table 7. Linear Regression results - CRIB conditioned by sex and mother's age

CRIB	Coefficient	SE	$t - value$	$P > \|t\|$	95% LBCI	95% UBCI
Maternal age	-0.0101	0.0085	-1.18	0.236	-0.0269	0.0066
Sex	-0.0656	0.1306	-0.5000	0.6150	-0.3218	0.1905
Const	2.2018	0.3293	6.69	0.000	1.5558	2.8479

Table 8. AUC, SE and CI - CRIB conditioned by sex and mother's age

AUC	Bias	SE	95% LBCI	95% UBCI
0.8785	-0.0000471	0.0162	0.8468	0.9103
			0.8409	0.9084
			0.8385	0.9058

of the CRIB index. These could be confirmed by the values of AUC's for the adjusted and not adjusted ROC curve. Figure 6 illustrates this result.

5 Conclusions

From the estimated values of AUC for the global ROC curve it is possible to observe that, without considering the effect of covariates, the CRIB scale discriminates between "alive" and "dead" newborns in 87.9% of cases. This value, when considering the newborn sex as covariate, is 86.19% and with maternal age as covariate, it is 87.9%. We can also see that, standard-error (SE) associated

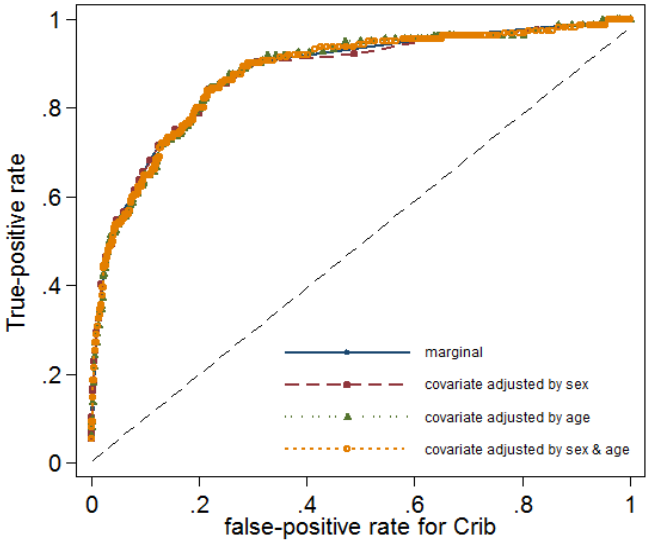

Fig. 6. All Specific ROC curves(output STATA)

to ROC curve, when considering the sex of babies as covariate, is greater when compared to the SE for the CRIB ROC curve and the SE for specific maternal age CRIB ROC curve.

From the results obtained, when we consider that there is a linear relationship between CRIB and the newborns gender, it is apparent that this covariate affects the discriminatory power of CRIB scale when used to classify babies. Our covariate adjustment model shows that the newborn sex has a negative effect on CRIB index, causing a decrease in the performance of CRIB scale. The standard error, in turn, is greater than the standard error of the ROC curve obtained without the covariate. Consequently, it also increases the dispersion of the pairs of values $(1-specificity, sensitivity)$. However, when analyzing the overall quality of the regression model, we do not reject the hypothesis that the contribution of the baby's sex for the CRIB scale is zero ($t = -0.49$, $p - value = 0.622$).

When we analyze the mother's age impact on CRIB index and assume that there is a linear relationship between the CRIB scale and this covariate, the results for this linear regression, despite the negative coefficient associated with it in the regression model, it seems that this covariate has no impact in the discrimination power of the CRIB scale, in the towards vicinity of its AUC and the AUC of the CRIB ROC curve. The same occurs for the SE, which remains unchanged. In fact, when analyzing the overall quality of the regression model, we do not reject the hypothesis that the contribution of the mother's age for the CRIB scale is zero ($t = -1.18$, $p - value = 0.239$).

Finally, after analyzing the results, considering the two covariates, from the AUC associated with specific ROC curve, we can conclude that CRIB index discriminate in 87.85%. This discriminatory power is higher than that found when considering only sex as a covariate but it is lower when we consider only the mother's age as a covariate. Looking at dispersion present in AUC of conditioned ROC curve for these two covariates, we found that it is higher than the ROC curve of CRIB and than the specified ROC curve for maternal age. However, it is inferior to the dispersion present in the CRIB ROC curve conditioned by newborn sex. When Analyzing the overall quality of the regression model, we do not reject the hypothesis that the contribuition of the newborn's sex and the mother's age to the CRIB scale is zero ($t_{sex} = -0.50$, $p - value = 0.615$ and $t_{age} = -1.18$, $p - value = 0.236$). So, we may conclude that the discriminatory power of CRIB scale is not affected when we consider that baby's sex and mother's age may influence the classification of babies in "dead" and "alive".

To answer the research question, we can say that the performance of CRIB to predict the risk of death of newborns with very low birth-weight and/or gestational age < 32 weeks is not affected by the sex of the babies nor the age of their mothers.

Acknowledgments. The authors would like to thank the availability of the data by the Portuguese VLBW infants network.
This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

1. Braga, A.C., Oliveira, P.: Diagnostic analysis based on ROC curves: theory and applications in medicine. International Journal of Health Care Quality Assurance, Emerald **16**(4), 191–198 (2003)
2. Dorling, J.S., Field, D.J., Manketelow, B.: Neonatal disease severity scoring systems. Arch. Dis. Child. Fetal Neonatal **90**, F11–F16 (2005)
3. López-de-Ullibarri, I., Cao, R., Cardaso-Suárez, C., Lado, M.J.: Nonparametric estimation of conditional ROC curves:application to discrimination tasks in computerized detection of early breast cancer. Computational Statistics & Data Analysis **52**(5), 2623–2631 (2008)
4. Tolsteson, A.N., Begg, C.B.: A general regression methodology for ROC curve estimation. Med. Decision Making **8**(3), 204–215 (1988)
5. Smith, D.J., Tompson, T.J.: Correcting for confounding in analising receiver operating characteristic curves. Biometrical Journal **38**, 357–863 (1996)
6. Pepe, M.S.: A regression modelling framework for receiver operating characteristic curve in medical diagnostic testing. Biometrika **84**, 595–608 (1997)
7. Pepe, M.S.: Three approaches to regression analysis of receiver operating characteristic curves for continuous tests results. Biometrics **54**, 124–135 (1998)
8. Pepe, M.S.: An interpretation for the ROC curve and inference using GLM procedures. Biometrics **56**, 352–359 (2000)
9. Faraggi, D.: Adjusting receiver operating characteristic curves and related indices for covariates. Journal of the Royal Statistical Society: Series D (The Statistician) **52**(2), 179–192 (2003)

10. Janes, H., Pepe, M.S.: Adjusting for covariates effects on classification accuracy using the covariate-adjusted ROC curve. UW Bioestatistics Workin Paper Series. Working paper 283 (2006). http://biostats.bepress.com/uwbiostat/paper283
11. Janes, H., Pepe, M.S.: Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. UW Bioestatistics Working Paper Series. Working paper 310 (2007). http://biostats.bepress.com/uwbiostat/paper310
12. Hanley, J.A., McNeil, B.J.: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology **148**, 839–843 (1983)
13. Zheng, Y., Heagerty, P.J.: Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics **4**, 615–632 (2004)
14. Cai, T.: Semiparametric ROC regression analysis with placement values. Biostatistics **5**, 45–60 (2004)
15. Alonzo, T.A., Pepe, M.S.: Distribution free ROC analysis using binary regression techniques. Biostatistics **3**, 421–432 (2002)
16. Pepe, M.S.: The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York (2003)
17. Mourão, M.F., Braga, A.C., Oliveira, P.N.: CRIB conditional on gender: nonparametric ROC curve. International Journal in Health Care and Quality Assurance **27**(8), 656–663 (2014)
18. Mourão, M.F., Braga, A.C., Oliveira, P.N.: Accommodating maternal age in CRIB scale: quantifying the effect on the classification. In: Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C., Rocha, J.G., Falcão, M.I., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2014, Part III. LNCS, vol. 8581, pp. 566–579. Springer, Heidelberg (2014)
19. Krzanowski, W.J., Hand, D.J.: ROC curves for Continuous Data. Chapman and Hall (2009)
20. Green, D.M., Swets, J.A.: Signal Detection Theory and Psychophysics (rev.ed) Robert, E. Krieger Publishing Company (1974)
21. Bamber, D.: The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology. **12**, 387–415 (1975)
22. Hanley, J.A., McNeill, B.J.: The meaning and use of the area under receiver operating characteristic (ROC) curve. Radiology **143**(1), 29–36 (1982)
23. Pepe, M.S., Cai, T.: The analysis of placement values for evaluating discriminatory measures. Biometrics **60**, 528–535 (2004)
24. Janes, H., Longton, G.M., Pepe, M.S.: Accommodating covariates in receiver operating characteristic analysis. Stata Journal **9**, 17–39 (2009)
25. Dodd, L.E., Pepe, M.S.: Semiparametric regression for the area under the receiver operating characteristic curve. Journal of the American Statistical Association **98**, 409–417 (2003)
26. Dodd, L.E., Pepe, M.S.: Partial AUC estimation and regression. Biometrics **59**, 614–623 (2003)