Application of Generalized Additive Models to the Evaluation of Continuous Markers for Classification Purposes

Mónica López-Ratón[1,*], Mar Rodríguez-Girondo[2], María Xosé Rodríguez-Álvarez[2], Carmen Cadarso-Suárez[1] and Francisco Gude[3]

[1]*Biostatistics Unit, Department of Statistics and Operational Research, University of Santiago de Compostela, Santiago de Compostela, Spain*

[2]*Statistical Inference, Decision and Operations Research Group, Department of Statistics and Operational Research, University of Vigo, Vigo, Spain*

[3]*Clinical Epidemiology Unit, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Galician Health Service (SERGAS), Santiago de Compostela, Spain*

Abstract: *Background*: Receiver operating characteristic (ROC) curve and derived measures as the Area Under the Curve (AUC) are often used for evaluating the discriminatory capability of a continuous biomarker in distinguishing between alternative states of health. However, if the marker shows an irregular distribution, with a dominance of diseased subjects in noncontiguous regions, classification using a single cutpoint is not appropriate, and it would lead to erroneous conclusions. This study sought to describe a procedure for improving the discriminatory capacity of a continuous biomarker, by using generalized additive models (GAMs) for binary data.

Methods: A new classification rule is obtained by using logistic GAM regression models to transform the original biomarker, with the predicted probabilities being the new transformed continuous biomarker. We propose using this transformed biomarker to establish optimal cut-offs or intervals on which to base the classification. This methodology is applied to different controlled scenarios, and to real data from a prospective study of patients undergoing surgery at a University Teaching Hospital, for examining plasma glucose as postoperative infection biomarker.

Results: Both, theoretical scenarios and real data results show that when the risk marker-disease relationship is not monotone, using the new transformed biomarker entails an improvement in discriminatory capacity. Moreover, in these situations, an optimal interval seems more reasonable than a single cutpoint to define lower and higher disease-risk categories.

Conclusions: Using statistical tools which allow for greater flexibility (e.g., GAMs) can optimize the classificatory capacity of a potential marker using ROC analysis. So, it is important to question linearity in marker-outcome relationships, in order to avoid erroneous conclusions.

Keywords: Discriminatory capability, ROC, AUC, optimal cutpoint, biomarker, plasma glucose.

1. BACKGROUND

Biomarkers serve various clinical functions, including diagnosis, prediction and prognosis of disease. Traditionally the receiver operating characteristic (ROC) curve [1, 2] and derived measures such as the Area Under the Curve (AUC) [3] are used for evaluating the discriminatory capability of a continuous marker in distinguishing between alternative states of health. Within the ROC framework, biomarker levels above and below a given cut-off value result in individuals being labeled as diseased or nondiseased, respectively. In many situations, however, the classification rule that minimizes the overall misclassification error is not necessarily the criterion used in ROC analysis [4]. For instance, in cases where the marker shows an irregular distribution, with a dominance of diseased subjects in non-contiguous regions, classification by reference to a cut-off value is neither feasible nor logical. Indeed, use of such an analysis would lead to erroneous conclusions, and a modification of the classification rule is therefore necessary [5]. An intuitive solution to this problem would be to estimate the probability of belonging to one of the states (e.g., diseased status) as a function of the values of the marker, with the predicted probabilities thus being regarded as a new continuous biomarker. We propose to estimate such probabilities by means of generalized additive models (GAMs) for binary data. GAMs are modern regression techniques that have the advantage of not assuming a parametric relationship between status and biomarker, and eliminate the need for the researcher to impose functional assumptions [6]. Furthermore, there are other important clinical tasks that are affected by the classification rule adopted, such as selection of optimal cut-off values [7-10]. The most widely-used criteria for cut-off value selection are based on accuracy measures, and so it is

*Address correspondence to this author at the Biostatistics Unit, Department of Statistics and Operational Research, Facultad de Medicina y Odontología, San Francisco s/n 15782, University of Santiago de Compostela, Santiago de Compostela, Spain; Tel: 981 56 31 00 ext 12282; Fax: 981 54 71 72; E-mail: monica.lopez.raton@gmail.com

essential that this decision be underpinned by the best classification rule based on a given marker. Accordingly, this study sought to describe a procedure for improving the discriminatory capacity of a continuous biomarker, by using generalized additive models for binary data. In addition, this paper outlines a new approach for dealing with the selection of cut-off values for a continuous biomarker.

2. METHODS

2.1. Statistical Methodology

Let Y be a continuous biomarker. An individual can be classified as diseased (D) or non-diseased ($\bar{D}$) on the basis of Y by choosing a cut-off value, c, such that: if $Y \geq c$, the subject is classified as diseased; and if $Y < c$, he/she is classified as non-diseased. Hence, each chosen cut-off value, c, will give rise to a true positive fraction (*TPF*) (or sensitivity), $TPF(c) = P[Y \geq c \mid D]$ and a false positive fraction (*FPF*) (or 1-specificity), $FPF(c) = P[Y \geq c \mid \bar{D}]$. In such a situation, the ROC curve is defined as the set of all *TPF-FPF* pairs that can be obtained by a varying cut-off value, c, $\left\{ \left(FPF(c), TPF(c) \right), c \in (-\infty, +\infty) \right\}$, or equivalently, as the function of the form $ROC(t) = S_D(S_{\bar{D}}^{-1}(t)), \ t \in (0,1)$, where $S_D(y) = P[Y \geq y \mid D]$ and $S_{\bar{D}}(y) = P[Y \geq y \mid \bar{D}]$ are the survival functions of Y in the diseased and healthy populations, respectively.

Several indices can serve as summaries of the discriminatory capacity of the ROC curve. The most commonly used of these is the area under the ROC curve (AUC), taking values from 0.5 (no power of discrimination) to 1 (perfect power of discrimination) [3].

As pointed out above, in many situations the *Y*-based classification rule that minimizes the overall misclassification error is not the criterion used in ROC analysis. Moreover, the best *Y*-based classifier with a cut-off as the decision threshold is that which is based on the conditional probability of one of the states (e.g., diseased), given the values of *Y* [4, 13]. Accordingly, the best classifier, $\tilde{Y}$, can be expressed as:

$$\tilde{Y} \equiv f(Y) = P[D \mid Y] \subset (0,1) \tag{1}$$

In practice, however, the function, f (.), of (1) is unknown, and its estimation would be required. In this study, the function f (.) is estimated by using a logistic GAM regression model as follows:

$$f(Y) = P[D \mid Y] = g^{-1}(\alpha + h(Y)) = \frac{\exp(\alpha + h(Y))}{1 + \exp(\alpha + h(Y))} \tag{2}$$

where g (.) is the logit link function (known), and h (.) is a smooth (unknown) function.

To date, several approaches have been proposed in the statistical literature for estimating the model (2), including methods based on penalized regression splines [14, 15] or the Bayesian versions of these [16]. Alternatively, the local scoring algorithm with kernel-type smoothers can also be used [17, 18].

In this paper, we used penalized regression with B(asic)-splines as smoothers [14], for the purpose of estimating the function $h(.)$. When estimating $h(.)$, a crucial step is choosing the smoothing parameter that is to control the smoothness of the resulting estimate. In this study, the optimal smoothing parameter is chosen automatically by use of the Un-Biased Risk Estimator (UBRE) criterion [19].

Once model (2) has been fitted, the estimated probabilities are used as the new classifier, with the corresponding ROC curve and AUC then being obtained.

It should be noted that GAM-logistic models are also suitable tools for estimating the odds ratio (OR) function, which generalizes the concept of OR per unit increase in the continuous marker. Assuming model (2), $OR(y, y_{ref})$ can be expressed as [11, 20]:

$$OR(y, y_{ref}) = \exp(h(y) - h(y_{ref})) \tag{3}$$

2.2. Validation on Theoretical Scenarios

The proposed methodology introduced in the previous subsection was applied to several controlled scenarios. Specifically, we analyzed different situations in which the distribution of marker *Y* among the healthy and diseased populations was assumed to be known, and assessed the improvement in the classificatory capacity of the transformed marker $\tilde{Y} = P[D \mid Y]$ over that of the crude marker. In all cases a 50% disease prevalence was assumed.

Scenario 1. Normal distributions with the same dispersion in healthy and diseased subjects:

$Y \mid \bar{D} \sim$ N(0, 0.5); $Y \mid D \sim$ N(0.5, 0.5) (see Figure **1**).

Scenario 2. Normal distributions with different dispersion in healthy and diseased subjects:

$Y | \bar{D} \sim$ N(0, 0.5); $Y | D \sim$ N(0.3,0.9) (see Figure **2**).

Scenario 3. Normal distribution in healthy subjects and mixture of normal distributions in diseased subjects: $Y | \bar{D} \sim$ N(1.5, 0.8); $Y | D \sim$ 0.5 x N(0, 0.7) + 0.5 x N(3, 0.6) (see Figure **3**).

2.3. Study Population

We used data drawn from a prospective study of patients who underwent surgical interventions at the *Hospital Clínico Universitario de Santiago* (Santiago de Compostela, Spain) in the period January 1996 through March 1997 [11]. The aim of this study was to ascertain factors associated with the appearance of postoperative infection (POI).

Of the 2353 individuals contained in the database, postoperative infection was detected in 460 during follow-up. Infection was strongly correlated with type of surgery performed. Classification of type of operation was based on Altemeier class [12], which categorizes surgery into the following 4 types: clean; clean-contaminated; contaminated; and dirty to a lesser or greater degree of bacterial contamination. This study evaluated whether glucose levels could predict appearance of infection in the immediate postoperative period. By way of example, and to prevent the presence of possible confounding variables, only the group of clean surgical interventions on nondiabetic individuals was considered for analysis purposes. The final sample comprised 836 patients, 45 of whom presented with POI.

2.4. Statistical Software

The software used for analysis purposes was the R statistics package [21]. Logistic GAMs were fitted using
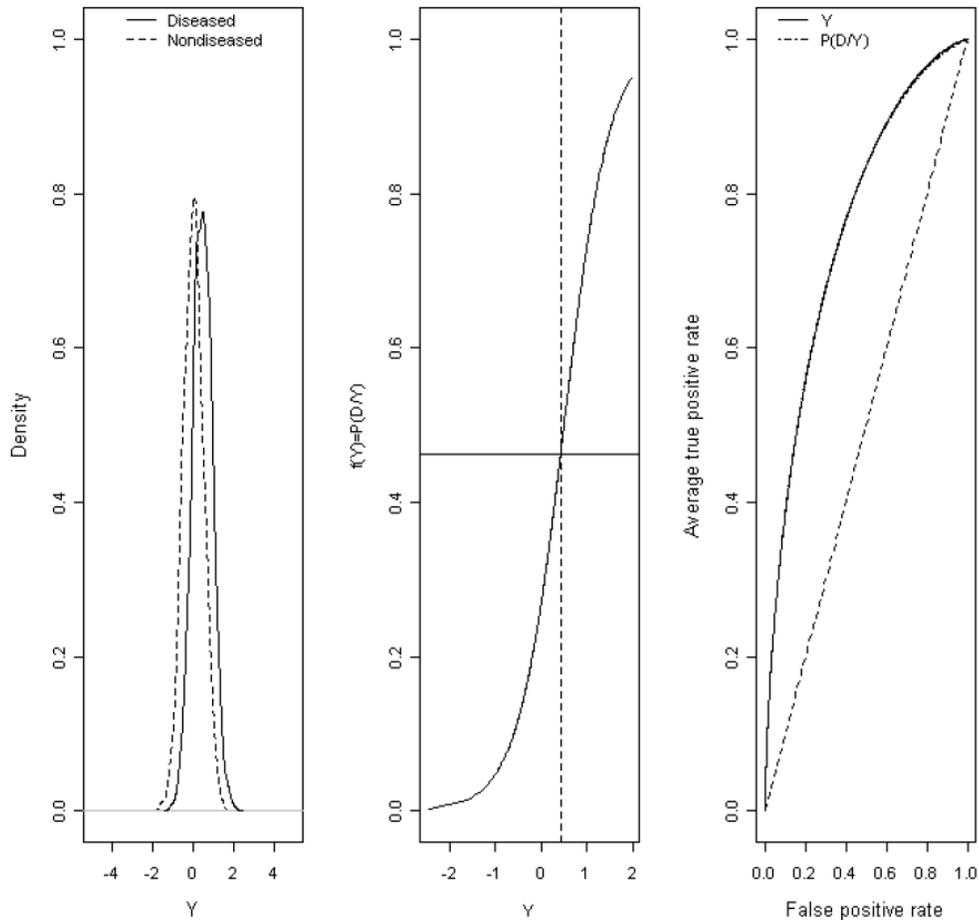
Figure 1: Theoretical scenario 1.

(**a**) Probability distribution of crude marker Y among diseased (solid line) and healthy subjects (dashed line) in Scenario 1. (**b**) Probability of being diseased according to the values of marker Y ($f(Y) = P[Y | D]$). The resulting ROC curves for the crude marker Y and transformed marker $\tilde{Y}$ ($\tilde{Y} = f(Y) = P[Y | D]$) are shown in Figure (**c**). As can be seen, in this case both ROC curves coincide.

the gam function of the mgcv package [22]. ROC curves and their corresponding AUCs were estimated using spline smoothing techniques [23]. This code was obtained from the authors of that paper. The rest of routines, including those necessary for the calculation of optimal cut-off points, are obtainable from the authors on request.

3. RESULTS

3.1. Performance Under Different Theoretical Scenarios

For illustration, we present the results obtained when the methodology proposed in subsection 2.1. was applied to the theoretical scenarios considered previously in subsection 2.2. Figure **1** shows the situation corresponding to Scenario 1. As can be seen, there is no interspersion in the marker's distribution among healthy and diseased subjects, so that the

probability of being ill increases linearly with any increase in the value of *Y*. This situation is consistent with a linear logistic regression model, and so the classic approach is acceptable, with the use of crude and transformed markers, *Y* and $\tilde{Y}$, yielding the same result in terms of ROC/AUC (AUC_Y = $AUC_{\tilde{Y}}$ = 0.76).

Under Scenarios 2 and 3, however, (see Figures **2** and **3**, respectively) there is interspersion in the marker's distribution among healthy and diseased subjects. Hence, the probability of being ill does not show a monotone increasing behavior vis-á-vis the values of *Y*. In such scenarios, therefore, the classification rule assumed in the ROC is not the one that will minimize classification errors, and when the marker is transformed by $\tilde{Y}$, classificatory capacity improves. This can be seen in Figures **2c** and **3c**, which depict the ROC curves corresponding to *Y* and $\tilde{Y}$ under both scenarios. As will be noted, regardless of the *FPF* value chosen, the corresponding *TPF* is always higher (or the same) for the transformed
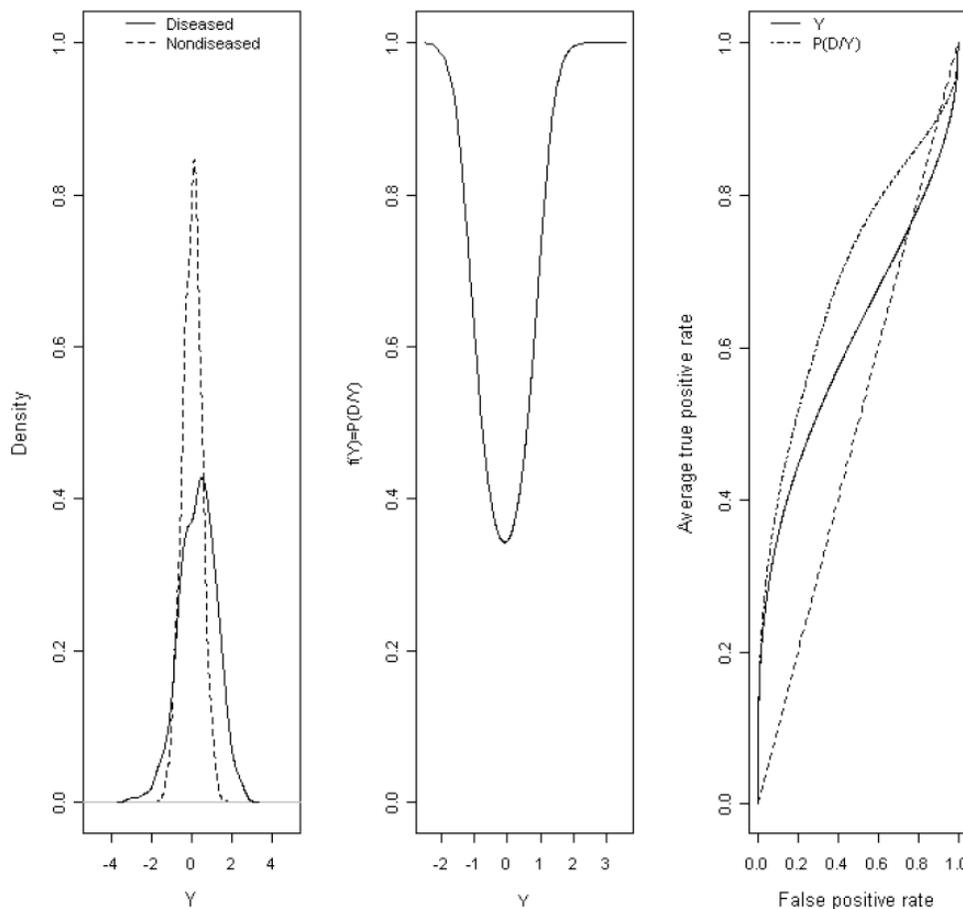
Figure 2: Theoretical scenario 2.

(**a**) Probability distribution of crude marker *Y* among diseased (solid line) and healthy subjects (dashed line) in Scenario 2. (**b**) Probability of being diseased according to the values of marker *Y* ($f(Y) = P[Y|D]$). The resulting ROC curves for the crude marker *Y* (solid line) and transformed marker $\tilde{Y}$ ($\tilde{Y} = f(Y) = P[Y|D]$) (dashed line) are shown in Figure (**c**).
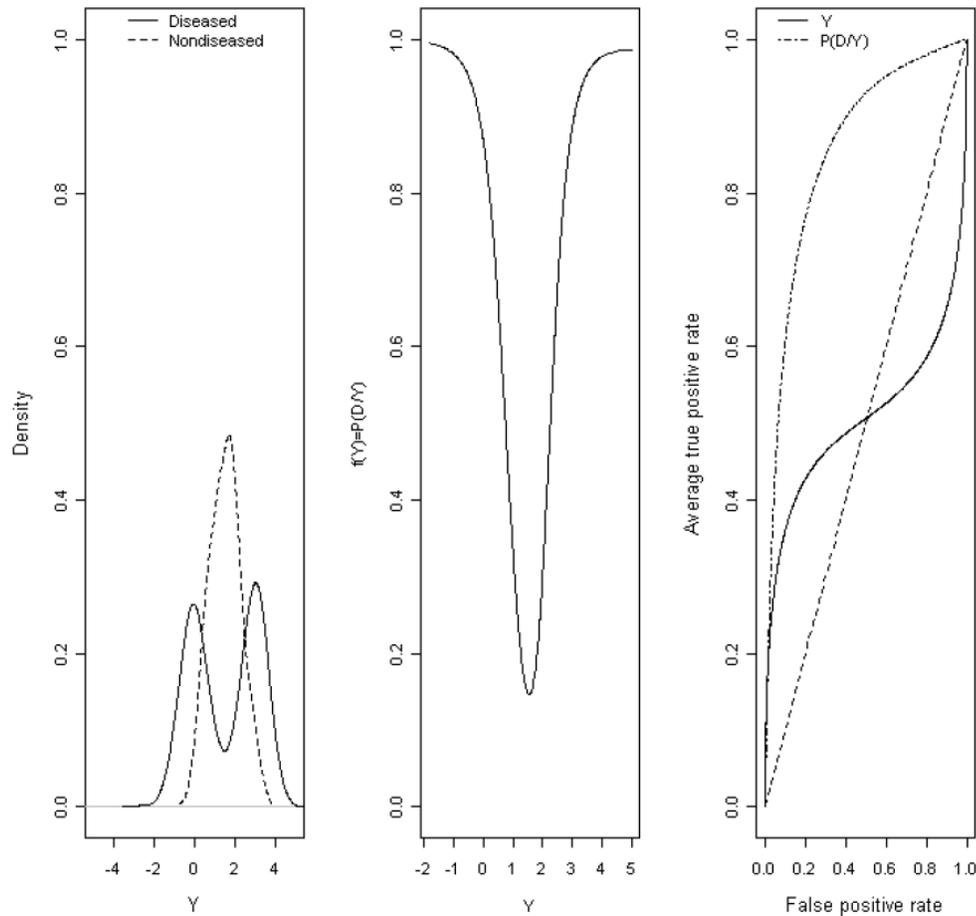
Figure 3: Theoretical scenario 3.

(**a**) Probability distribution of crude marker *Y* among diseased (solid line) and healthy subjects (dashed line) in Scenario 3. (**b**) Probability of being diseased according to the values of marker *Y* ($f(Y) = P[Y|D]$). The resulting ROC curves for the crude marker *Y* (solid line) and transformed marker $\tilde{Y}$ ($\tilde{Y} = f(Y) = P[Y|D]$) (dashed line) are shown in Figure (**c**).

marker. Lastly, whereas *AUC* values for the crude marker *Y* are AUC_Y = 0.62 and AUC_Y = 0.51 in Scenarios 2 and 3, respectively, these change to $AUC_{\tilde{Y}}$ = 0.70 and $AUC_{\tilde{Y}}$ = 0.86 when using the transformed marker $\tilde{Y}$.

In general, the suitability or unsuitability of using the transformed marker for discriminatory purposes is dictated by the form adopted by the function $f(Y) = P[D|Y]$. If this function is strictly monotone increasing (Scenario 1), then using the transformed marker yields no gain in terms of discrimination, since the ROC curve is invariant to this type of transformation [24, 25].

Impact on Selecting Optimal Marker Threshold

Situations in which the risk marker/disease relationship is not monotone increasing (or decreasing) give rise to a new challenge linked to the choice of the "optimal" cut-off point. In such cases, rather than

proposing a single cut-off value as the classification rule, it would seem more reasonable to use an optimal interval that will define individuals with lower (higher) risk of presenting with the disease. Hence, unlike the classical approach, which consists of identifying a single cut-off point, *c*, for the marker, the proposed methodology enables a range of marker values to be obtained which are applicable to a wider variety of risk situations.

Based on the theoretical scenarios presented in the above subsection 2.2, this study proposes the use of transformed marker $\tilde{Y}$ for the choice of optimal cut-off point(s). Cut-off values are calculated in two stages, as follows: a) the optimality criterion is applied to the transformed marker $\tilde{Y} = f(Y) = P[D|Y]$, thereby obtaining the "optimal" probability or risk cut-off: and, b) the crude marker value/s that corresponds/correspond to this probability is/are then computed. For this purpose, the inverse function of *f* is calculated, $f^{-1}(\tilde{Y})$,

for which an explicit form does not always exist but which can be calculated by means of numerical approximation techniques.

In the literature, there are numerous criteria for selecting optimal cut-off values. By way of illustration, the criterion used in this study was that which ensured an *FPF* = 0.20. This choice enables direct observation of the gain in sensitivity resulting from the use of the transformed marker.

As can be seen from Figure **4a**, in Scenario 1 the choice of cut-off point is independent of the use of *Y* or $\tilde{Y}$, with a single cut-off being obtained in both cases, i.e., the value 0.42, corresponding to a *TPF* = 0.56. Under Scenarios 2 and 3, in contrast, (see Figures **4b** and **4c**) the choice of cut-off changes according to whether the decision is based on the crude or transformed marker. In Scenario 2, the optimal cut-off point obtained by the classic method is 0.42, which corresponds to a *TPF* = 0.45. When the choice of the cut-off point is based on the values of the transformed marker, the optimal probability value is 0.48 which

corresponds to a *TPF* = 0.51, i.e., use of the transformed marker yields a 6 % sensitivity gain. The optimal probability value corresponds to an interval of crude marker values ranging from -0.74 to 0.56. The classification rule derived from determination of this optimal interval would lay down that individuals within this interval are to be deemed healthy and the remainder, diseased.

Under Scenario 3, the phenomenon, albeit analogous, is more pronounced in terms of sensitivity gain. Thus, if one were to base oneself on the crude marker and take the marker value showing an *FPF* = 0.20 as being optimal, a cut-off point of 2.17 associated with a *TPF* = 0.46 would be obtained; on the other hand, if the transformed-marker point with an *FPF* = 0.20 were taken as optimal, a value of 0.68 would be obtained, which has a *TPF* = 0.77 and yields an optimal crude value interval (0.46, 2.51). In other words, the latter case would result in a 31 % sensitivity gain. In brief, in any case where the risk is not monotone (Scenarios 2 and 3), applying the optimal cut-off point selection criterion to the transformed
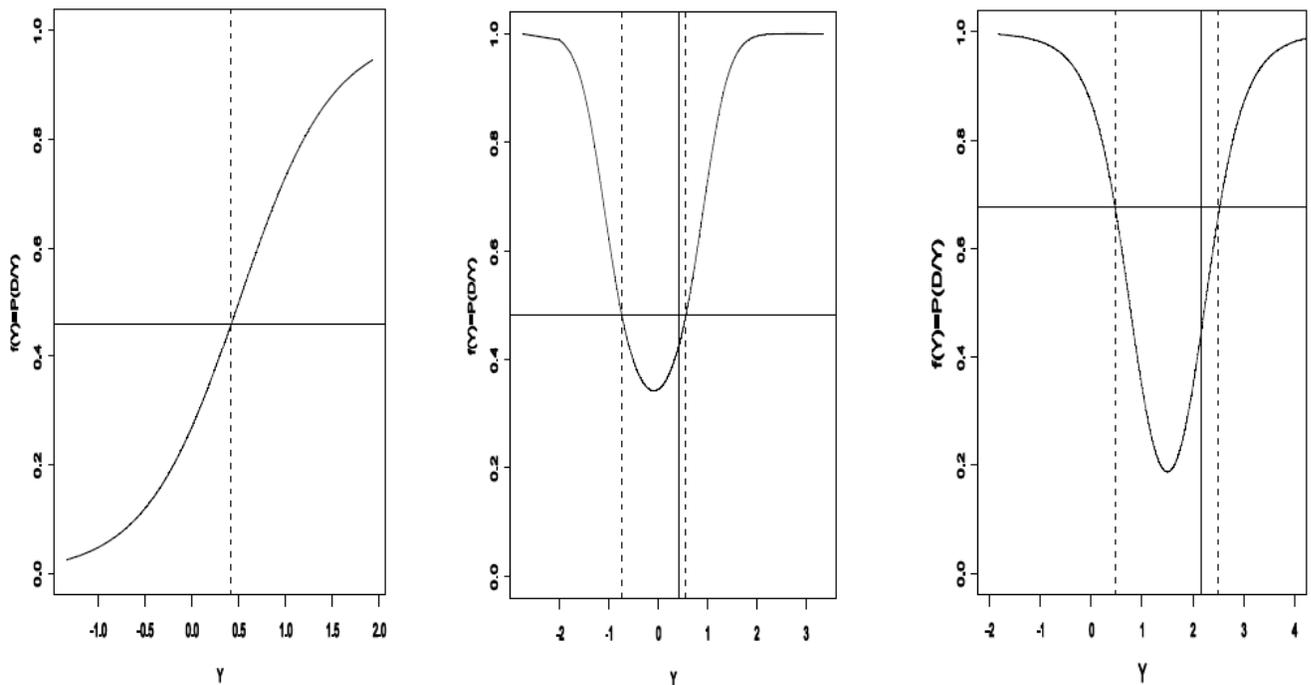
Figure 4: Choice of optimal cut-off points of marker Y in theoretical scenarios.

Choice based on crude marker *Y*, and transformed marker $\tilde{Y} = f(Y) = P[Y|D]$ in theoretical scenarios 1, 2 and 3 (figures (**a**), (**b**) and (**c**), respectively). In each figure: solid horizontal line depicts optimal cut-off point obtained on application of criterion of optimality to transformed marker $\tilde{Y}$; dashed vertical lines depict extremes of the interval of optimal cut-off points derived from the above optimal cut-off point; and, solid vertical line depicts optimal cut-off point obtained on application of criterion of optimality to *Y*. Under Scenario 1, a single optimal cut-off point is obtained, which is independent of the application of the criterion of optimality to *Y* or $\tilde{Y}$. Under Scenario 2, the optimal cut-off point obtained on application of criterion of optimality to *Y*, is contained in the interval of the optimal cut-off point derived from the optimal risk on the basis of the transformed marker $\tilde{Y}$. Under Scenario 3, the pertinent optimal cut-off point obtained on application of criterion of optimality to *Y*, does not belong to the optimal interval derived from the optimal risk.

marker ensures that more sensitive values are obtained for any given value of specificity.

3.2. Application to Real Data

The methodology introduced in subsection 2.1. is applied here to the real data from a prospective study previously presented in subsection 2.3 performed for evaluating the discriminatory capacity of plasma glucose as postoperative infection biomarker. Specifically, a GAM model was fitted, given by the expression:

$$P\left[POI\mid Glu\cos e\right]=g^{-1}(\alpha+h(Glu\cos e))=\frac{\exp(\alpha+h(Glu\cos e))}{1+\exp(\alpha+h(Glu\cos e))}$$

In above expression (4), *Glucose* is the continuous variable that represents plasma glucose levels, *POI* is the binary indicator variable of the presence (*POI* = 1) or absence (*POI* = 0) of postoperative infection, and $h(.)$ is an unknown smooth function.

To assess glucose's classificatory capacity, ROC curves and the AUC were calculated for both crude and GAM-transformed glucose markers. Similarly, glucose cut-off points associated with a value of *FPF* = 0.20 were designated as optimal. This criterion was applied to: (a) the crude marker (*Glucose*); and (b) to the marker transformed by the model (4).

Figure **5a** shows the kernel-type estimate of marker density among healthy and diseased subjects. As will be seen, glucose distribution in healthy and diseased subjects displayed a high level of interspersion, with characteristics similar to those displayed under Scenario 2. Individuals with *POI* registered extreme glucose values (high and low) compared with those who failed to develop postoperative infection (intermediate values).

Figure **5b** depicts the *ln OR* function for the reference value y_{ref} = 95 mg/dL, for the model given by expression (3). As can be seen, rather than being
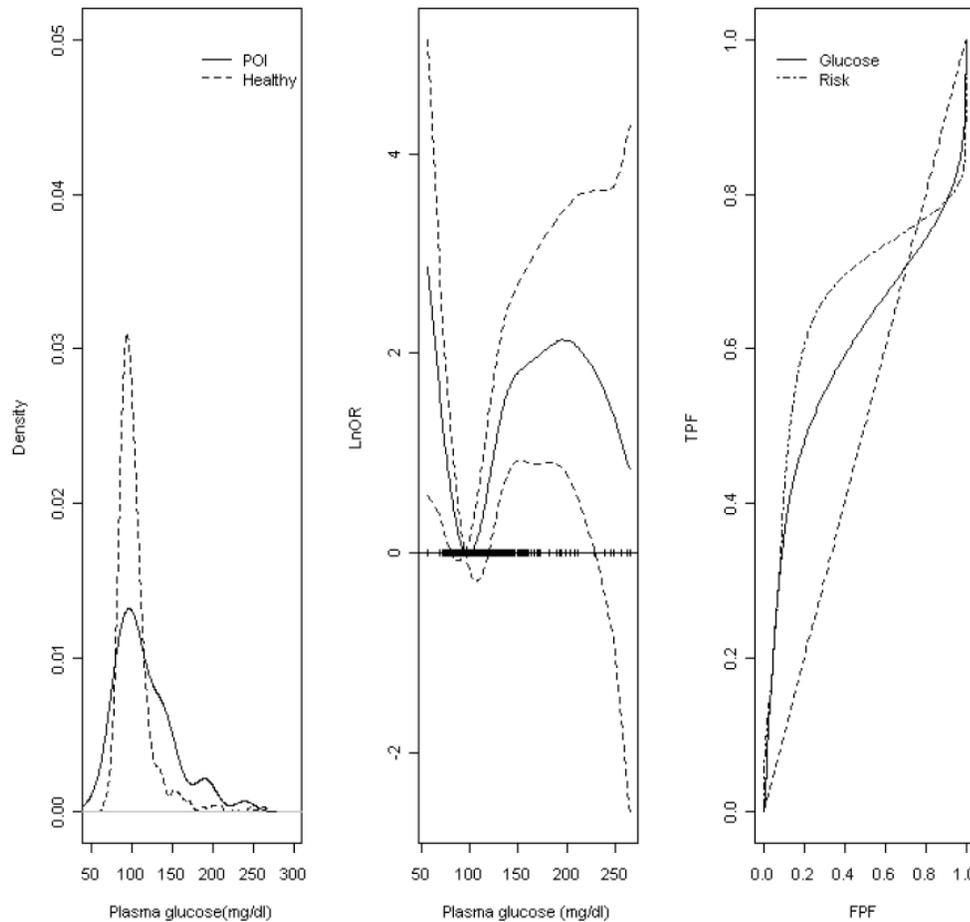
Figure 5: Analysis of plasma glucose as postoperative infection (POI) marker.

(**a**) Kernel-type estimate of plasma-glucose marker density in the presence and absence of postoperative infection (*POI*); (**b**) estimation of *LnOR(y; yref)* (y_{ref} = 95) between glucose and *POI*; and, (**c**) ROC curves for glucose and for glucose transformed by the GAM fitted.

linear, the relationship between plasma glucose and *POI* shows a "spoon" shape [11, 26]. In other words, situations of hypo- and hyperglycaemia pose a higher risk of suffering from *POI* than do intermediate values.

The ROC curves for the crude glucose marker and the GAM-transformed marker (see (4)) are shown in Figure **5c**. The AUC obtained for the glucose marker was 0.60 and its corresponding 95% bootstrap confidence interval (CI) [27] was (0.49, 0.70). In the light of this result, there would seem to be no evidence to indicate that glucose values can be used for discerning *POI.* However, the AUC (CI) for the transformed marker was 0.69 (0.60, 0.78) with the value of 0.5 lying outside this interval. Although the AUC obtained was not high, these results nonetheless allow for plasma glucose to be regarded as a possible biomarker for *POI.*

Lastly, when the optimality criterion was applied to the crude marker values, the resulting glucose cut-off value was 112 mg/dL (*FPF*=0.20, *TPF*=0.44). When this same optimality condition was applied to the transformed glucose marker, an optimal probability value of 0.05 (*FPF*=0.20, *TPF*=0.52) was obtained, which corresponds to crude glucose values of 82 mg/dL and 117 mg/dL. As in theoretical Scenarios 2 and 3, the choice of the optimal cut-off point varies according to whether the optimality criterion is based on the glucose values or on the risk associated with each such value. In this case, basing oneself on the transformed marker results in a 8% gain in sensitivity.

4. DISCUSSION

This study shows: firstly, how the use of GAMs improves the classificatory capacity of a potential continuous marker, particularly in situations where there is a high level of interspersion in the marker's distribution among the subpopulations of healthy and diseased subjects; and secondly, how this type of modeling may modify the choice of suitable cut-off points.

Situations of nonlinearity in the marker/outcome relationship tend to arise relatively frequently in the sphere of clinical-epidemiologic research. For instance, it is frequent for U- or J-shaped relationships to be found when analyzing the appearance of cardiovascular events and plasma glucose [28] or hormone levels [29]. In the different theoretical distribution scenarios analyzed, as well as in the application to real data outlined above, classificatory capacity is seen to improve in terms of ROC/AUC when

the transformed marker is used in situations that reflect a nonlinear risk marker/disease relationship. An additional advantage of the method proposed here is that it provides a tool for calculating more appropriate cut-off points, inasmuch as it takes the relationship between marker values and clinical outcomes into account. Hence, in such U- or J-shaped relationships, two values will be selected that will delimit an optimal interval within which there is a lower probability of being ill.

The criterion used, by way of illustration, for the choice of cut-off points in this study was for a designated *FPF* value to be set. Similarly, other optimal cut-off selection criteria could have been used, such as the approach that simultaneously maximizes sensitivity and specificity [30] or the method based on maximization of the Youden index [10, 31]. Using the proposed criterion, we obtained an interval of lower *POI* risk for a plasma glucose range of 81 to 117 mg/dL. These cut-off points are in line with the results of published studies on risk of infection and glucose values [32]. To our knowledge, this is the first time an optimal interval -as opposed to a single point- has been proposed for enhancing the accuracy of a continuous marker. However, caution must be exercised when choosing the threshold or range of thresholds. The threshold selection criterion should be based on the clinical context, and may involve weighing the expected costs against the benefits associated with a high-risk designation. Other clinical settings may involve an entirely different balance between disease severity, treatment efficacy and side-effects, leading to different choices according to the importance of sensitivity versus specificity. One of the strengths of this proposal is that the proposed methodology can be easily used by clinicians and epidemiologists. We used spline smoothing methods to estimate ROC curves and related measures but other available tools, such as empirical or kernel smoothing techniques, can equally be used.

The results obtained in this study can be extended to other biomedical scenarios, such as survival analysis. In this context, the outcome variable is deemed to be time-dependent, and the concepts of ROC curve and AUC apply to this situation [33, 34]. The fitting of Cox regression models [35] used for survival analysis and the modeling of hazard ratios may likewise be performed by flexible methods analogous to those described for logistic regression [28, 36], so that nonparametric modeling can also be used to obtain more suitable markers in this context.

A further interesting task linked to the methodology presented here is its extension to a multivariate context. ROC analysis can also be used in conjunction with GAM logistic models in the context of the use of combined markers for diagnosis purposes [37].

5. CONCLUSIONS

It is important to stress the need to question linearity in marker-outcome relationships: firstly, because failure to do so may lead to erroneous conclusions; and secondly, because the use of statistical tools that make for greater flexibility (e.g., GAMs) can optimize the classificatory capacity of a potential marker from the standpoint of ROC analysis.

ACKNOWLEDGEMENTS

REFERENCES

[1] Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283-98.
http://dx.doi.org/10.1016/S0001-2998(78)80014-2

[2] Swets JA, Pickett RM. Evaluation of diagnostic systems: Methods from signal detection theory. New York: Academic Press 1982.

[3] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.
http://dx.doi.org/10.1148/radiology.143.1.7063747

[4] McIntosh MW, Pepe MS. Combining several screening tests: optimality of the risk score. Biometrics 2002; 58: 657-64.
http://dx.doi.org/10.1111/j.0006-341X.2002.00657.x

[5] Lustres-Pérez V, Rodríguez-Álvarez MX, Pazos-Pata M, Cadarso-Suárez C, Fernández-Pulpeiro E. The application of Receiver Operating Characteristic (ROC) methodology in biological studies of marine resources: sex determination of Paracentrotus lividus (Lamarck, 1816). SORT 2010; 34: 239-48.

[6] Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman and Hall 1990.

[7] Mazumdar M, Glassman JR. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. Stat Med 2000; 19: 113-32.
http://dx.doi.org/10.1002/(SICI)1097-0258(20000115)19:1<113::AID-SIM245>3.0.CO;2-O

[8] Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 1994; 86: 829-35.
http://dx.doi.org/10.1093/jnci/86.11.829

[9] Lausen B, Schumacher M. Evaluating the effect of optimized cutoff values in the assessment of prognostic factors. Comput Stat Data Anal 1996; 21: 307-26.
http://dx.doi.org/10.1016/0167-9473(95)00016-X

[10] Klotsche J, Ferger D, Pieper L, Rehm J, Wittchen HU. A novel nonparametric approach for estimating cut-offs in continuous risk indicators with application to diabetes epidemiology. BMC Med Res Methodol 2009; 9: 63.
http://dx.doi.org/10.1186/1471-2288-9-63

[11] Figueiras A, Cadarso-Suárez C. Application of nonparametric models for calculating odds ratios and their confidence intervals for continuous exposures. Am J Epidemiol 2001; 154: 264-75.
http://dx.doi.org/10.1093/aje/154.3.264

[12] Altemeier W. Surgical infections: incisional wounds. In: Bennet JV, Brachman P, editors. Hospital infections. Boston: Little, Brown and Company 1979.

[13] Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypothesis. Philos Trans R Soc Lond A 1933; 231: 289-337.
http://dx.doi.org/10.1098/rsta.1933.0009

[14] Eilers P, Marx B. Flexible smoothing with B-splines and penalties. Stat Sci 1996; 11: 89-121.
http://dx.doi.org/10.1214/ss/1038425655

[15] Wood SN. Thin plate regression splines. J R Stat Soc Series B Stat Methodol 2003; 65: 95-114.
http://dx.doi.org/10.1111/1467-9868.00374

[16] Lang S, Brezger A. Bayesian P-splines. J Comput Graph Stat 2004; 13: 183-212.
http://dx.doi.org/10.1198/1061860043010

[17] McCullagh P, Nelder J. Generalized linear models. 2nd ed. London: Chapman and Hall 1989.
http://dx.doi.org/10.1007/978-1-4899-3242-6

[18] Wand MP, Jones MC. Kernel smoothing. London: Chapman and Hall 1995.
http://dx.doi.org/10.1007/978-1-4899-4493-1

[19] Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc 2004; 99: 673-86.
http://dx.doi.org/10.1198/016214504000000980

[20] Zhao LP, Kristal AR, White E. Estimating relative risk functions in case-control studies using a nonparametric logistic regression. Am J Epidemiol 2006; 144: 598-609.
http://dx.doi.org/10.1093/oxfordjournals.aje.a008970

[21] R Development Core Team. R: A language and environment for statistical computing, version R.3.1.3. R Foundation for Statistical Computing, Vienna, Austria 2015. URL http://www.R-project.org/

[22] Wood SN. Generalized additive models, an introduction with R. Boca Raton, Florida: Chapman and Hall/CRC 2006.

[23] Du P, Tang L. Transformation-invariant and nonparametric monotone smooth estimation of ROC curves. Stat Med 2009; 28: 349-59.
http://dx.doi.org/10.1002/sim.3465

[24] Swets JA, Tanner WPJ, Birdsall TG. Decision processes in perception. Phychol Rev 1961; 68: 301-40.
http://dx.doi.org/10.1037/h0040547

[25] Egan JP. Signal detection theory and ROC Analysis. New York: Academic Press 1975.

[26] Sáez M, Cadarso-Suárez C, Figueiras A. np.OR: an S-Plus function for pointwise nonparametric estimation of odds-ratios of continuous predictors. Comput Methods Programs Biomed 2003; 71: 175-79.
http://dx.doi.org/10.1016/S0169-2607(02)00076-7

[27] Efron B. Bootstrap methods: Another look at the jackknife. Ann Stat 1979; 7: 1-26.
http://dx.doi.org/10.1214/aos/1176344552

[28] Cid-Álvarez B, Gude F, Cadarso-Suárez C, *et al*. Admission and fasting plasma glucose for estimating risk of death of diabetic and nondiabetic patients with acute coronary syndrome: nonlinearity of hazard ratios and time-dependent comparison. Am Heart J 2009; 58: 989-97.
http://dx.doi.org/10.1016/j.ahj.2009.10.004

[29] Bertone-Johnson ER, Tworoger SS, Hankinson SE. Recreational physical activity and steroid hormone levels in postmenopausal women. Am J Epidemiol 2009; 170: 1095-104.
http://dx.doi.org/10.1093/aje/kwp254

[30] Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. Phys Ther 1999; 79: 939-50.

[31] Greiner M, Pfeiffer D, Smith RD. Principals and practical application of the receiver operating characteristic analysis for diagnostic tests. Prev Vet Med 2002; 45: 23-41.
http://dx.doi.org/10.1016/S0167-5877(00)00115-X

[32] Van den Berghe G, Wouters P, Weekers F, *et al*. Intensive insulin therapy in the critically ill patients. N Engl J Med 2001; 345: 1359-67.
http://dx.doi.org/10.1056/NEJMoa011300

[33] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000; 56: 337-44.
http://dx.doi.org/10.1111/j.0006-341X.2000.00337.x

[34] Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005; 61: 92-105.
http://dx.doi.org/10.1111/j.0006-341X.2005.030814.x

[35] Cox DR. Regression models and life-tables (with discussion). J R Stat Soc Series B Stat Methodol 1972; 34: 187-220.

[36] Cadarso-Suárez C, Meira-Machado L, Kneib T, Gude F. Flexible hazard ratio curves for continuous predictors in multi-state models: an application to breast cancer data. Stat Modelling 2010; 10: 291-314.
http://dx.doi.org/10.1177/1471082X0801000303

[37] Lado MJ, Cadarso-Suárez C, Roca-Pardiñas J, Tahoces PG: Using generalized additive models for construction of nonlinear classifiers in computer-aided diagnosis systems. IEEE Trans Inf Technol Biomed 2006; 10: 246-53.
http://dx.doi.org/10.1109/TITB.2005.859892