Journal of Physical Sciences,
Vol. 1, No. 1, 77 – 85, April 2010

Mixed Variable Logistic Regression Model for Assessing Diagnostic Markers in Prostate Cancer.

Fidelis .I. Ugwuowo
Department of Statistics, University of Nigeria, Nsukka, Enugu state, Nigeria.
fiugwuowo@yahoo.com

Abstract

When a diagnostic test is based on some observed variable, an assessment of the overall value of the test can be made through the use of receiver operating characteristic (ROC) curve. We present the methodology for assessing some dichotomous and continuous variable in a diagnostic process. The approach uses logistic regression (LR) model to obtain the best linear combination of markers. The area under the ROC curve of this combination is maximised among all possible linear combinations. We further demonstrate using confusion matrix and Youden Index (YI) that the discriminating power of this multiple marker combination is higher than for all other combinations. The corresponding optimum critical threshold value to the Youden Index is derived for all possible combinations. Finally, an illustration of this methodology is given using prostate cancer diagnostic data from University of Nigeria Teaching Hospital (UNTH) Enugu.

Key words: *Diognostic Marker; Logistic Regression; Prostate Cancer; Receiver Operating Characteristic Curve; Youden Index.*

1. Introduction

Mixed variable logistic regression (LR) model for binary response data has frequently been applied in analysis of data with nested levels (Ten Have, et al., 1999; and Hedeker, 1997). One example given by Ten Have, et al, is the fitting of what they referred to as three-level probit models to educational binary data with students and classes or with classes and schools serving as nested clustering factors. The other areas of application of LR model in analysing binary data which arise in studying relationships between diseases and environment or genetic characteristics, can be found in Breslow and Day (1980, ch.4), Prentice and Pyke (1979) and Farewell (1979). One advantage of this model as pointed out by Cox and Snell (1989) is that it presupposes a stable statistical relation such that once a vector of explanatory variables is given, then the probability that an individual belongs to one of the two groups is determined. The distribution of the variables is therefore irrelevant. This feature renders the LR model more robust than the linear discrimination analysis which has been shown to be more efficient in combining markers if the normality assumption on the variables is satisfied (Ruiz-Valasco (1991), Efron (1975).

In this paper, we provide the optimal linear combination of markers involving some dichotomous and continuous variables such that the area under the corresponding ROC curve of the combination is maximised. Many authors have recommended the use of ROC curve in the assessment of diagnostic markers. A marker's usefulness is generally assessed based on its 'sensitivity' and 'specificity' defined respectively as the probability that the test result is positive given that the subject is truly diseased and the probability that the test result is negative given that the subject is truly non-diseased (Pepe, 1997). This paper is devoid of distributional assumptions or transformations in estimating the ROC curve (Su and Liu, 1993; Zou & Hall, 2000, 2002; Faraggi, 2003). We have adopted a simple and direct modelling approach, which we consider an advantage.

For statistical analysis, a recommended index of accuracy associated with a ROC curve is the area under the curve (Swets and Pickett 1982). The area under the population ROC curve represents the probability that when the variable is observed for a randomly selected individual from the diseased population and a randomly selected individual from the normal population, the result will indicate that the diseased value is higher than the normal value. Bamber (1975) has proved that the area under the ROC curve could be evaluated as $A = P(X > Y)$ where X represents the diseased population and Y the healthy population. Beam and Wieand (1991) developed a statistical method for the

comparison of a discrete with several continuous diagnostic tests. They noted that the comparison of areas under the ROC curve could be intrinsically biased against the discrete test. Some regression related methods for estimating the ROC curve and evaluating the effect of factors that may influence test accuracy could be found in Pepe (1998,2000), and Faraggi (2003).
Also in use as a measure of markers discriminating accuracy is the Youden Index, defined as max $\{P + q - 1\}$, where P is the specificity and q is the sensitivity. The maximum is taken over all P's and q's on the ROC curve or equivalently over all possible threshold values. The critical threshold value C^* , which corresponds to the Youden Index, is frequently used as a criterion for determining whether subjects are considered healthy or diseased if their observed marker value is less than or equal to, or greater than C^* respectively. Faraggi (2003) and Greiner, et al. (2000) have provided a recent discussion on this and other criteria for obtaining the critical threshold value.

Nakas et al., (2003) showed that goodness-of-fit of uniformity of the distribution of the false positive (true positive) rates can be used instead of test based on the area index. He used a semi-parametric approach based on a completely specified distribution of marker measurements for either the healthy or diseased and extended it to the one and two marker case when neither distribution is specified (non-parametric case). He concluded that ROC based tests is more powerful than goodness-of-fit test for location differences between the distribution of healthy and diseased subjects. He further asserted that ROC-based tests are less powerful when location-scale differences exist. Various methods of evaluating and comparing the performances of two or more diagnostic markers have been presented based on differences of areas under ROC curves, some examples can be found in Delong, et al., (1988), Hanley & McNeil (1982,1983) and McClish (1987). Some modifications of these comparisons at some fixed level and weighted averages of sensitivity were also considered by Linnet (1987) and Wieand, et al., (1989). This approach in combining dichotomous and continuous variable through the use of LR model has not been considered in the literature. Su and Liu, (1993) who used Fisher's linear discriminating function noted that different markers are usually sensitive to different aspects of disease in real situation. They further stated that it is important to use two or more good diagnostic markers simultaneously so that one may obtain a new diagnostic marker with higher sensitivity. This paper is aimed at showing this assertion using a different approach.

The paper is planned such that section 2 introduces the motivating example that discusses the data collected on prostate cancer patients. Section 3 is devoted to the fitting of the LR model while section 4, discusses the estimation of ROC curve using the best linear combination. In section 5, the methodology developed is applied to the prostate cancer data, together with other methods of prediction accuracy. The results and discussion arising from the analysis are given in section 6, which is then followed by concluding remarks.

2. Motivating Study

Prostate cancer like any other cancer is a very serious disease causing considerable mortality and morbidity among the population. This disease could be localised if diagnosed at the incipient stages. The treatment regime to be adopted for patients who have been diagnosed as having cancer of the prostate is crucially dependent upon whether or not the cancer has spread to the surrounding nodes. However, a laparotomy (a surgical incision into the abdominal cavity) may be performed to ascertain the extent of nodal involvement but due to the rising cost of medical treatment, it has become necessary for an alternative method of diagnosis that reduces the cost of treatment to be used. Some variables, which can be measured, are indicative of nodal involvement; these are x-ray examination result, tumour size and serum acid phosphatase (Serum prostate specific antigen). Studies have shown that serum acid phosphatase usually rise in men who have prostate cancer and other infections of the prostate (Renard, et al., 2003).
Cancer of the prostate is conventionally treated as a two-category (binary) response variable in most diagnostic procedure; i.e. a sufferer is either diagnosed as having the nodes affected or not having the nodes affected. This is more so since the correct treatment regime depends on the extent of nodal involvement. In a related work on pancreatic cancer, Pepe (2000) developed a regression methodology for identifying factors

that can influence the discriminatory capacity of a biomarker screening test or, more generally, that of a medical diagnostic test and applied it to the particular problem of comparing two serum antigen markers.

This study involves a sample of 259 prostate cancer patients at University of Nigeria Teaching Hospital (UNTH) Enugu. Three tests were conducted on all the patients to determine the extent of nodal involvement. The outcome of the three diagnostic markers were recorded for continuous variables and coded for discrete variables. The results of x-ray examination were assigned 0 if negative and 1 if positive, the size of the tumour as determined by a rectal examination is assigned the value 0 if small and 1 if large while the level of serum acid phosphatase were measured in king Armstrong units. The first two variables are recorded as dichotomous while the third is recorded as a continuous variable. The data collected were divided into two populations or groups. The first population comprises patients who were diagnosed of nodal involvement after a surgical intervention. The second population comprises of patients who were not diagnosed of nodal involvement. The two categories of patients were identified with 136 most severe cases having nodal involvement and 123 others without nodal involvement. It is therefore of interest to study the predictive capacity of a linear combination of these mixed variables in prostate cancer diagnosis.

3. Modelling Approach

Define Y_{ij} to be a binary variable for the j-th biomarker in i-th individual. For each individual, we also have a covariate vector X_{ij} with some dichotomous and continuous variable. The outcome of the test for every individual in the sample is either diseased or non-diseased. A LR model is commonly used to model the relationship between Y_{ij} and X_{ij}. If P_i is the probability of ith individual coming from diseased population, then according to Lee, E.T. (1992),

$$P = \Pr(Y_{ij} = 1 / X_{ij})$$

$$= \frac{\exp\left(\beta_0 + \sum_{j=1}^{r} \beta_j x_{ij}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{r} \beta_j x_{ij}\right)} \quad (1)$$

where Y_{ij} is the binary outcome for jth biomarker in individual i.

X_{ij} is the covariate vector for jth biomarker in individual i.

β_0 is the scale parameter and β_j, j =1...r is the other parameters.

We can then obtain the logit of (1) as,

$$\log_e \frac{P_i}{1 - P_i} = \beta_0 + \sum_{j=1}^{r} \beta_j x_{ij} \quad (2)$$

The parameters β_j's are estimated through the maximisation of the log-likelihood as suggested by Cox (1970) and Hosmer & Lemeshow (1989).

Some statistical tests to determine the adequacy of the LR model were carried out. The first was the likelihood ratio test where the hypothesis that the parameters of the model are zero was tested, i.e. test $H_o : \beta_1 = \beta_2 = \Lambda = \beta_r = 0$. This is done using the G statistic given by $2(l_r - l_0)$ which has a χ^2 distribution with r degrees of freedom (Belsley, 1991). l_r is the log-likelihood for the full model with the r parameters determined and l_0 is the log-likelihood with the constant term only for the 'null' model. If the likelihood ratio test is found to be significant at an $\alpha = 0.05$ then we can proceed to test each of the individual parameters. This could be done using Wald test (Rao, 1973), with a test statistics given by,

$$Z_w = \frac{\hat{\beta}_j}{s.e\,\hat{\beta}_j} \sim N(0,1) \quad (3)$$

This is a two-sided test which would reject the null hypothesis if $|Z_w| > Z_{\alpha/2}$.

To determine the accuracy of fit of the model, it is often useful to consider misclassification rates (Anderson & Phillips (1981)) between the two groups. Table 1 show the notation in obtained frequencies when a binary classifier is used to

predict the class of unseen observations in a confusion matrix (Gordis, 1996).

Table 1. Notation for the binary classification matrix.

Predicted / Observed	Group1	Group2
Group1	n_{11}	n_{12}
Group2	n_{21}	n_{22}

We can obtain an expository value of sensitivity and specificity from this matrix by using the formula,

$$\text{Sensitivity} = \frac{n_{22}}{n_{21}+n_{22}} \quad \text{and} \quad \text{Specificity} = \frac{n_{11}}{n_{11}+n_{12}} \tag{4}$$

where the sensitivity is the diseased proportion that were correctly classified and specificity is the non-diseased proportions that were also correctly classified as such. Both values are expected to be high for the classification to be reliable.

The final consideration of model adequacy is by Lemeshow & Hosmer goodness-of –Fit test. The test statistic given by Lemeshow & Hosmer (1982) is

$$C = \sum_{k=1}^{g} \frac{(O_k - E_k)^2}{n_k \bar{P}_k (1-\bar{P}_k)} \tag{5}$$

where $E_k = \sum_{j=1}^{n_k} P_j$ is the expected number of successes for the kth group.

O_k is the observed number of successes in the kth group and $\bar{P}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} P_j$ is the estimated probability of the kth group and g is the total number of groups.

Under the null hypothesis that the model is adequate, the distribution of C is approximated by the chi-square distribution with g-2 degrees of freedom. Finally, the correlation coefficients between the variables were obtained to determine their degree of association. If any two variables are strongly correlated, one will be selected for inclusion in the model.

4. Roc Curve Estimation

As has been observed earlier, the LR model makes predictions as a probability rather than a binary value hence the estimation of ROC curve is more simplified.

Given that for the sample of individuals under consideration, define

$$y_{ij} = \begin{cases} 1, if \quad diseased \\ 0, otherwise \end{cases}$$

Upon each individual in the sample we record the covariate vector X_{ij}, with some dichotomous and continuous variable. As usual we assume that higher values of the variable are associated with the diseased population and lower values with the non-diseased. To use the outcomes obtained from these subjects, we assumed that the two populations are identically and independently distributed (iid) with survivor functions for the diseased group given by $F(c) = p(X_{ij} > c \mid \text{diseased})$ and for the non-diseased group $H(c) = p(X_{ij} > c \mid \text{non-diseased})$, where c is the threshold value. Then equivalently,

$$F(c) = 1 - p(y_{ij} = 1 \mid X_{ij} = c) \quad \text{and}$$
$$H(c) = 1 - p(y_{ij} = 0 \mid X_{ij} = c) \tag{6}$$

The ROC curve is the monotone increasing function in [0,1], obtained from the diagnostic markers by generating values of sensitivity {F(c)} and the corresponding 1-specificity {H(c)} to form a locus [{F(c), H(c)}, c ∈ (-∞, ∞)] (Pepe, 2000; Zou, et al., 2002, and Heagerty et al., 2000).

Hence using the trapezoidal rule (Bamber, 1975), the area under the ROC curve A(c) can then be estimated given the threshold values. The higher the ROC curve is in the quadrant [0,1] x [0,1], the better is its capacity for discriminating diseased from non-diseased individuals.

5. Application to the Prostate Cancer Data

In this section, we aim at applying the proposed methodology to the prostate cancer data

introduced in section (2). We have identified the covariate vector $X_{ij} = (x_{i1}, x_{i2}, x_{i3})$, with X_{i1} and X_{i2} being dichotomous and X_{i3} continuous variable. Using equation (1) and (6), the marker's sensitivity and 1-specificity for the threshold value c is given as

$$F(c) = 1 - p(y_{ij} = 1 \mid x_{1j} = 1, x_{2j} = 1, x_{3j} = c) = \frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 c)}{1 + \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 c)} \quad (7)$$

and,

$$H(c) = 1 - P(y_{ij} = 0 \mid x_{1j} = 0, x_{2j} = 0, x_{3j} = c) = \frac{\exp(\beta_0 + \beta_3 c)}{1 + \exp(\beta_0 + \beta_3 c)} \quad (8)$$

The ROC curve for the linear combination of markers using the LR model is shown in Fig. 1. The interpretation of the area could be given as the estimated probability that a randomly selected individual with nodal involvement will be assigned a higher predicted probability by the logistic model than another randomly selected individual without nodal involvement. However, the values of the dichotomous variable will usually assume lower values for non-disease occurrence and higher values for disease occurrence.

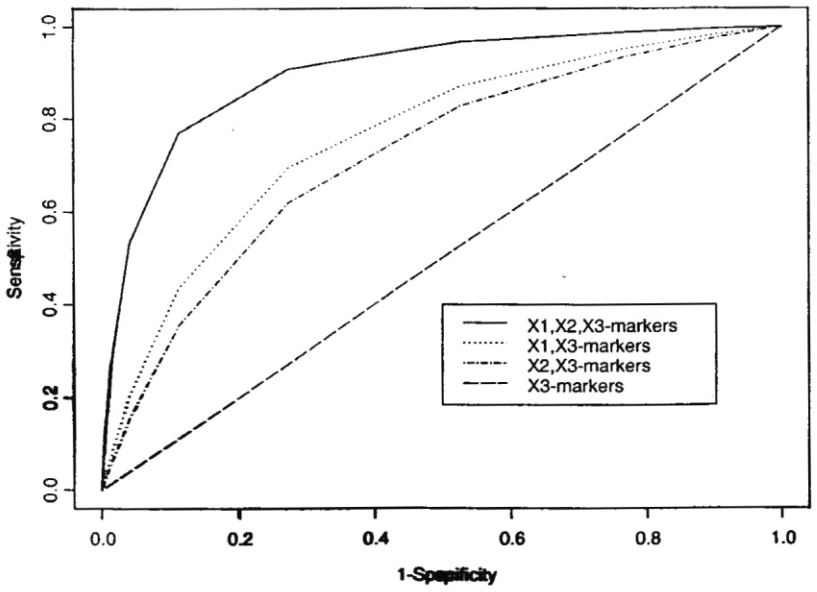

Fig. 1. ROC Curves with the best linear combination of the variables.

5.1 The Youden Index.

The Youden index (YI), (Youden, 1950, Faraggi, 2003) has also been suggested as an appropriate measure of accuracy for the diagnostic markers. This can be obtained given the threshold value c as

$$YI(c) = \max_c \{F(c) + H(c)\} - 1 \quad (9)$$

where

$$F(c) + H(c) = \frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 c)}{1 + \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 c)} + \frac{\exp(\beta_0 + \beta_3 c)}{1 + \exp(\beta_0 + \beta_3 c)} \quad (10)$$

The critical threshold value C* is the value of c which brings equation (10) to its maximum value. This can be obtained by setting the log of the first derivative of (10) with respect to c to 0 and solving the resulting equation, giving the solution

$$C^*(c) = \frac{-10\beta_0 - (\beta_1 + \beta_2) - In4}{10\beta_3} \quad (11)$$

Hence, on substitution of the critical threshold, we then obtain the index as,

$$YI(c) = \{F(C^*) + H(C^*)\} - 1 = \left\{ \frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 C^*)}{1 + \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 C^*)} + \frac{\exp(\beta_0 + \beta_3 C^*)}{1 + \exp(\beta_0 + \beta_3 C^*)} \right\} - 1 \quad (12)$$

This index can vary from minus one to plus one, with plus one being perfect accuracy and minus one being the worst accuracy. An evaluation of the Youden index (YI) for different threshold values for the predicted probability from the best linear LR model is shown in Fig. 2. We can observe that the curve clearly decreases very sharply with increase in the threshold value.

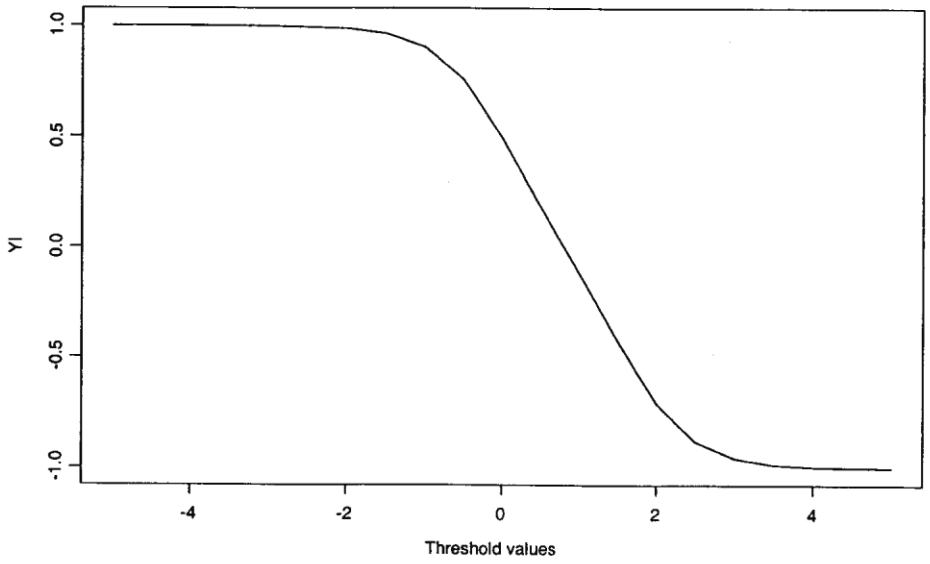

Fig. 2. Youden Index (YI) as function of threshold value c.

6. Results and Discussion

The test for model adequacy using the likelihood ratio test gave a chi-square value of 18.256 at p = 0.0004 which indicates significance at 5% level. Table 2 shows the values of the estimated parameters, the standard error and the result of the Wald test with the significance level indicated. The result shows that all the variables should be included in the model since they are all significant. However, Lee (1992) did show that the significance of p-value should not be the only bases for inclusion of a covariate in the LR model.

. The percent correct classification as given in Table 3 is 72.97%, which is also appreciably high. The corresponding sensitivity and specificity obtained from the confusion matrix is 0.8608 and 0.5248 respectively. These values are pretty high which goes to show that the correct classification for nodal involvement and non-nodal involvement is very high. In addition to the above tests, the Lemeshow & Hosmer goodness-of-Fit test showed a value of C = 7.4279 with a P value of 0.4912 which indicates a high significant value.

The estimates of the area (A) under the ROC curve using the trapezoid rule (Barber, 1975), the Youden Index (YI) and the critical threshold (C*) for the combination of markers are given in Table 4. The area under the ROC curve for the three markers (X_1, X_2, X_3) is 0.8689 which is maximised over all other combinations. The curve also produced the highest sensitivity at every specificity. The ROC curve for single variable X_3 showed a straight line, which implies that with single marker under this model one would expect to obtain 50% proportion correct accuracy by guessing with no prior information. Hence it will be a bad prediction for individuals with nodal involvement and those without nodal involvement.

Table 2. Summary results of the estimated parameters; the standard error, the Wald test and the significance values of the test.

Variables	β_j	S.E	Wald	d.f	Sig.
Constant	0.1069	(0.9894)	0.0117	1	0.9140
X-ray result (x1)	1.8050	(0.6925)	6.7941	1	0.0091
Tumour size (x2)	1.4675	(0.6549)	5.0218	1	0.0250
Acid phosphatase level (x3)	-2.1657	(1.1419)	3.5969	1	0.0579

Table 3. Prediction results in a confusion matrix.

		Predicted		
		Group1	Group2	percent correct
Observed				
	Group1	53	48	52.48%
	Group2	22	136	86.08%
			Overall	**72.97%**

Table 4. Summary result of the area under the ROC curve (A), the Youden Index (YI) and the critical threshold (C*)

Markers	Area(A)	YI	C*
X_1, X_2, X_3	0.8985	-0.0378	0.8689
X_1, X_3	0.7637	-0.0569	0.5301
X_2, X_3	0.7211	-0.0607	0.4522
X_3	0.5000	-0.0692	0.1134

Table 5. The correlation coefficients amongst the three variables X-ray result (X_1), Tumour size (X_2), and Serum acid phosphatase (X_3).

	X_1	X_2	X_3
X_1	1.0000	0.1979	0.0725
X_2	0.1979	1.0000	-0.0741
X_3	0.0725	-0.0743	1.0000

7. Concluding Remarks

The model fitted indicates that on the in [odds] scale, there is a linear relationship between the probability of nodal involvement and the three explanatory variables. This relationship has a different slope and intercept depending on the levels of the biomarkers. Table 5 shows the correlation coefficient between the variables, which are quite low, paving the way for inclusion of all the variables in the LR model.

There has been in use numerous indices for summarising the information contained in the ROC curve but the most popular quantitative index of diagnostic accuracy is the area under the ROC curve. It is generally believed that in one marker situation, the larger the area the more information it contains. Generalising this criterion, we call a linear combination coefficient of markers the best if the area under the ROC curve generated by that combination is the largest among all other linear combinations. In conclusion, one can say that the diagnostic decision is best made using the three biomarkers. However, the x-ray result combined with the acid phosphatase level appears to give more information than the latter's combination with

tumour size. It is worthy of note that the use of a single biomarker could be very misleading since its sensitivity at every specificity is the lowest. The result is further confirmed by the Youden Index, which shows a remarkable decrease in value (Table 4) from the linear combination of three diagnostic markers to one diagnostic marker. Although the values of the Youden index appear to be low, their ranking still corresponds with the areas under the ROC curve.

This formulation has several advantages: (1) they are more robust since the probability that an individual belongs to one of the two groups is determined once the explanatory variables are given. (2) They permit simple construction of a ROC curve. However, one disadvantage of the approach is the limitation on the threshold values of the dichotomous variable. A further consideration of the estimation of ROC curve when all the biomarkers are dichotomous is ongoing.

Acknowledgements

The contribution of TWAS-UNESCO and the host, Indian Statistical Institute, Calcutta that enabled the author to conduct this research work as a 2001-2003 TWAS associate is hereby acknowledged. Thanks are due to Professors Ashis SenGupta and Anup Dewanji both of ISI Calcutta for their useful comments.

References

Anderson, J.A. & Phillips, P.R. (1981) Regression, discrimination and measurement Models for ordered categorical variables. *Applied Statistics*, 30, pp. 22-31.

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

Beam, C.A. & Wieand, H.S. (1991) A Statistical Method for the Comparison of a discrete Diagnostic test with several continuous diagnostic tests. *Biometrics*, 47, pp. 907-919.

Belsley, D. (1991) *Conditioning diagnostics: Collinearity and weak data in regression.* John Wiley and sons.

Breslow, N. & Day N.E. (1980) Statistical methods in cancer research 1, the analysis of Case-control studies Lyon: IARC

Cox, D.R. (1970) *Analysis of Binary Data.* Methuen, London.

Cox, D.R. & Snell, E.J. (1989) *Analysis of Binary Data* (2nd Ed.) London: Chapmanand Hall.

Delong, E.R., Delong, D.M. & Clarke-Person, D.L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A Non-parametric approach. *Biometrics,* 44 , pp. 837-845.

Efron, B. (1975) The efficiency of Logistic Regression Compared to Normal Discriminant Analysis. *Journal of American Statistical Association*, 70, pp. 892-898.

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

Farewell, V. (1979) Some results on the estimation of logistic models based on retrospective data. *Biometrika*, 66, pp. 27-32.

Gibbons, R. D. & Hedeker, D. (1997) Random-effects probit and logistic regression models for three-level data. *Biometrics*, 53, pp. 1527-1537.

Gordis, L. (1996) *Epidemiology.* W.B. Saunders Company.

Greiner, M., Pfeiffer, D. & Smith, R.M. (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preview of Veterinary Medicine*, 45, pp. 23-41.

Hanley, J. A. & McNeil, B.J. (1982) The meaning and use of area under a receiver operating characteristic (ROC) curve. *Radiology,* 143, pp. 29-36.

Hanley, J.A. & McNeil, B.J. (1983) A method of comparing the area under two ROC curves derived from the same cases. *Radiology,* 148, pp. 839-843.

Heagerty, P.J., Lumley, T. & Pepe, M. S. (2000) Time-Dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56, pp. 337-344.

Hosmer, D. & Lemeshow, S. (1989) *Applied logistic regression.* John wiley & sons.

Lee, E.T. (1992) *Statistical methods for survival data Analysis* 2nd edn.: wiley.

Lemeshow, S. & Hosmer, D.W. (1982) A Review of Goodness-of-fit Statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, 115, pp. 92-106.

Linnet, K. (1987) Comparison of Quantitative Diagnostic Tests: Type 1 Error, Power, and Sample Size, *Statistics in Medicine*, 6, pp. 147-158.

McClish, D.K. (1987) Comparing the Areas under More than Two Independent ROC Curves. *Medical Decision Making*, 7, pp. 149-155.

Nakas, C., Yiannoutsos, C.T., Bosch, R.J. & Moyssiadis, C. (2003) Assessment of Diagnostic Markers by Goodness-of-fit Tests. *Statistics in Medicine,* 22, pp. 2503-2513.

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

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

Pepe, S.M. (2000) An interpretation for the ROC Curve and inference using GLM Procedures. *Biometrics,* 56, pp. 352-359.

Prentice, R.L. & Pyke, R. (1979) Logistic disease incidence models and case-control studies. *Biometrika,* 66, pp. 403-411.

Rao, C.R. (1973) *Linear Statistical Inference and Its Application*, 2nd ed. Wiley, New York.

Reiser, B. & Guttman, I. (1986) Statistical Inference for P(Y<X) :the Normal case. *Technometrics,* 28, pp. 253-257.

Renard, D., Greys, H., Molenberghs, G., Burzykowski, T., Buyse, M., Vangeneugden, T. & Bijnens, L. (2003) Validation of a longitudinally measured surrogate marker for a time-to-event endpoint. *Journal of Applied Statistics*, 30, pp. 235-247.

Ruiz-Velasco, S. (1991) Asymptotic Efficiency of Logistic Regression Relative to Linear Discriminant Analysis. *Biometrika*, 78, pp. 235-243.

Simonoff, J.S., Hochberg, Y. & Reiser, B. (1986). Alternative estimation procedures for Pr (X<Y) in categorised data. *Biometrics,* 42, pp. 895-907.

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

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

TenHave, T.R., Kunselman, A.R. & Tran, L. (1999) A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine*, 18, pp. 947-960.

Wieand, H.S., Gail, M.H., James, B.R. & James, K.L. (1989) A Family of Non-parametric Statistics for Comparing Diagnostic Markers with Paired or Unpaired Data. *Biometrika,* 76, pp. 585-592.

Youden, W.J. (1950) Index for rating diagnostic tests. *Cancer,* 3, pp. 32-35.

Zou, K.H. & Hall, W. J. (2000) Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*, 5, pp. 621-631.

Zou, K.H. & Hall, W. J. (2002) Semi-parametric and parametric transformation models for comparing diagnostic marker's with paired design. *Journal of Applied Statistics*, 29, pp. 803-816.