TG-ROC PLOTS WITH CONFIDENCE

Edson Zangiacomi MARTINEZ[1]
Francisco LOUZADA-NETO[2]

▪ *ABSTRACT: TG-ROC (two-graph receiver operating characteristic) plots are defined by a graph of the sensitivity and specificity against a range of cut-off points related to the possible outcomes of a diagnostic test. This approach provides a didactical picture for helping the selection of an optimal cut-off point, and it has several advantages compared with the usual ROC curve. The optimal cut-off point may be defined by clinical criteria related to the practical utility of the test and not only by mathematical rules, and the visualization of the performance measures provided by TG-ROC plots can be a useful tool to the medical researcher. We proposed the presentation of TG-ROC plots jointly with its respective confidence bounds, that allow a description of the sample variability of Se and Sp, and thus, to visualize the confidence of the diagnostic performance for each cut-off point candidate to be optimal. An approach for constructing confidence bounds is described, and a Monte Carlo study evidenced that these confidence bounds has coverage probability satisfactorily near to nominal value .95. This simulation used normal probability models for described the distribution of the continuous diagnostic test.*

▪ *KEYWORDS: TG-ROC plots; diagnostic tests; Monte Carlo method; interval estimation.*

1 Introduction

Receiver operating characteristic (ROC) analysis is widely recognized as the standard approach for describing the quality of diagnostic information and diagnostic decisions. ROC plots provide a global picture of the performance of a

[1]Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, USP, Av. Bandeirantes 3900, 14049-900, Ribeirão Preto, SP, Brazil. E-mail: *edson@fmrp.usp.br*
[2]Departamento de Estatística, Universidade Federal de São Carlos – UFSCar, caixa postal 676, 13565-905, Brasil. Email: *dfln@power.ufscar.br*

diagnostic procedure, according to pairs of sensitivity and specificity estimated over a set of established cut-offs points (Martinez *et al.*, 2003). A variant representation of ROC plot was proposed by Greiner *et al.* (1995), denominated TG-ROC (two-graph receiver operating characteristic) plots. This alternative is a graphical tool for demonstrating the performance of a continuous diagnostic test and to assist the choice of an optimal cut-off value. This approach is simply a plot of the sensitivity (Se) and specificity (Sp) against a range of different cut-offs points defined about the possible outcomes of the diagnostic test. Some examples of the use of the TG-ROC plots are given in the studies by Xu *et al.* (1997), Gubbels *et al.* (2000), Nogueira (2004) and Paweska *et al.* (2008).

TG-ROC plots are used as a complement or alternative to the ROC plot for helping in the selection of cut-off points in medical decision making problems. The establishment of the most optimal cut-off point is frequently based on values where Se and Sp are simultaneously large, but this is not always appropriate. In many situations, when a false negative result is more undesirable than a false positive result, it is more appropriate to focus on the Se. Thus, an optimal cut-off point may be defined by clinical criteria related to the practical utility of the test and not only by mathematical rules, and the visualization of the performance measures provided by TG-ROC plots can be a useful tool to the medical researcher.

When a researcher makes decisions about the optimal cut-off point, free of mathematical rules, it is desired to know the sample variability of Se and Sp at each value into a range of possible outcomes of the diagnostic test under investigation. This variability is represented by confidence bounds, with a given coverage probability. In this article, we discussed a free-distribution approach for providing pointwise confidence bounds for TG-ROC plots. Using a Monte Carlo approach, we studied the coverage probability of these confidence bounds.

2 Methods

2.1 Two-graph-ROC analysis (TG-ROC)

Firstly, we will describe the steps for the construction of the empirical TG-ROC curves and we will introduce a notation that will be used in this article. Let X and Y be two independent random variables, $X \in \Psi$ and $Y \in \Psi$, for continuous outcomes of the test under investigation, considering healthy and diseased individuals, respectively, and let $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_m)$ be realizations of X and Y, where n and m are the respective sample sizes. We denote Ψ as the set of all possible outcomes of the diagnostic test, considering $\psi = (x_1, x_2, ..., x_n, y_1, y_2, ..., y_m)$ a vector of $n+m$ observations, and $\psi_{(t)}$ is the t-th order statistics of ψ, $t = 1, ..., n+m$. For a possible value $\psi\prime \in \Psi$, not necessarily $\psi\prime \in \psi$, and for a fixed t, let the following decision rule: if $\psi\prime \geq \psi_{(t)}$, the individual is classified as positive and if $\psi\prime < \psi_{(t)}$, the individual is classified as negative.

Thus, we define

$$\widehat{S_e(t)} = \frac{\sum_{i=1}^{m} I\left(y_{(i)} \geq \psi_{(t)}\right)}{m} \quad \text{and} \quad \widehat{S_p(t)} = \frac{\sum_{i=1}^{n} I\left(x_{(i)} < \psi_{(t)}\right)}{n}$$

as the estimates for Se and Sp defined on a cut-off $\psi_{(t)}$. Consider $I(condition)$ as a function equal to 1 if the condition is true and equal to zero if the same condition is false.

The TG-ROC curves are thus defined as a plot of the both curves $\widehat{S_e(t)}$ and $\widehat{S_p(t)}$, for ψ, $t = 1, ..., n+m$, against the continuous outcomes of diagnostic test (in the horizontal axe). Greiner *et al.* (1995) denominated the curves constructed in this manner as non-parametric TG-ROC curves. They also considered an alternative of estimating TG-ROC curves based on parametric models, but this is out of the scope of the paper and will only be briefly commented in Section 4.

2.2 Free-distribution confidence interval for TG-ROC plots

Let $X = (X_1, X_2, ..., X_n)$ be a sample of independent random variables with continuous distribution $F_X(x)$, representing the outcomes of the test under investigation, for the healthy individuals. A confidence interval for $F(x)$ without assumptions about the distribution of X is provided by $X_{(1)}, X_{(2)}, ..., X_{(n)}$, the order statistics from $F(x)$. Let $W_j = F_X(X_{(j)})$, $j = 1, ..., n$, the probability that X is at most equal to j-th order statistics. Therefore, W_j is the S_p of the test, at a cut-off point given by $X_{(j)}$.

Independently of the distribution of the variable X, it is not difficult to show that W_j has uniform distribution over the interval $[0, 1]$, or say, $W_j \sim U(0, 1)$, for $j = 1, ..., n$ (see, for example, Casella and Berger, 2002, p.54). The probability that W_j is at most q is thus given by

$$P\left(W_j \leq q\right) = \frac{1}{B\left(j, n-j+1\right)} \int_0^q u^{j-1} \left(1-u\right)^{n-j} du,$$

the cumulative distribution function of a beta-distributed random variable, with parameters j and $n-j+1$, and where $B(\cdot, \cdot)$ denotes the beta function. A proof for this result is given by Kendall and Stuart (1967, p.518). A confidence interval for W_j (and consequently for $F_X(x)$) is constructed using the expression above, where the limits are calculated according to the inverse beta distribution.

The construction of a confidence interval associated with a curve for the values of Se (or S_p) is analogous. Since we aimed to construct simultaneous confidence intervals for both curves, and Se and Sp are treated as independent measures, we should determine that the confidence coefficients for the intervals are equal to square root of the chosen nominal coefficient.

The ideas described above provide the basis for the construction of the Hilger's two-stage pointwise confidence bounds (Hilgers, 1991), which search for simultaneous confidence limits for Se and Sp at each cut-off point established by ROC analysis.

3 Results

3.1 A Monte Carlo study

We performed a Monte Carlo study aiming to estimate the coverage probabilities of free-distribution confidence bounds for TG-ROC curves described in Section 2. We assumed that the random variables X and Y are normally distributed with prefixed parameters. For each different sample size, 20,000 samples were generated. We then construct the TG-ROC curves with their respective free-distribution simultaneous confidence bounds, with nominal coverage probability fixed at $100(1-\alpha)\% = 95\%$. Thus, for the Se and Sp curves, the estimated Monte Carlo coverage probabilities are given by the proportion of generated confidence intervals which contain the true Se (or Sp) curve at different cut-off values. This intensively computational estimation method was implemented in IML (Interactive Matrix Language) procedure in the commercial package SAS ® version 9.1.

The results are summarized in Table 1 for $X \sim N(0,1)$ and $Y \sim N(1,1)$, and in Table 2 for $X \sim N(0,1)$ and $Y \sim N(2,1)$. The estimated coverage probabilities for Se and Sp are satisfactorily closed to 0.975 (the square root of 0.95), even when the sample sizes are small and the cut-off values are extreme values. For Se and Sp curves considered simultaneously, the coverage probabilities are almost always slightly greater than 0.95. However, the largest relative difference between the Monte Carlo coverage probability and the nominal probability of 0.95 is equal to 0.024, that is, it seems that free-distribution confidence bounds for TG-ROC curve are satisfactorily accurate.

3.2 The serum levels of CA125 data (Torres *et al.*, 2002)

The CA 125 antigen (Verheijen *et al.*, 1999) is a marker widely used for the diagnosis and follow-up of ovarian cancer. In women presenting pelvic tumours, elevations in serum levels of antigen reactive with CA 125 (measured in units per milliliter, or U/ml) are indicative of malignancy. A study (Torres *et al.*, 2002) about the malignancy of pelvic tumours was performed at the Pelvic Oncology Ambulatory of Women's Health Care Center, State University of Campinas (UNICAMP), Brazil, considering a sample of $m = 91$ women with benign pelvic masses and $n = 67$ women with malignant tumours. The gold standard was obtained by histopathological examination.

Figure 1 shows the TG-ROC curves for the serum levels of CA 125 (in the logarithmic scale), that is, the curves for $\widehat{S_e(t)}$ and $1 - \widehat{S_p(t)}$, $t = 1, ..., 158$, for each observed serum level of CA 125, with the simultaneous 95% confidence bounds generated by the free-distribution method. Table 3 shows the 95% free-distribution confidence bounds for several observable cut-off points for CA 125 serum levels (U/ml). The cut-off point of 30.4 U/ml for the CA 125 levels corresponds to a relatively high combined value of S_e and S_p in predicting the malignancy of pelvic tumours, where the 95% confidence bound for S_e is (.658,.879) and the 95% confidence bound for S_p is (.619,.824).

Tabela 1 - Monte Carlo coverage probabilities of simultaneous free-distribution confidence bounds based in beta distribution for TG-ROC curves, under the model $X \sim N(0,1)$ and $Y \sim N(1,1)$. The nominal coverage probability for Se and Sp curves was simultaneously .95.

sample		cut-off value								
size	measure	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
$n = 10$	S_e	.987	.980	.971	.971	.966	.969	.971	.979	.980
$m = 10$	S_p	.974	.984	.986	.988	.988	.990	.987	.988	.988
	S_e and S_p	.962	.965	.957	.960	.954	.959	.958	.968	.969
$n = 20$	S_e	.980	.974	.974	.971	.967	.961	.964	.972	.978
$m = 10$	S_p	.969	.986	.987	.992	.992	.994	.987	.990	.990
	S_e and S_p	.950	.960	.961	.964	.960	.955	.951	.962	.968
$n = 20$	S_e	.979	.975	.974	.969	.971	.965	.966	.975	.980
$m = 20$	S_p	.974	.968	.989	.990	.990	.991	.990	.990	.990
	S_e and S_p	.954	.944	.963	.960	.961	.956	.956	.964	.970
$n = 50$	S_e	.975	.972	.973	.975	.973	.968	.962	.967	.975
$m = 20$	S_p	.973	.965	.991	.991	.989	.993	.991	.990	.990
	S_e and S_p	.949	.938	.964	.966	.962	.961	.953	.956	.966
$n = 50$	S_e	.976	.974	.975	.974	.972	.970	.964	.967	.977
$m = 50$	S_p	.976	.976	.971	.990	.993	.993	.993	.988	.990
	S_e and S_p	.952	.951	.946	.965	.965	.963	.957	.955	.966
$n = 100$	S_e	.974	.975	.974	.972	.975	.973	.960	.961	.968
$m = 20$	S_p	.973	.965	.992	.992	.990	.993	.995	.991	.992
	S_e and S_p	.948	.941	.967	.964	.965	.966	.955	.952	.961
$n = 100$	S_e	.977	.976	.974	.972	.975	.974	.962	.959	.966
$m = 10$	S_p	.958	.988	.988	.997	.998	.998	.989	.996	.996
	S_e and S_p	.935	.964	.962	.968	.972	.972	.950	.955	.962

Tabela 2 - Monte Carlo coverage probabilities of simultaneous free-distribution confidence bounds based in beta distribution for TG-ROC curves, under the model $X \sim N(0,1)$ and $Y \sim N(2,1)$. The nominal coverage probability for Se and Sp curves was simultaneously .95.

sample		cut-off value								
size	measure	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
$n = 10$	S_e	.982	.982	.986	.975	.970	.962	.958	.962	.966
$m = 10$	S_p	.974	.980	.985	.990	.992	.996	.987	.994	.995
	S_e and S_p	.956	.963	.971	.966	.962	.959	.946	.957	.961
$n = 20$	S_e	.981	.982	.977	.973	.974	.972	.962	.956	.960
$m = 10$	S_p	.971	.983	.987	.993	.994	.998	.988	.997	.998
	S_e and S_p	.952	.966	.964	.966	.969	.970	.950	.953	.958
$n = 20$	S_e	.972	.982	.978	.973	.973	.970	.963	.959	.961
$m = 20$	S_p	.974	.970	.987	.990	.991	.994	.995	.995	.997
	S_e and S_p	.947	.953	.965	.963	.964	.964	.959	.954	.958
$n = 50$	S_e	.982	.982	.974	.971	.973	.975	.974	.965	.959
$m = 20$	S_p	.975	.969	.990	.991	.989	.994	.998	.995	.999
	S_e and S_p	.957	.951	.964	.962	.963	.969	.972	.960	.958
$n = 50$	S_e	.977	.984	.975	.974	.974	.974	.973	.963	.960
$m = 50$	S_p	.975	.975	.972	.990	.994	.997	.998	.989	.998
	S_e and S_p	.952	.959	.948	.965	.969	.971	.971	.952	.957
$n = 100$	S_e	.984	.976	.974	.974	.974	.972	.975	.973	.959
$m = 20$	S_p	.975	.968	.992	.992	.990	.993	.999	.995	.999
	S_e and S_p	.959	.945	.966	.966	.964	.966	.974	.968	.959
$n = 100$	S_e	.982	.974	.976	.975	.974	.972	.975	.973	.961
$m = 10$	S_p	.971	.987	.988	.997	.998	.999	.992	.998	1
	S_e and S_p	.954	.962	.963	.972	.972	.970	.967	.972	.961

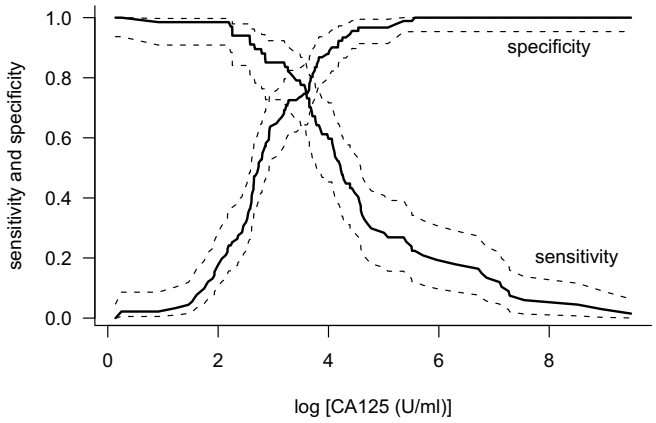

Figura 1 - TG-ROC curves for the serum levels of CA 125 (in a logarithmic scale), with simultaneous 95% confidence bounds generated by the free-distribution method.

4 Discussion

ROC curves are a well-known tool for visually describing the performance of a test. Confidence bounds for ROC curves was proposed by Hilgers (1991), Ma and Hall (1993), Jensen *et al.* (2000) and Schafer (2006) based on different perspectives. On the other hand, in the medical literature many authors has been presented the results of the performance assessment of diagnostic tests by TG-ROC curves, but the use of confidence bounds for this approach is not common in these aplications. The present article presents a simple procedure for constructing confidence bounds for TG-ROC curves which can be easily implemented using existing software, such as SAS or R.

Greiner *et al.* (1995) pointed that an advantage of TG-ROC approach over the ROC curve is to allow a more direct reading of the value of a cut-off point associated to a specific combination of Se and Sp. However, ROC curves may be preferable where it is desired to compare the global performance of two or more diagnostic tests, especially when measured on different scales. In this situation, the individual accuracy of the selected cut-off points is not important, but the potentiality of each test in discriminating between diseased and healthy individuals. A usual global performance measure of a test is the area under the ROC curve, interpreted by Begg (1991) as the probability, given one randomly selected diseased individual and one randomly selected healthy individual, that the test results are correctly ranked. This index is useful for comparisons among tests whose ROC curves are not crossed, and it is not visible in TG-ROC curves.

Greiner *et al.* (1995) cited that a parametric version of the TG-ROC curves provides a statistical power superior to that of the non-parametric method described

Tabela 3 - 95% free-distribution confidence bounds for several observable cut-off points for serum levels of CA 125 (U/ml).

CA 125 cut-off value (U/ml)	$\widehat{S_e}$	95% confidence bounds	$\widehat{S_p}$	95% confidence bounds
5.0	.985	(.909 .997)	.077	(.034 .163)
6.9	.985	(.909 .997)	.131	(.080 .245)
10.5	.940	(.841 .979)	.264	(.176 .381)
15.0	.896	(.782 .953)	.483	(.374 .605)
18.9	.851	(.727 .923)	.637	(.527 .748)
30.4	.791	(.658 .879)	.725	(.619 .824)
36.1	.761	(.625 .856)	.747	(.643 .843)
39.6	.686	(.560 .807)	.802	(.704 .887)
45.5	.642	(.498 .756)	.857	(.768 .929)
70.5	.478	(.338 .605)	.934	(.866 .979)
100.3	.343	(.218 .471)	.967	(.914 .995)
301.6	.209	(.109 .325)	1	(.954 1)

above. A parametric approach takes into account the estimation of parameters for the curves relative to Se and Sp, considering a known probability distribution. Thus, monotone data transformation may lead to a satisfactory approximation to a normal distribution. However, in many practical situations the variables X and Y have not the same distribution, consequently, an unique data transformation is not sufficiently for producing outcomes with normal distribution for both healthy and diseased individuals. This may be clinically explicable, because we can consider that the test outcomes has in many times the same distribution for healthy and diseased individuals, except for a "perturbation" in this distribution for the diseased individuals due to the proper disease. In fact, the normal probability plots in Figure 2 show that the serum levels of CA 125 (transformed in a logarithmic scale) have a distribution approximately closed to a normal for healthy women (Figure 2(a)), but this is not true for diseased individuals (Figure 2(b)). In conclusion, although the parametric method is more advantageous, its application is not always straightforward.

5 Concluding remarks

TG-ROC curves are proposed as a didactical tool for demonstrating the accuracy of a diagnostic test, and this graphical approach is useful for aiding the decisions about the optimal cut-off value, as pointed by Greiner *et al.* (1995). The estimation of the sample variability of Se and Sp for each possible outcome is an important step in the determination of the optimal cut-off, and this variability may be expressed by confidence bounds for TG-ROC curves. In this article, we demonstrated that an exact confidence interval for TG-ROC curves can be easily

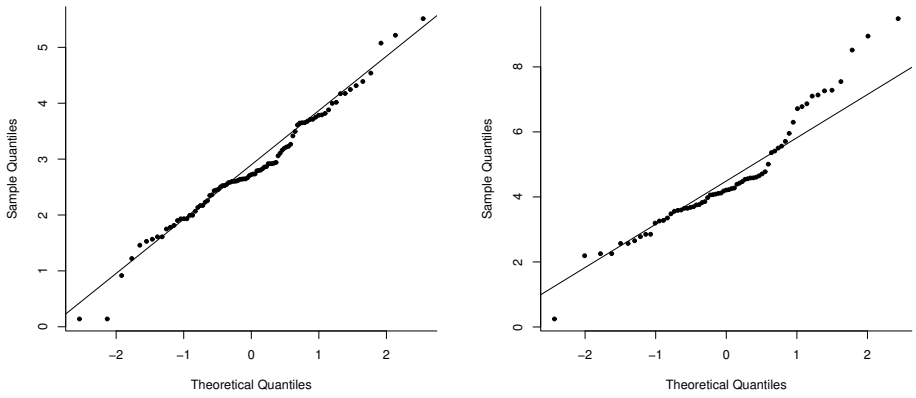

Figura 2 - Normal probability plots for serum levels of CA 125 (transformed in a logarithmic scale) for $m = 91$ women with benign tumours (a) and $n = 67$ women with malignant tumours (b)

constructed, based in the inverse of beta distribution, independently the original distribution of data.

MARTINEZ, E. Z.; LOUZADA-NETO, F. Curvas TG-ROC com confiança. *Rev. Mat. Estat.,* São Paulo, v.26, n.1, p.87-97, 2008.

- RESUMO: A curva TG-ROC (two-graph receiver operating characteristic) é um gráfico da sensibilidade (S) e da especificidade (E) de um teste diagnóstico em relação a um intervalo de possíveis pontos de corte. Este gráfico é uma didática ferramenta para o auxílio da determinação de um ponto de corte ótimo para um teste diagnóstico cujos resultados são expressos por uma variável contínua, e traz algumas vantagens em relação à tradicional curva ROC. A determinação do ponto de corte ótimo deve ser feita de acordo com critérios clínicos, sendo a visualização da curva TG-ROC um apoio para esta tomada de decisão. Nós propomos a apresentação da curva TG-ROC em conjunto com suas respectivas regiões de confiança, que permitem demonstrar a variabilidade amostral de S e E, e conseqüentemente, uma visualização da confiança de cada valor candidato a um ponto de corte ótimo. É descrito um método para a construção das regiões de confiança, e um estudo Monte Carlo evidencia que estas regiões de confiança possuem probabilidades de cobertura próximas à nominal (0,95). Esta simulação usa modelos de probabilidade normal para descrever a distribuição dos valores contínuos de um teste diagnóstico.

- PALAVRAS-CHAVE: Curvas TG-ROC; testes diagnósticos; método Monte Carlo; estimação por intervalo.

Referências

BEGG, C. B. Advances in statistical methodology for diagnostic medicine in the 1980's. *Stat. Med.*, Hoboken, v.10, n.12, p.1887-1895, 1991.

CASELLA, G.; BERGER, R. L. *Statistical inference*. 2nd. ed. Duxbury: Pacific Grove, 2002. 660p.

GREINER, M.; SOHR, D.; GÖBEL, P. A modified ROC analysis for the selection of cut-off vales and the definition of intermediate results of serodiagnostic tests. *J. Immunol. Methods*, Amsterdam, v.185, n.1, p.123-132, 1995.

GUBBELS, M.J.; D'OLIVEIRA, C.; JONGEJAN, F. Development of an indirect Tams1 enzyme-linked immunosorbent assay for diagnosis of Theileria annulata infection in cattle. *Clin. Diag. Lab. Immunol.*, Washington, v.7, n.3, p.404-411, 2000.

HILGERS, R.A. Distribution-free confidence bounds for ROC curves. *Methods Inf. Med.*, Stuttgart, v.30, n.2, p.96-101, 1991.

JENSEN, K.; HANS-HELGE, M.; SCHÄFER, H. Regional confidence bands for ROC curves. *Stat. Med.*, Hoboken, v.19, n.4, p.493-509, 2000.

KENDALL, M.G.; STUART, A. *The advanced theory of statistics*. 2nd. ed. London: Griffin, 1967. v.2. 723p.

MA, G.; HALL, W. J. Confidence bands for receiver operating characteristic curves. *Med. Decis. Mak.*, v.13, n.3, p.191-197, 1993.

MACSKASSY, S. A.; PROVOST, F.; ROSSET, S. ROC confidence bands: an empirical evaluation. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML), 22., 2005, Bonn, Germany. *Preceedings...* Bonn, Germany. 2005.

MARTINEZ, E. Z.; LOUZADA NETO, F.; PEREIRA, B. B. A curva ROC para testes diagnósticos. *Cad. Saúde Col.*, Rio de Janeiro, v.11, p.7-31, 2003.

NOGUEIRA, Y.L. Estimate of the validity of a new method for the isolation of rabies virus. *Rev. Saúde Pública*, Rio de Janeiros, v.38, n.2, p.315-322, 2004.

PAWESKA, J.T.; VAN VUREN, P.J.; KEMP, A.; BUSS, P.; BENGIS, R.G.; GAKUYA, F.; BREIMAN, R.F.; NJENGA, M.K.; SWANEPOEL, R. Recombinant nucleocapsid-based ELISA for detection of IgG antibody to Rift Valley fever virus in African buffalo. *Vet. Microbiol.*, Amsterdam, v.127, n.1-2, p.21-8, 2008.

SCHÄFER, H. Efficient confidence bounds for ROC curves. *Stat. Med.*, Hoboken, v.13, n.15, p.1551-1561, 2006.

TORRES, J.C.C.; DERCHAIN, S.F.M.; FAÚNDES, A.; GONTIJO, R.C.; MARTINEZ, E.Z.; ANDRADE, L.A.L.A. Risk-of-malignancy index in preoperative evaluation of clinically restricted ovarian cancer. *São Paulo Med. J.*, São Paulo, v.120, n.3, p.72-76, 2002.

VERHEIJEN, R.H.M.; VON MENSDORFF-POUILLY, S.; VAN KAMP, G.J.; KENEMANS, P. CA 125: fundamental and clinical aspects. *Semin. Cancer Biol.*, v.9, p.117-124, 1999.

XU, H.; LOHRB, J.; GREINER, M. The selection of ELISA cut-off points for testing antibody to Newcastle disease by two-graph receiver operating characteristic (TG-ROC) analysis. *J. Immunol. Methods*, Amsterdam, v.208, n.1, p.61-64, 1997.

Received in 28.01.2008.

Approved after revised in 09.04.2008.