J Vet Clin 19(3) : 312-315(2002)

Application of Receiver Operating Characteristic (ROC) Curves for Clinical Diagnostic Tests

Son-Il Pak[1], Hee-Seung Koo*, Cheol-Yong Hwang and Hwa-Young Youn****

Department of Veterinary Medicine, Kangwon National University, Chuncheon 200-701, Korea
**College of Natural Resources, Colorado State University, Fort Collins, CO 80523, USA*
***College of Veterinary Medicine, Seoul National University, Korea*

Abstract : Diagnostic tests often require the determination of cut-off values that discriminate uninfected from infected individuals. The receiver operating characteristic (ROC) curve has been frequently used to attain this purpose and gives a representation of diagnostic accuracy (sensitivity and specificity) of a prediction model when varying the cut-point of a decision rule on a whole spectrum. We have written and tested a visual basic application program in EXCEL for maximum likelihood estimation of a binormal ROC curve, which also computes univariate statistics of a diagnostic test employed. Examples applying for computed tomographic images in radiology and methicillin-resistant *Staphylococcus aureus* research are given to illustrate this approach. This stand-alone module is available from the first author on request.

Key words : Receiver operating characteristic (ROC) curve, Clinical Diagnosis, Application

Introduction

In medicine screening and diagnostic tests are commonly used to detect underlying disorders and separate a certain disease group from non-diseased group. Ideally, perfect diagnostic test would always be positive in diseased group and negative in non-diseased group. For many diagnostic tests, however, some overlap of test results is seen among patients with and without the disease. Therefore researchers would select optimum cut-off values to differentiate these two groups most accurately. When other cut-off values are used, the degree of overlap between the test results from the two patient groups also changes and thus, sensitivity and specificity change.

The changes in sensitivity and specificity introduced by the use of other cut-off values can be graphically presented - the receiver operating characteristic (ROC) curve. Several authors described the basic principles of ROC analysis in biomedical contexts.[9,10,13] Greiner *et al.*[6] reviewed relevant features constructing ROC curve and related approaches with emphasis on cut-off selection in a variety clinical settings and test comparison. Briefly, ROC curve is a graph of the sensitivity (true-positive, TP) on the vertical axis against the 1-specificity (false-positive, FP) on the horizontal axis. Each point represents the true-positive and false-positive rate for a given study.

A number of indices have been used to summarize the information contained in the ROC curve,[14] and one of which is the area under the ROC curve (AUC), a quantitative measure of the tests capacity to classify two populations into separate group, where a value of 0.5 is obtained if the test does no better than chance because the true-positive rate equaled the false-positive rate. A value of 1 is obtained if the test is perfect as the curve moves towards the left and top boundaries of the plot.[8] A straight line arising from the origin at a 45-degree angle has an area under the curve of 0.5 and represents accuracy no better than flipping a coin. A perfect predictive model has an area under the curve of 1.0. As accuracy and discrimination improve, the ROC curve moves upward and to the left. ROC curves allow one to compare different predictive models used in the same population of patients.

To authors knowledge, ROC curves have been increasingly used in many fields, but still limited in veterinary medicine.[2,5] Two different types results-continuous scale and dichotomous-can be obtained from the screening tests. These differing types of outcome variables naturally lead to different performance measures. In this study we present the application of ROC curve for the ordinal scale outcomes, using the datasets from the studies of radiology and microbiology.

Materials and Methods

Analytical Module Employed

We developed a stand-alone module (ROCFITVBA), which was coded by using the visual basic application (VBA) program in EXCEL for maximum likelihood estimation of a binormal ROC curve and its associated parameters from a set of categorical rating-scale data. This program is a slightly modified version of the program ROCFIT by Metz *et al.*[11] and transformed to VBA. The mathematical basis of the ROCFIT is described well in elsewhere[3,4].

In this module the accuracy of rating data graphical repre-

[1]Corresponding author.
E-mail : paksi@kangwon.ac.kr

sentation between FP and TP was constructed by summing the conditional probability from top to bottom of data table. This cumulative normal probability plot shows the TP that would be obtained to yield a particular ROC curve in every value of the FP from 0 to 1. During the running program, each cell frequency is divided by the total number of frequency and by each disease status (eg, normal vs. abnormal) for estimating a conditional probability. It also calculates cumulative probabilities by summing the conditional probability top to bottom. Thus, the program requires only the absolute frequencies of ratings in each category.

ROCFITVBA used Newton-Raphson method to solve the log of likelihood equations with initial value from the parameter estimates of the least-squares solution because of nonlinearity of equations. Final estimates of the expected operating points on the fitted ROC curve, with lower and upper bounds along the fitted curve of the asymmetric 95% confidence intervals for those operating point estimates, calculated assuming that errors in z(k) estimates, the normal deviate value of P(FP), are normally distributed. Thus, estimates of the expect operating points on the fitted curve, on normal-deviate axe have abscissa values = -z(k) and ordinate values = $a - b \times z(k)$, resulting in the ROC curve shows straight lines on z-coordinates. Chi-square was reported as an index of goodness-of-fit for examining ROC curve. However, the chi-square goodness-of-fit measure does not calculate if expected cell frequency is less than 5, or if only 3 rating categories are employed.

Calculation of AUC

To derive the AUC we used two approaches described by Hanley and McNeil[7] based on Mann-Whitney test with no assumptions about the distribution of MSSA and MRSA patients and maximum likelihood. For the latter method we developed a spreadsheet program using the ROCFITVBA.

Example data

We used two complete datasets. First was from radiological study of computed tomographic images for the detection of neurological problems[7]. Second dataset was from the study of microbiology aimed at timely and accurate diagnosis of the methicillin-resistant *Staphylococcus aureus* (MRSA) infection[12].

Results

Table 1 presents illustrative data showing how a single reader rated the CT images obtained in a sample of 109 patients with neurological problems[7]. They classified each image into one of five categories: definitely normal, probably normal, questionable, probably abnormal, and definitely abnormal.

Antibiotic resistant pathogens are becoming increasingly prevalent in the hospitals. One of these is MRSA which can easily spread among hospitalized patients and colonized for a very long time, acting as a source of infection. It is reported that treating patients with MRSA is expensive and difficult: the average cost of handling such patient is about 7 times greater than treating methicillin-susceptible *S. aureus* (MSSA) cases[1]. Therefore, to help clinicians effectively treat patients it is essential to detect and distinguish those infected with MRSA from those infected with MSSA. Table 2 is summarized result from Shang *et al.*[12]. who investigated MRSA risk contributors in different geographic locations to

Table 1. Results from a single rater reading computed tomographic images (Hanley and McNeil, 1982) (n=109)

True Disease status	Definitely normal	Probably normal	Questionable	Probably abnormal	Definitely abnormal
Normal	33	6	6	11	2
Abnormal	3	2	2	11	33
Totals	36	8	8	22	35

Area under the curve (standard error)
Mann-Whitney method: 89.3% (3.2%)
Maximum likelihood method: 91.1% (3.0%)

Table 2. Classification results for MRSA and MSSA using neural network model (Shang et al., 2000) (n=472)

True Disease status	Definitely MSSA	Likely MSSA	Questionable	Likely MRSA	Definitely MRSA
MSSA	261	64	6	9	9
MRSA	8	4	26	23	62
Totals	269	68	32	32	71

Area under the curve (standard error)
Mann-Whitney method: 92.8% (1.6%)
Maximum likelihood method: 95.1% (1.0%)

provide physicians with guidelines for treatments. The authors used 5 rating categories of 0.1, 0.3, 0.5, 0.7, and 0.9 as cut-off points, which correspond to definitely MSSA, likely MSSA, questionable, likely MRSA, and definitely MRSA.

Discussion and Conclusion

Several authors described the basic principles of ROC analysis in biomedical contexts[9,10,13,15]. The index of accuracy, the proportion of patients for which the diagnosis is correct, has two components: the true positive or sensitivity (the probability that an indicator is positive when a particular disease is present) and the true negative (TN) or specificity (the probability that an indicator is negative when a particular disease is not present). Their complements, the false positive (type II error) fraction = 1-TP and false positive (type I error) fraction = 1-TN, are also used. The plot of TP fraction as a function of FP fraction is called a ROC curve, showing the trade-off between TP and FP errors. A number of indices have been used to summarize the information contained in the ROC curve[14], and of which the area under the ROC curve is a quantitative measure of the tests capacity to classify two populations into separate group. This area measures the probability that the inhibition percentage of the two samples (for example, randomly paired samples of diseased and non-diseased individuals) will allow samples to be correctly identified[8]. A value of 0.7-0.8 in the fitted smooth ROC curve can be considered as reasonable discrimination, and over 0.8 as good[16,17]. The Mann-Whitney statistic often is computed to test whether the levels of some quantitative variable in one population tend to be greater than in a second population, without assuming how the values are distributed in the two populations.

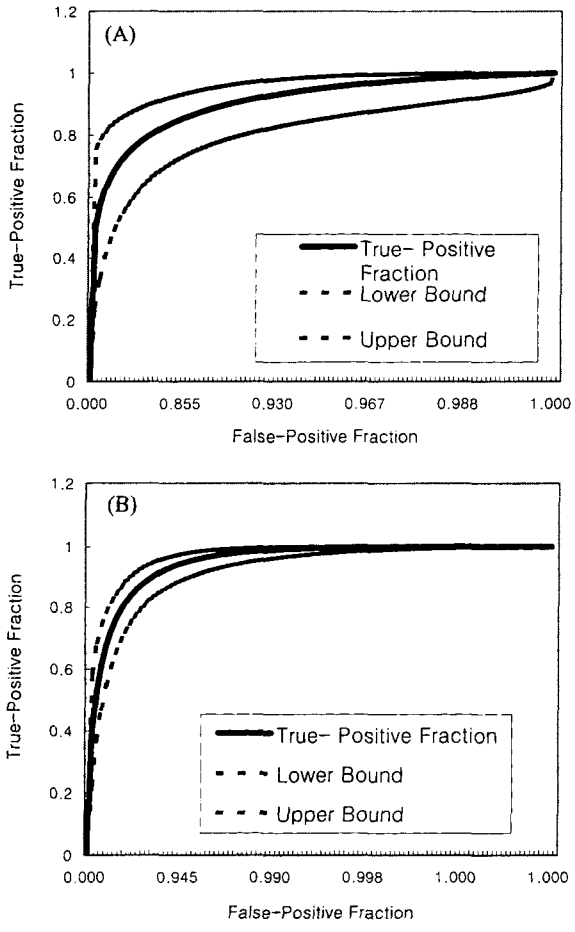

Fig 1. ROC curves for Table 1 (A) and Table 2 (B).

The ROC curve is a popular tool used to evaluate the accuracy of medical diagnosis and prognosis and provides a graphic representation of the tradeoff between true positive rates and false positive rates for predicting disease. We presented and compared two methods to estimate the AUC. The commercial software is not cheap enough to use for researchers calculating only AUC. Our spreadsheet program was inspired by this thought. The parameter estimates obtained from Mann-Whitney statistic, maximum-likelihood method and commercial software are quite similar with no great difference, resulting the spreadsheet program we developed can be considered as useful tool. If program is stopped due to singularity of inverse matrix (ie, determinant is zero), total frequency of response categories is too small for maximum likelihood theory. In this case, one could re-group infrequently used categories with adjacent response categories until the problem is solved. If this does not solve the problem, use initial estimates. Although initial values are not optimal, they are acceptable as a first approximation[14,15]. The program can be obtained from the first author of this paper on request.

References

1. Boyce JM, Potter-Bynoe G, Chenevert C, King T. Environmental contamination due to methicillin-resistant *Staphylococcus aureus*. Infect Control Hosp Epidemiol 1997; 18: 622-627.
2. Detilleux J, Arendt J, Lomba F, Leory P. Methods for estimating area under receiver-operating characteristic curves: illustration with somatic-cell scores in subclinical intramammary infections. Prev Vet Med 1999; 41: 75-88.
3. Dorfman, DD, Alf E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals-rating-method data. J Math Psychol 1969; 6: 487-496.
4. Dorfman, DD. Maximum-likelihood estimation of parameters of signal detection theory- a direct solution. Psychometrika 1968; 33: 117-124.
5. Greiner M, Sohr D, Göbel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. J Immunol Methods 1995; 185: 123-132.

6. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 2000; 45: 23-41.
7. Hanley JA, McNeil BJ. The meaning and use of the area under an ROC curve. Radiology 1982; 143: 29-36.
8. Kraemer HC. Evaluating medical tests, objective and quantitative guidelines. London, Sage Publications. 1992; 63-95.
9. McNeil BJ, Keeler E, Adelstein JS. Primer on certain elements of medical decision making. N Engl J Med 1975; 293: 211-215.
10. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283-298.
11. Metz CE, Shen J-H, Wang P-L, Kronman HB. ROCFIT. Department of radiology and the Franklin Mclean memorial research institute, The University of Chicago, Chicago, 1993.
12. Shang JS, Lin YE, Goetz AM. Diagnosis of MRSA with neural networks and logistic regression approach. Health Care Man Sci 2000; 3: 287-297.
13. Swets JA, Pickett RM, Whitehead SF, Getty DJ, Schnur JA, Swets JB, Freeman BA. Assessment of diagnostic technologies. Science 1979; 205: 753-759.
14. Swets JA. Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. Psychol Bullet 1986; 99: 181-198.
15. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285-1293.
16. Ward CD. The differential positive rate, a derivative of receiver operating characteristic curves useful in comparing tests and determining decision levels. Clin Chem 1986; 32: 1428-1429.
17. Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia: WB Saunders, 1980.

임상진단 검사에서 ROC 곡선의 응용

박선일 · 구희승* · 황철용 · 윤화영****

강원대학교 수의학과

*콜로라도 주립대학교 자연과학대학

**서울대학교 수의과대학

초 록 : 질병에 이환된 개체로부터 이환되지 않은 개체를 구분하기 위해 사용되는 대부분의 진단검사는 판별의 기준점 (cut-off value)을 필요로 한다. ROC (receiver operating characteristic) 곡선은 이러한 목적으로 흔히 사용되고 있으며 진단의 기준점을 다양하게 변화시킬 때 진단검사의 정확도 (민감도와 특이도)를 제시해주는 지표로 활용되고 있다. 저자들은 수의학관련 연구자들이 이 방법을 효과적으로 사용할 수 있도록 EXCEL에 내장된 비쥬얼 베이직으로 binormal ROC 곡선의 최대우도비를 계산해주는 프로그램을 작성하였다. 방사선 분야의 자료와 미생물학 자료를 예제로 들어 이 프로그램의 활용성을 높이고자 하였고 이 분야에 관심이 있는 연구자는 저자에게 연락하여 이 프로그램을 얻을 수 있다.

주요어 : ROC, 임상진단검사, 응용