pISSN 1598-298X / eISSN 2384-0749
*J Vet Clin* 33(2) : 97-101 (2016)
http://dx.doi.org/10.17555/jvc.2016.04.33.2.97

Journal of Veterinary Clinics
www.ksvc.or.kr

# 진단검사의 특성 평가를 위한 Receiver Operating Characteristic (ROC) 곡선의 활용

박선일[1] · 오태호*
강원대학교 수의과대학 및 동물의학종합연구소, *경북대학교 수의과대학

# Application of Receiver Operating Characteristic (ROC) Curve for Evaluation of Diagnostic Test Performance

**Son-Il Pak[1] and Tae-Ho Oh***

*College of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chuncheon 200-701, Korea*
**College of Veterinary Medicine, Kyungpook National University, Daegu 702-701, Korea*

(Received: December 09, 2015 / Accepted: April 14, 2016)

**Abstract :** In the field of clinical medicine, diagnostic accuracy studies refer to the degree of agreement between the index test and the reference standard for the discriminatory ability to identify a target disorder of interest in a patient. The receiver operating characteristic (ROC) curve offers a graphical display the trade-off between sensitivity and specificity at each cutoff for a diagnostic test and is useful in assigning the best cutoff for clinical use. In this end, the ROC curve analysis is a useful tool for estimating and comparing the accuracy of competing diagnostic tests. This paper reviews briefly the measures of diagnostic accuracy such as sensitivity, specificity, and area under the ROC curve (AUC) that is a summary measure for diagnostic accuracy across the spectrum of test results. In addition, the methods of creating an ROC curve in single diagnostic test with five-category discrete scale for disease classification from healthy individuals, meaningful interpretation of the AUC, and the applications of ROC methodology in clinical medicine to determine the optimal cutoff values have been discussed using a hypothetical example as an illustration.

***Key words*** **:** receiver operating characteristic curve (ROC), diagnostic test performance, accuracy.

## 서 론

ROC (Receiver Operating Characteristic) 곡선은 2차 세계대전 당시 신호검출이론(signal detection theory)에 근거하여 레이더 영상에서 적군을 감별하기 위한 수단으로 공학에서 처음으로 개발된 이후(4,20) 심리학, 생명과학, 경제학, 사회학 등 다양한 분야에 응용되어 왔다(2,9,14). 의학 분야에서는 1960년대에 영상진단 분야에 처음으로 도입된 이래(11) 임상의학에서는 질병 진단과 예측, 환자 분류를 위한 연구, 생물정보학, 실험실 검사, 영상진단, 역학 등에 광범위하게 사용되고 있다(1,3,5,10,16,17,19,20).

ROC 곡선은 진단검사 결과에 근거하여 다수의 의사결정 기준점(operating condition, cutoff, decision threshold, 양성 판정기준)에 대하여 가양성률(false positive rate, FPR, 1-specificity)과 진양성률(true positive rate, TPR, sensitivity)을 나타낸 그림이다(13,15). 가양성률과 진양성률 간의 관계에 대한 ROC 곡선을 통하여 첫째, 진단검사의 정확도 즉 질병에 감염된 집단(환자군)과 정상 집단(대조군)을 올바르게 분류하는 능력을 평가하며 둘째, 민감도(sensitivity, Se)와 특이도(specificity, Sp)를 동시에 고려한 상태에서 집단 분류를 위한 최적의 의사결정 기준점을 찾고 셋째, 경쟁관계에 있는 진단검사의 판별 능력이나 집단 분류를 위한 통계적 모형의 정확도를 상호 비교하며 넷째, 민감도(특이도)가 특이도(민감도) 보다 더 중요한 상황에서 특이도(민감도)를 고정할 때 목표로 하는 민감도(특이도)를 달성할 수 있는 기준점을 선택하는 등 다양한 목적으로 활용할 수 있다. 특히 검사결과가 연속형 척도인 경우 ROC 곡선에서 양성 판정을 위한 기준점(cutoff)을 $-\infty$에서 $\infty$로 변화시킬 때 민감도와 가양성률의 변화를 그래프로 파악함으로써 진단검사의 활동 특성(operating characteristic)을 정량적으로 평가할 수 있다. 본 연구에서는 가상의 예시 자료를 사용하여 ROC 곡선을 작성하는 방법과 요약통계량을 추정하는 방법론을 소개한다.

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

## 결 론

### 진단검사 결과 요약 및 ROC 곡선

ROC 곡선을 사용하여 진단검사의 정확도를 추정하기 위해서는 표준검사(gold standard test)를 이용하여 각 검체의 질병 상태에 대한 정보를 정확히 알고 있어야 한다. 예를 들

**Table 1.** The decision matrix showing responses of a binary diagnostic test, according to the status of each individual samples

| Category of test result | Disease status (gold standard) | | |
|---|---|---|---|
| | Positive (D+) | Negative (D–) | Total |
| Positive (T+) | a: true positives | b: false positives | a + b: test positives |
| Negative (T–) | c: false negatives | d: true negatives | c + d: test negatives |
| Total | a + c: diseased | b + d: non-diseased | a + b + c + d: sample size |

Sensitivity = true positive fraction = a/(a + c). Specificity = true negative fraction = d/(b + d).
Prevalence = (a + c)/(a + b + c + d)

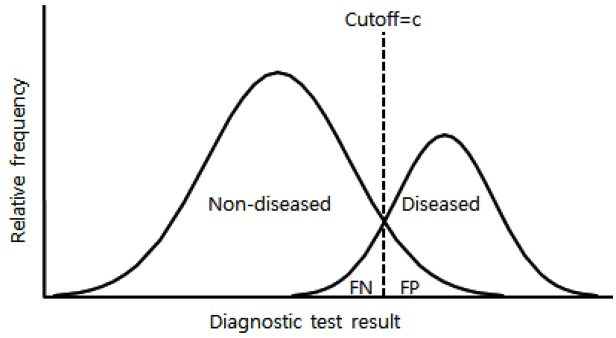


**Fig 1.** Probability density function of a hypothetical diagnostic test results for two populations (non-diseased and diseased).

어 질병상태의 이분형(binary) 참값을 알고 있는 검체에 대하여 환자 분류를 위해 사용한 진단검사(index test) 결과가 이분형일 때 Table 1과 같이 요약할 수 있다. 마찬가지로 진단검사 결과가 연속형인 경우 환자 분류를 위하여 기준점(cutoff)으로 c를 사용하는 경우(Fig 1) c 이하를 비질병군(non-diseased), c 이상을 질병군(diseased)으로 분류하고 각 군에서의 검체 수를 적용하면 4개의 셀을 갖는 동일한 표가 작성된다. 진단검사의 정확도는 질병군에 대하여 검사 결과 양성으로 올바르게 분류할 확률인 민감도와 비질병군에 대하여 검사 결과 음성으로 올바르게 분류할 확률인 특이도로 요약한다.

ROC 곡선은 진단검사 결과가 명목형(nominal), 순위형(ordinal) 혹은 연속형(continuous) 척도로 측정된 경우 가양성률을 x축, 진양성률을 y축으로 나타낸 그림이다(Fig 2). 범주형 검사결과(test result)의 개수를 $h$라고 할 때 경험적(empirical, non-parametric) ROC 곡선은 $h-1$개의 좌표를 갖는다. 예를 들어 병리조직학적 검사에서 악성 종양(n = 255)과 양성 종양(n = 334)으로 확진된 환자를 대상으로 진단검사 A를 사용하여 환자를 5개의 범주로 분류한 가상의 결과가 Table 2와 같다고 할 때 악성 종양을 판정하기 위하여 C1-C4의 4개 기준점에 해당하는 좌표(가양성률과 민감도의 조합)에서 민감도와 가양성률을 계산할 수 있으며, 이를 선으로 연결하면 ROC 곡선이 작성된다.

### ROC 곡선의 면적 해석

진단검사의 정확도는 ROC 곡선의 면적(area under the ROC curve, AUC)으로 요약되며 이 값은 0-1의 범위를 갖고 면적이 넓을수록 진단검사의 판별능력(discriminatory ability) 즉 정확도(accuracy)가 높다는 것을 의미한다. AUC

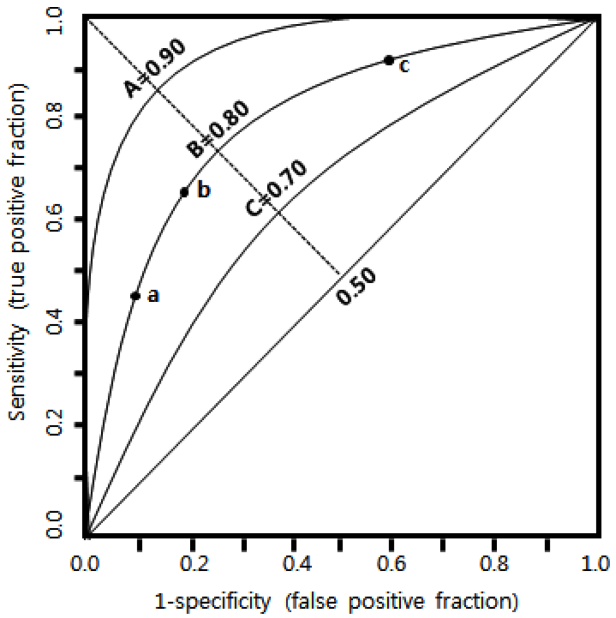


**Fig 2.** Three hypothetical ROC curves representing the diagnostic test accuracy. Each point on the curve represents the true-positive fraction (sensitivity) and false-positive fraction (1-specificity) associated with a diagnostic test value. The area under the curve (AUC) for A, B and C is 0.9, 0.8 and 0.7, respectively. An AUC value close to 0.5 indicates no discriminative value and is represented by a diagonal line extending from the lower left corner to the upper right. The bigger the AUC is, the better the overall performance of the diagnostic test. Therefore, as diagnostic test accuracy improves (better discriminating power), the ROC curve moves toward the upper left corner, and the AUC approaches 1. a = strict cutoff; b = moderate cutoff; c = lenient cutoff.

= 1.0은 민감도가 1이고 가양성률이 0이므로 완벽한 검사(perfect test)를 의미하며, AUC = 0은 모든 검사 대상자에 대하여 환자군을 정상군으로, 정상군을 환자군으로 완벽하게 잘못 분류하는 상황이다. 실질적인 의미에서 AUC의 하한값은 AUC = 0.5인 경우인데 이는 진양성률과 가양성률이 동일한 경우로 감염여부를 판정할 때 진단검사 결과가 동전을 던져 판정하는 것과 마찬가지로 유용한 정보를 제공하지 못한다는 것을 의미한다(noninformative test). 따라서 진단검사가 최소한의 유용성을 갖기 위해서는 $AUC > 0.5$이므로 진단검사의 유용성에 대한 통계적 검정에서는 AUC = 0.5에 관심을 두게 된다. ROC 곡선의 면적이 $0.5 < AUC \leq 0.7$일 때 낮은 정확도(less), $0.7 < AUC \leq 0.9$일 때 중등도(moderate), $0.9 < AUC < 1.0$일 때 높은(high) 정확도를 갖는다고 해석한다(15).

**Table 2.** Results from a hypothetical data for illustrative ROC plot

| Category of test result (G) | Disease status (gold standard) | | Decision rules (cutoff, C1-C4) | |
|---|---|---|---|---|
| | Malignant (D+) | Benign (D−) | Sensitivity | FPR |
| G1: Very likely benign | 11 | 98 | C1: 0.96 | C1: 0.71 |
| G2: Probably benign | 32 | 87 | C2: 0.83 | C2: 0.45 |
| G3: Possibly malignant | 52 | 54 | C3: 0.63 | C3: 0.28 |
| G4: Probably malignant | 75 | 63 | C4: 0.33 | C4: 0.10 |
| G5: Very likely malignant | 85 | 32 | | |
| Total | 255 | 334 | | |

Cutoff 1 classifies G1 category as negative and all other categories are positive. Cutoff 2 classifies G1 and G2 as negative and all other categories are positive. Cutoff 3 classifies G1, G2, and G3 as negative and all other categories are positive. Cutoff 4 classifies G1, G2, G3, and G4 as negative and G5 is the only category classified as positive. FPR = false positive fraction.

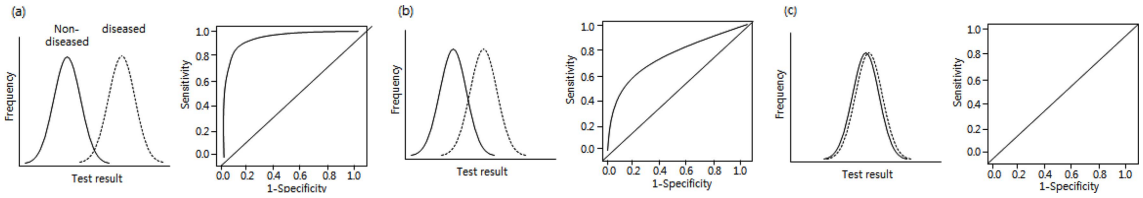


**Fig 3.** Distributions of test results for populations of diseased and non-diseased individuals and the corresponding ROC curves.

ROC 곡선에서 우측 상단의 민감도 = 가양성률 = 1인 좌표(1,1)에서 좌측 하단의 민감도 = 가양성률 = 0인 좌표(0,0)을 연결하는 대각선을 “chance diagonal” 이라고 하며 질병군과 비질병군을 전혀 구분하지 못한다(noninformative test, AUC = 0.5). 따라서 진단검사가 최소한의 유용성을 갖기 위해서는 ROC 곡선이 대각선 보다 높게 위치해야 하며, 진단적 가치가 낮은 검사일수록 ROC 곡선이 좌측에서 우측으로 이동하며 대각선에 접근하면 최악의 상황이라고 할 수 있다(13,15). 반면에 정확도가 높은 유용한 검사(집단을 분류하는 능력이 높은 검사)는 대각선에서 멀리 떨어지며, 유병률이 50%일 때 가양성률이 최소(특이도 최대)인 좌표 (0,0)에서 진양성률(민감도)이 최대가 되는 좌표 (0,1)로 수직으로 접근하며 이 경우 $Se(c) = 1$이고 $FP(c) = 0$이 되어 완벽한 검사라고 할 수 있다. 즉 ROC 곡선의 위치는 질병군과 비질병군에 대한 검사 결과가 중복되는 정도에 따라 결정되며, 이러한 관계를 도식화하면 Fig 3과 같다. 두 집단을 올바르게 분류하는 정도가 높을수록 즉 두 집단의 분포가 겹치는 정도가 낮을수록 진단의 유용성이 높고(Fig 3의 A), 반면에 두 집단에서의 검사결과가 겹치는 경우 ROC 곡선은 좌표 (0,0)에서 좌표 (1,1)으로 이어지는 45도의 직선의 형태를 보이며 모든 양성판정 기준점 $c$에서 $Se(c) = FP(c)$가 되어 유용성이 전혀 없는 검사라고 할 수 있다(Fig 3의 C).

### 모수 및 비모수 ROC 곡선

ROC 곡선 및 이와 관련된 요약 통계량을 추정하기 위한 방법은 매우 다양하며 흔히 비모수 기법(non-parametric method)과 모수(parametric method) 기법으로 구분한다(7, 16,20). 비모수 기법인 경험적 ROC 곡선(empirical, non-parametric; Fig 4의 A)은 검사결과의 분포에 대하여 가정을 전제로 하지 않는 방법으로 민감도와 위양성률을 비모수적

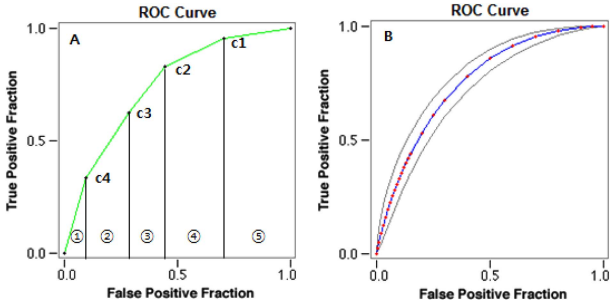


**Fig 4.** The empiric (A) and fitted (or smooth) (B) ROC curves with 95% confidence interval constructed from the Table 2. The discrete points on the empirical ROC curve are marked with dots. Four labeled points on empiric curve correspond to four cutoffs used to estimate sensitivity and false positive fraction. The area under curve (AUC) for empiric ROC curve is 0.744 and for fitted curve is 0.756. The nonparametric area under the empirical ROC curve is the summation of the areas of the trapezoids formed by connecting the points on the ROC curve, which is equivalent to the Mann-Whitney version for the rank sum test.

인 방법을 추정한 경우 관찰된 자료 쌍(기준점)을 선으로 연결하기 때문에 톱니(jagged) 혹은 계단(staircase) 형태를 보인다(20). 이 값은 로지스틱 회귀분석의 concordance index (c-index)와 Mann-Whitney U test (Wilcoxon rank-sum test)의 결과와 동일하다(13). 즉 ROC 곡선의 면적은 곡선의 각 지점에 대한 부등변사각형의 면적(areas of trapezium)을 모두 합하여 계산하는데 이를 trapezoidal rule이라고 하며(6), Fig 4의 A에서 5개 trapezoid(①-⑤)의 면적을 모두 합하면 경험적 AUC를 얻는다.

모수기법의 경우 실제 연구에서는 관찰하지 않았지만 표본크기를 확장하여 무수히 많은 지점을 연결함으로써 부드

러운 곡선(smooth) 모양을 적합하게 되는데 이를 binormal ROC 곡선 혹은 smooth ROC 곡선(Fig 4의 B)이라고 한다. 이 방법은 서로 다른 평균과 표준편차를 갖는 상호 독립인 두 집단의 검사결과가 정규분포를 따르거나 대수변환, 제곱근 변환, Box-Cox 변환 등을 통하여 정규성(normality)을 충족할 때 최대우도법(maximum likelihood)으로 추정한다(13,18,20). Table 2의 자료에 대하여 ROC 곡선을 작성하면 Fig 4와 같고, AUC는 모수적 방법의 경우 0.756, 비모수적 방법에서는 0.744로 추정된다.

**ROC 곡선의 부분 면적**

AUC는 진단검사의 총체적인 정확도(global summary measures)를 나타내는 지표이지만 임상적으로 의미가 없는 높은 위양성률에서 면적의 상당 부분을 기여하는 경우가 있다(8). 또한 두 ROC 곡선의 면적은 동일하지만 상호 교차하는 경우 특정한 구간에서는 검사 A가 우수하지만 나머지 구간에서는 검사 B가 더 우수할 수 있는데 이 경우 진단검사의 정확도를 요약하는 다른 방법으로 ROC 곡선의 부분면적(partial AUC, PAUC)을 계산한다(13,20). 이를테면 연구목적에 따라 기준점을 위양성률이 낮은 구간(sensitivity at fixed FPR, $FPR < 0.1$) 혹은 민감도가 높은 구간(FPR at fixed sensitivity, $Se > 0.95$) 등 연구자가 관심을 두고 있는 특정 구간에서의 정확도를 부분면적으로 계산하게 된다.

**ROC 곡선 응용**

ROC 곡선은 다음과 같은 장점을 갖는다. 첫째, 민감도와 특이도 단독으로는 검사의 특성을 파악하기 어렵지만 이들 지표 간의 관계를 통하여 다양한 의사결정의 기준점에서 진단검사의 특성을 비교하고 최적의 기준점을 찾는데 활용할 수 있다(21). 예를 들어 Table 4에서 연구의 목적이 종양이 확실한 환자만을 검출하는 것이 목적이라면 위양성률이 최소가 되어야 하므로 특이도가 높은 기준점인 C4로 설정한다면 민감도는 33%이고, 특이도는 90%가 된다. 한편 연구의 목적이 종양이 의심되는 모든 환자를 스크리닝하는 것이라면 높은 민감도가 필요하므로 기준점을 C1으로 설정하면 민감도는 96%이고, 특이도는 39%가 된다. Fig 2에서 검사 B는 세 가지 기준점을 예시하고 있는데 a는 가양성률이 가장 낮고 종양이 확실한 경우에만 양성으로 판단하므로 가장 엄격한 기준점이 되며(strict), b는 중등도(moderate), c는 상대적으로 느슨한(lax) 기준이라고 할 수 있다. 즉 ROC 곡선에서 엄격한 기준점일수록 좌표 (0,0)에 접근하며, 느슨할수록 좌표 (1,0)에 접근한다. 최적의 기준점을 찾는 방법으로 Youden index 계산, ROC 곡선의 좌표 (0,1)에서 특정 지점 간의 거리(distance)를 이용하는 방법, 진단검사에 소요되는 금전적인 비용(cost)을 고려한 상태에서 최적의 기준점을 선정하는 공식 등을 사용할 수 있다(8,12,13). 둘째, ROC 곡선은 2개 이상의 진단검사의 특성을 비교하는 용도로 사용할 수 있다. 또한 민감도와 특이도 단독으로는 어느 검사가 더 우수한지 파악하기 어렵지만 AUC를 이용하면 진단검사의 특성을 정량적으로 비교해주는 장점이 있다. 진단검사 3종(A, B, C)의 ROC 곡선을 비교한 Fig 2에서 검사 A의 정확도는 90%로 모든 가양성률 지점에 대하여 진양성률이 가장 높고, 모든 진양성률 지점에 대하여 가양성률이 가장 낮기 때문에 정확도가 가장 높고, 검사 C는 정확도가 70%로 가장 낮다고 할 수 있다. 셋째, ROC 곡선은 민감도와 특이도의 함수이므로 유병률에 영향을 받지 않는데 이는 모집단의 유병률에 관계없이 표본을 선발해도 된다는 것을 의미한다(8). 또한 경험적 ROC 곡선은 검사결과의 측정단위(척도)와는 무관하기 때문에 검사 간 정확도를 시각적으로 직접적인 비교가 가능하며, 대수변환이나 제곱근 변환이 가능하다는 장점이 있다. 넷째, AUC는 예측모형(prediction model)이나 예후모형(prognostic model)을 평가하는데 사용할 수 있다. 예측모형은 확률모형을 이용하여 위험에 노출된 특정한 환자가 질병에 감염될 확률을 추정하는 것이고, 예후모형은 질병에 감염된 환자가 회복하거나 사망하는 등 예후를 예측하는 모형이다. 이러한 모형에 대하여 경쟁관계에 있는 모형들의 정확도를 상호 비교하는 가장 일반적인 방법으로 c-index에 해당하는 AUC를 사용하며, 이 값은 해당 모형이 사건 발생과 비발생을 구분하는 정도에 대한 정량적인 수치를 제공한다.

## 감사의 글

본 연구는 농림축산식품부 가축질병대응기술개발사업(Animal Disease Management Technology Development, 과제번호: 315038-2, 강원대 C1012360-01-01)과 강원대학교 동물의학종합연구소의 지원으로 이루어진 것임.

## 참고문헌

1. Burfeind O, Sannmann I, Voigtsberger R, Heuwieser W. Receiver operating characteristic curve analysis to determine the diagnostic performance of serum haptoglobin concentration for the diagnosis of acute puerperal metritis in dairy cows. Anim Reprod Sci 2014; 149: 145-151.
2. Collinson P. Of bombers, radiologists, and cardiologists: time to ROC. Heart 1998; 80: 215-217.
3. Denis-Robichaud J, Dubuc J, Lefebvre D, DesCteaux L. Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. J Dairy Sci 2014; 97: 3364-3370.
4. Doi K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. Phys Med Biol 2006; 51: R5-27.
5. Enachescu V, Ionita M, Mitrea IL. Comparative study for the detection of antibodies to Neospora caninum in milk and sera in dairy cattle in southern Romania. Acta Parasitol 2014; 59: 5-10.
6. Eng J. Receiver operating characteristic analysis: a primer. Acad Radiol 2005; 12: 909-916.
7. Faraggi D, Reiser B. Estimation of the area under the ROC curve. Stat Med 2002; 21: 3093-3106.
8. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013; 4: 627-635.
9. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inform 2005; 38: 404-415.
10. Li J, Jiang B, Fine JP. Multicategory reclassification statistics

for assessing improvements in diagnostic accuracy. Biostatistics 2013; 14: 382-394.
11. Lusted LB. Logical analysis in roentgen diagnosis. Radiology 1960; 74: 178-193.
12. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8: 283-298.
13. Obuchowski NA. ROC analysis. AJR Am J Roentgenol 2005; 184: 364-372.
14. Swets JA. Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. Psychol Bull 1986; 99: 181-198.
15. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285-1293.
16. Wan S, Zhang B. Smooth semiparametric receiver operating characteristic curves for continuous diagnostic tests. Stat Med 2007; 26: 2565-2586.
17. Yang ZH, Li L, Pan ZS. Development of multiple ELISAs for the detection of antibodies against classical swine fever virus in pig sera. Virol Sin 2012; 27: 48-56.
18. Zhou XH, Li CM, Yang Z. Improving interval estimation of binomial proportions. Philos Trans A Math Phys Eng Sci 2008; 366: 2405-2418.
19. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. Stat Med 1997; 16: 2143-2156.
20. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 2007; 115: 654-657.
21. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561-577.