

의학적 의사결정 도구들에 대한 고찰: Bayesian analysis and ROC analysis

원광대학교 치과대학 구강악안면방사선학교실
이병도

Medical decision making tools : Bayesian analysis and ROC analysis

Byung-Do Lee

Department of Oral and Maxillofacial Radiology, School of Dentistry, Wonkwang University

ABSTRACT

During the diagnostic process of the various oral and maxillofacial lesions, we should consider the following: "When should we order diagnostic tests? What tests should be ordered? How should we interpret the results clinically? And how should we use this frequently imperfect information to make optimal medical decisions?" For the clinicians to make proper judgement, several decision making tools are suggested. This article discusses the concept of the diagnostic accuracy (sensitivity and specificity values) with several decision making tools such as decision matrix, ROC analysis and Bayesian analysis. The article also explain the introductory concept of ORAD program. (*Korean J Oral Maxillofac Radiol* 2006; 36 : 1-5)

KEY WORDS : Diagnosis; Sensitivity; Specificity

서 론

의사결정 (decision making)이란 문제를 해결하거나 목표를 달성하기 위해서 선택 가능한 대안들 중에서 최적의 대안을 선택하는 과정으로 정의할 수 있으며,¹ 의사 결정자는 가능한 한 모든 대안을 검토하여 좋은 결과를 도출할 필요가 있을 것이다.² 그러나 모든 대안을 검토하는 것은 현실적으로 어렵기 때문에 접근 가능한 안들 중에서 가장 좋다고 판단되는 대안을 선택하여 실행하는 것이 일반적인 의사 결정 과정이다. 질병의 진단과정 중에도 위와 유사한 의사결정 과정이 진행된다. 환자가 특정한 주소를 가지고 내원하면 질병을 진단하기 위한 검사시행 여부를 결정하고, 검사를 시행한다면 어떤 검사법을 사용할 것인지를 선별하게 된다. 이 때 질병을 진단하는 의사 (doctor) 들은 가능한 모든 검사를 시행하기 보다는, 본인의 의학 지식과 경험에 의존하여 검사 시행 유무 및 검사 방법을

결정하게 되며 의사 결정에 도움을 주는 도구가 사용되면 효과적이다. 이러한 보조도구 (decision making tools)에는 decision tree, decision matrix, Bayesian analysis, ROC analysis, cost-benefit analysis 등이 있으며,³ decision matrix, Bayesian analysis, ROC 분석법 등은 민감도 및 특이도와 깊은 관련이 있다.⁴

본 소고의 목적은 의사결정 보조 도구들에 대해 간략하게 고찰하고 진단과정을 형식화 (formulation)해 봄으로써 진단과정을 이해하고 효과적인 의학적 의사결정을 내리는 데 도움을 주고자 하는데 있다. 설명의 편의를 위하여 목차와 같은 순서대로 기술하였으며, 주로 참고문헌상의 설명과 그림을 인용하였다.

목 차

1. Decision matrix
2. Prior probability, Decision criterion
3. ROC analysis
4. Bayesian analysis
5. ORAD program

접수일 : 2005년 12월 5일; 심사일 : 2005년 12월 6일; 채택일 : 2006년 1월 26일
Correspondence to : Prof. Byung-Do Lee
Department of Oral and Maxillofacial Radiology, College of Dentistry, Wonkwang University, 344-2, Shinyong dong, Iksan city, Chunbuk 570-711, Korea
Tel) 82-63-850-1912, Fax) 82-63-857-4002, E-mail) eebydo@wonkwang.ac.kr

1. Decision matrix (Table 1)⁴

질병 검사법의 정확성을 평가하는 방법에는 민감도 (sensitivity)와 특이도 (specificity)가 있으며, gold standard를 기준으로 진단의 옳고 그름을 판별한다. 민감도와 특이도는 decision matrix로부터 산출이 가능하며 양성 예측치 (positive predictive value)와 음성 예측치 (negative predictive value)도 얻어진다. 예를 들어 치근단 방사선사진을 이용하여 치아우식증을 진단한다고 가정해 보자. 치아 우식증이 실제로 존재하는 치아들이 치근단 방사선사진에서 치아우식증으로 판독되는 확률이 민감도이고, 치아우식증이 존재하지 않는 정상적인 치아들이 정상적인 치아로 판독되는 확률이 특이도로 표현 된다. 이러한 민감도와 특이도는 질병 검사법 (예: 치근단 방사선사진촬영법)의 정확도를 표현하는 방법으로 사용되는데, 경우에 따라서는 검사법의 정확도보다는, 특정 환자가 치아우식증 등의 질병을 실제로 가지고 있을 확률에 더욱 관심이 가는 경우가 있으며 이러한 경우 양성 예측치와 음성 예측치가 이용된다.

2. Prior probability, Decision criterion^{2,5}

질병의 진단 과정을 형식화 하는데는 prior probability, decision criterion 등의 개념 이해가 필요하다. 병을 진단하는 과정에서 Fig. 1과 같은 모형을 생각해볼 수 있으며 이는 질병의 발생 확률을 수치화한 것이다. 만약 환자가 병을 실제로 가지고 있을 확률이 action threshold 보다 크게 되면 진단이나 치료 술식 등이 필요하며 exclusion threshold

Table 1. General decision matrix

Test result	Presence of disease		Total
	Present (D+)	Absent (D-)	
Abnormal (T+)	TP	FP	TP+FP
Normal (T-)	FN	TN	FN+TN
Total	TP+FN	FP+TN	

TP: true positive, FP: false positive, FN: false negative, TN: True negative. Sensitivity: TP/ TP+FN, Specificity: TN/ FP+TN, Positive Predictive value: TP/ TP+FP, Negative Predictive value: TN/ FN+TN

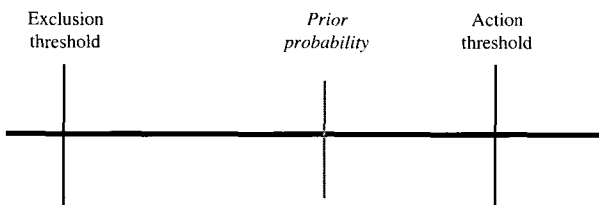


Fig. 1. The purpose of the diagnostic test is to obtain additional information in order to update the prior probability to a posterior probability. Prior probability: the prevalence of the disease in the population being tested. Exclusion: the absence of the disease assumable. Action: deciding whether to do a test or not to do it

threshold 하방이면 검사를 시도해 볼 가치가 없다. 여기에서 prior probability는 진단 검사가 시행되기 이전의 조사받은 집단의 유병율을 의미하는데, 통계적인 정보, 교과서내용, 진단하는 사람의 경험 등에 의해서 그 수치가 달라질 수 있으며 action threshold와 exclusion threshold 사이에 존재하는 prior probability의 위치를 가늠함으로써, 검사법을 선택하는데 도움이 될 수 있다. Fig. 2를 살펴보자. prior prob-

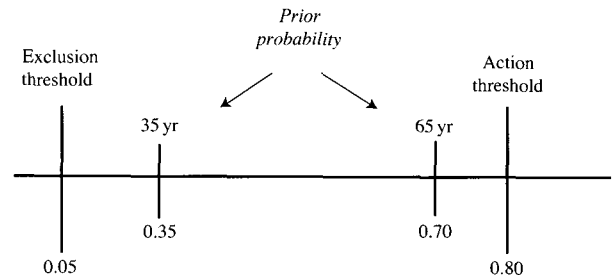


Fig. 2. Assumed "action" (0.80) and "exclusion" (0.05) thresholds for hypothetical example. Prior probability for certain disease (ex: myocardial ischemia) is significantly lower for a 35-year-old woman (0.35) than for a 65-year-old man (0.70).

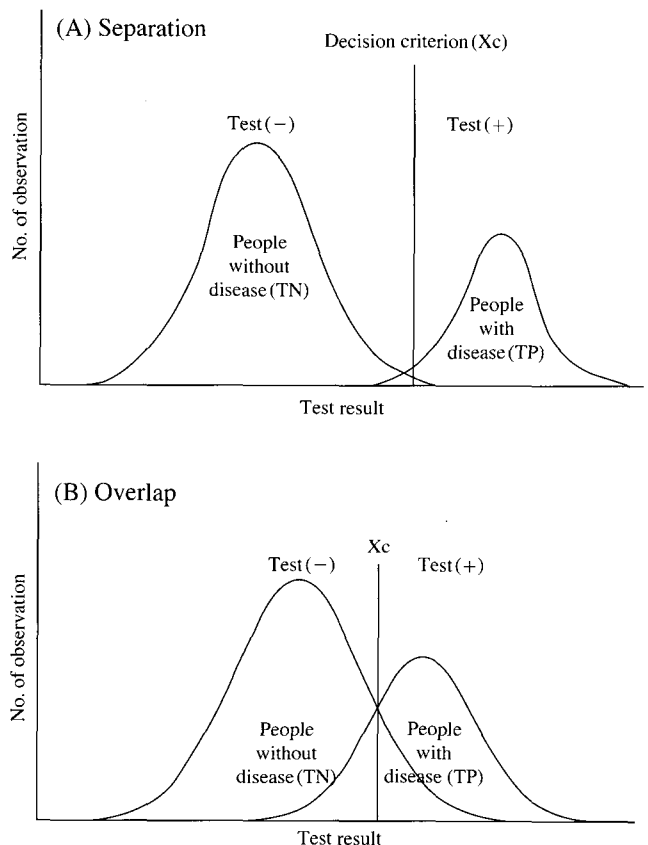


Fig. 3. The degree of separation between the abnormal and normal populations when a specific decision criterion is used determines the discriminating ability of the test. The selected decision criterion defines a positive or negative test.

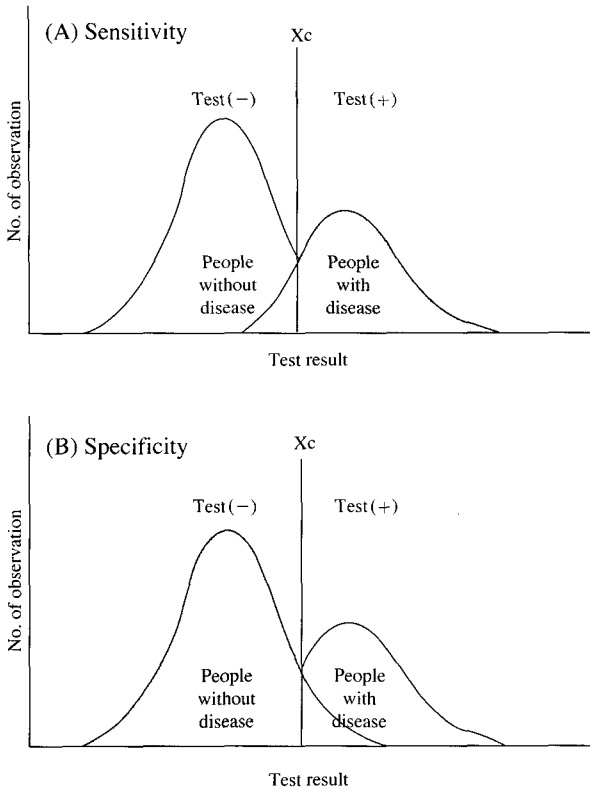


Fig. 4. The discriminating ability of a test using the defined decision criteria is characterized by the sensitivity (A) and the specificity (B).

ability가 낮고(0.35) action threshold가 높음으로(0.80), 그 간격이 넓다면, 검사력이 우수한 검사법을 선택하여야 할 것이고(35세 환자), 그 간격이 좁다면 값이 저렴한 검사 방법 혹은 비침습적인 방법을 선택하는 것이 좋다고 생각한다(65세 환자, prior probability 0.70, action threshold 0.80).

검사법의 우열을 평가시 피검자 수의 분포가 Gaussian 형태를 가지는 Fig. 3과 같은 도형을 가정해볼 수 있다. Fig. 3A와 같이 질병이 존재하는 군과 존재하지 않는 군이 분리될 수 있는 경우는 완벽한 질병 진단 검사법이 사용되면 가능하다. 그러나 완벽한 검사법보다는 불완전한 검사법이 많기 때문에 Fig. 3B와 같이 질병이 존재하는 군과 존재하지 않는 군이 중복된 형태가 일반적으로 관찰되며, 이러한 중복부위(overlap)의 크기는 질병 검사법의 판별력을 대변하게 된다. 즉 중복 부위가 커지게 되면 판별 능력이 우수하지 못하다는 것을 의미하고, 중복 부위가 적어질수록 질병 군으로부터 정상군을 쉽게 분리할 수 있는 능력이 있음을 암시한다고 할 수 있으며, 이러한 진단법의 판별력은 민감도와 특이도 등의 정확도로 표현될 수 있다 (Fig. 4). 이때 기준이 되는 decision criterion은 검사자의 주관에 따라 이동을 할 수 있으며 위양성(false positive)을 증가시킨다면 위음성(false negative)을 감소시킬 수 있다

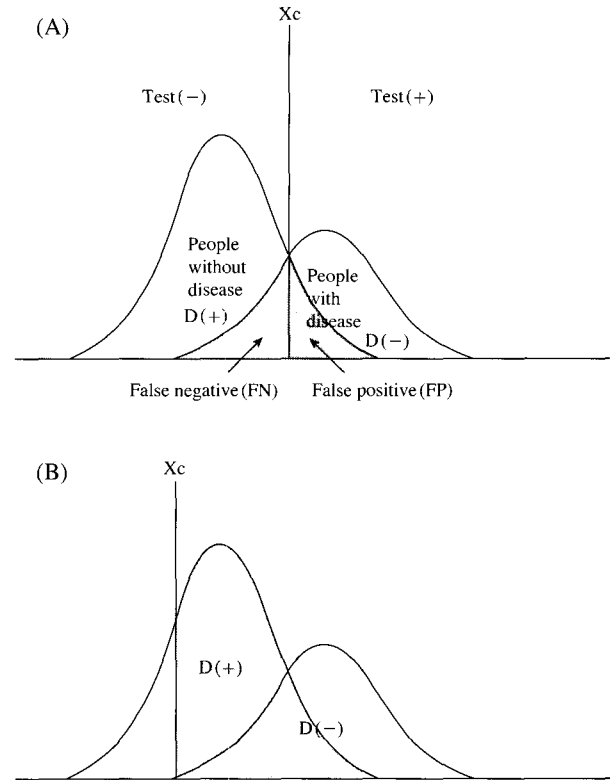


Fig. 5. (A) People with disease whose test results happen to be to the left of X_c will be FN, and normals whose results are to the right of X_c will be FP. (B) FN have been eliminated by moving X_c to the lower end of D(+) curve. As a result of eliminating FN, FP increased.

(Fig. 5). 검사결과 등에 대해 의사 결정을 내릴때는 위양성과 위음성의 장단점을 비교하여 장점이 많은 방향으로 결정하는 것이 추천된다. 예를 들어 방사선사진 상에서 초기 인접면 우식증이 의심되는 경우 위양성이 높게 의사 결정이 이루어진다면 건전한 치질에 와동 형성이 이루어질 위험성이 있으므로 주의해야 한다.

3. ROC (Receiver Operating Characteristics)^{3,5}

질병 진단시 “질병이 존재한다” 혹은 “존재하지 아니다”라고 단정할 수 있는 경우도 있지만, 질병이 존재하는 지가 모호하여 연속선(continuum) 상에서 판단하는 경우가 있다. 예를 들어 방사선사진상에서 치아우식증 존재 유무를 판독하고자 할때 1. 확실히 존재한다, 2. 존재하는 것 같다. 3. 잘 모른다. 4. 존재하지 않는 것 같다. 5. 확실히 존재하지 않는다 등의 5 단계로 구분하여 민감도와 특이도를 각각 계산할 수 있으며,⁶ 이를 그래프화 한 것이 ROC 곡선이다. 이 때 판독자의 주관에 따라 decision criterion이 좌우로 이동되며 이에 따라 민감도와 특이도가 각각 증감된다(Fig. 6A). 즉 ROC분석 방법은 올바르게 진단을 하는

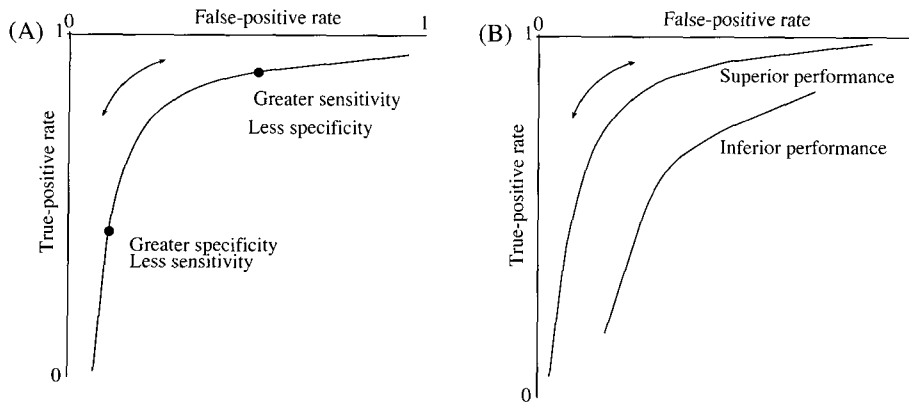


Fig. 6. (A) The reciprocal relationship between sensitivity and specificity as a function of changing the decision criterion can be expressed by the ROC curve. (B) Intrinsic test performance is directly related to overlap between normal and abnormal populations. The ROC curve is a good way to demonstrate the relative performance of diagnostic tests: the superior test ROC curve can be seen to be closer to the upper left-hand corner of the graph.

경우(true positive)와 그렇지 못한 경우(false positive)의 연관관계를 다양하게 표현하는 방법이라고 할 수 있겠다.

ROC 곡선 하방 면적(Az)은 질병 검사법의 정확도를 나타내는 척도로서 하방 면적이 클수록 우수한 질병 진단법이라고 할 수 있으며(Fig. 6B) 곡선이 좌상 측에 위치할수록 민감도와 특이도가 높다. 이러한 ROC 곡선은 spreadsheet 프로그램이나 특정 소프트웨어(Rokit, The University of Chicago)등을 이용하면 생성이 되며 ROC 곡선을 이용한 수많은 의학 논문이 작성되고 있다.

4. Bayes 분석방법(Bayesian analysis)^{2,4,7}

방사선사진 판독시 질병 존재 유무의 판단이 모호한 경우가 있는데 이때 Bayes 검사법을 이용하여 특정 질병의 발생 확률을 수치화 시킬 수 있다. Bayes 이론은 영국 수학자인 Thomas Bayes (1702-1761)에 의해 제시된 것으로서 의학적 의사결정에 중요한 위치를 차지하고 있으며, 검사법의 민감도와 특이도, 질환의 prior probability 등을 알고 있는 상태에서 새로운 자료값(예: new evidence such as symptom)이 수식에 추가됨으로써 질병 발생 확률이 계산된다.²

이 검사법은 1959년 경부터 임상적인 의사결정에 이용되기 시작했으며, Cushing's syndrome, 선천적인 심장병, 갑상선 기능검사법 등에 광범위하게 사용되었다.⁷ 그 간단한 수식은 Fig. 7에 소개되고 있으며 간단한 응용 예가 제시되어 있다. 갑상선기능항진증이 전 인구의 5%에서 발생한다고 가정해보고 불편증이 집단 인구의 10% 정도 존재하거나 갑상선기능항진증 환자군에서는 불편증이 30% 정도 존재한다고 가정하는 경우, 불편증을 가진 환자가 내원시 그 환자가 갑상선기능항진증을 가지고 있을 확률을 계산해보면 0.15로 계산이 된다. 즉 불편증(symptom)이란 새로운 특징으로 인해 갑상선기능항진증 진단 확률을 3배(0.05에서 0.15) 높일 수가 있다.

Bayes 분석법은 다양한 수식으로 표현이 되고 있으며, 흔히 사용되는 수식 2가지를 더 소개하면 Fig. 8⁵과 Fig. 9⁸

$$P(D/S) = \frac{P(D) P(S/D)}{P(S)}$$

P(D/S): Posterior Probability, the probability of the disease after a given test result is known

P(S): Probability of Symptom

P(D): Probability of Disease

P(S/D): Probability of specific symptom S being present, given that the patient has the disease D

In the case of Hyperthyroidism (D), insomnia (S)

P(D)=0.05 (5% of the population)

P(S)=0.1 (10% of the population)

P(S/D)=0.3 (30% of the population)

P(D/S)=(0.05 × 0.3)/0.1=0.15

Fig. 7. A simple example of Bayes equation.

$$P(D|T+) = \frac{(TPR * \text{Prior Probability})}{(TPR * \text{Prior Probability}) + [FPR * (1 - \text{Probability})]}$$

P[D|T+]: posterior probability, represents the probability of having the disease if the test is positive

TPR: true-positive rate or sensitivity

FPR: false-positive rate or (1-specificity).

Fig. 8. Bayes' theorem expressed with the concept of sensitivity, specificity and prior probability.

와 같다.

5. ORAD 프로그램^{7,9}

Bayes 분석법은 의학 분야에서 다양하게 응용되고 있으며^{10,11} White 교수(UCLA, Oral Radiology) 등이 공동 개발한 ORAD 프로그램 역시 Bayes 분석법과 알고리즘 접근법 등이 응용되었다. Web상의 ORAD 프로그램의 일례를 살펴보면(Fig. 10), 악골에 발생될 수 있는 140개 질환들을 서로 감별진단하기 위해 병소의 방사선학적 특징 중 위치, 크기, 경계 등 16가지의 방사선학적 특징을 이용자가 입력하도록 되어 있으며, 입력이 끝난 후 감별진단을 하도록

$$P(D+/T+) = \frac{P(T+/D+) \times P(D+)}{P(T+/D+) \times P(D+) + P(T+/D-) \times P(D-)}$$

P(D+/T+) = The conditional probability that given a positive diagnostic procedure (T+), disease is present (D+).

P(T+/D+) = The true-positive ratio, or the probability that given a disease (D+), the diagnostic procedure will result in positive findings (T+).

P(D+) = Incidence of disease (D+) in the population (or surfaces).

P(T+/D-) = The false-positive ratio, or the probability of a positive result (T+) when the disease is not present.

P(D-) = The probability that disease (D-) is not present. This is equivalent to 1-P(D+).

Fig. 9. Bayes' theorem may be written Using McNeil's notation⁸.

Table 2. A simple example of application of ORAD program

Condition	Prior probability	Age				Gender	
		0-16	12-25	26-50	>50	M	F
A (cyst)	2	10	30	20	40	60	40
B (tumor)	1	50	30	10	10	10	90
C (infection)	3	0	0	30	70	50	50

In the case of 35-year old man (for example)

	Relative probability	Age	Gender	Product
A (cyst)	2	20	60	2,400 (2 × 20 × 60)
B (tumor)	1	10	10	100 (1 × 10 × 10)
C (infection)	3	30	50	4,500 (3 × 30 × 50)

Probability of each condition
 C (infection) : 4,500/7,000=0.643, A (cyst) : 2,400/7,000=0.343, B (tumor) : 100/7,000=0.014

주소(D) <http://www.orad.org/cgi-bin/orad/patient.pl>

Patient Information

What is the sex of your patient?

What is the race of your patient?

What is the age of your patient?

Does your patient have pain or paresthesia?

Please estimate the number of lesions.

Which jaw contains the lesion?

The lesion center is in what region?

The contents of the lesions are:

What is the maximum size of the lesion?

The borders of the lesion are

The loculation of the lesion is:

Where is the origin of the lesion?

The relationship of the lesion to teeth is:

Does the lesion expand the bony cortex?

Does the lesion cause root resorption?

Does the lesion cause tooth displacement or impaction?

Shall we consider prevalence?

Fig. 10. An example of view page of ORAD program on WEB.

클릭하면 제일 가능성이 높은 질병들로부터 순서적으로 나열된다. 이러한 방법이 사용된 배경에는 악골 질환들이 때로는 특징적이지 못하기 때문에 판독하기가 어려운 경우들이 종종 있으며 의학적 의사결정 도구들이 감별진단에 도움을 줄 수 있기 때문이다. 그 간단한 응용 예가 Table 2에 설명되어 있다. 즉 낭종, 종양, 감염 등의 prior probability가 각 2, 1, 3이라고 가정하고, 연령대별, 성별에서 발생할 확률이 표와 같다면 35세 환자의 경우 감염, 낭종, 종양 등의 발생확율이 각각 64.3, 34.3, 1.4%로 계산되어 감별진단상의 순위를 정할 수가 있다.

결 론

의학 분야에서 ROC 분석과 Bayes 이론 등을 이용한 다양한 연구가 진행되어 왔으며, ORAD 프로그램 등은 악골

병소 감별진단에 도움을 줄 수 있는 유용한 프로그램이라고 생각한다. 위의 개념들을 구강악안면방사선과 전공자 등이 응용하여 악골 질환의 prior probability 정리 및 감별진단에 도움이 되도록 의사결정 도구들이 사용되었으면 하는 바램이 있으며, 아울러 Bayes 분석방법을 이용하여 악골 질환의 진단에 유용한 분석법을 연구해 보는 것도 의미가 있을 것이라고 생각한다.

참 고 문 헌

1. 김삼수, 김형섭, 배용환. 의사결정과 기업경영. 서울: 명경사; 2004. p.25.
2. Patton DD. Introduction to clinical decision making. Semin Nucl Med 1978; 8: 273-82.
3. Lusted LB. General problems in medical decision making with comments on ROC analysis. Semin Nucl Med 1978; 8: 299-306.
4. Douglass CW, McNeil BJ. Clinical decision analysis methods applied to diagnostic tests in dentistry. J Dent Educ 1983; 47: 708-14.
5. Chang PJ. The rational selection and interpretation of diagnostic tests. In: Erkonen WE. Radiology 101; The basics and fundamentals of imaging. Philadelphia: Lippincott-Raven; 1998. p.57-67.
6. Nair MK, Nair UP. An in-vitro evaluation of Kodak Insight and Ektaspeed Plus film with a CMOS detector for natural proximal caries: ROC analysis. Caries Res 2001; 35: 354-9.
7. White SC. Computer-aided differential diagnosis of oral radiographic lesions. Dentomaxillofac Radiol 1989; 18: 53-9.
8. McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med 1975; 293: 211-5.
9. White SC. Decision-support systems in dentistry. J Dent Educ 1996; 60: 47-63.
10. 조현자, 곽은영, 최철순. Bayes 이론을 이용한 유방촬영술의 판독 방법에 관한 연구. 대한방사선의학회지 1991; 27: 901-3.
11. 오유환, 박승용, 강은용, 박재성, 이기열, 김홍인 등. 단일폐결절의 악성도 가능성의 평가: 일반 판독방법과 Bayesian 분석방법의 비교연구. 대한방사선의학회지 1998; 38: 67-74.