# A Closer Look at Diagnosis in Clinical Dental Practice: Part 2. Using Predictive Values and Receiver Operating Characteristics in Assessing Diagnostic Accuracy

## (Examen approfondi du diagnostic en pratique clinique dentaire : Partie 2. Usage des valeurs prédictives et des fonctions d'efficacité du récepteur pour évaluer l'exactitude diagnostique)

• **Iain A. Pretty,** BDS(Hons), MSc, PhD •
• **Gerardo Maupomé,** PhD •

## Sommaire

*Lorsqu'un clinicien planifie d'utiliser un test ou une technique diagnostique, il est important qu'il établisse la probabilité qu'une personne est bel et bien atteinte d'un état pathologique ou d'une maladie; cette détermination dépend des valeurs prédictives qui sont influencées par diverses caractéristiques de la technique diagnostique. À cet égard, la sensibilité et la spécificité sont limitées parce qu'elles décrivent les résultats d'une technique d'une manière dichotomique : le résultat est soit positif, soit négatif. Pourtant, beaucoup de techniques cliniques ne sont pas dichotomiques, comme le sondage des poches parodontales ou l'évaluation des radiographies de caries, et dans ces situations, on examine une gamme de caractéristiques en vue d'atteindre un certain degré de certitude au sujet de la présence ou de l'absence de maladie. Le présent article examine les valeurs prédictives et l'analyse des fonctions d'efficacité du récepteur, soit un algorithme qui combine diverses caractéristiques statistiques des techniques diagnostiques pour évaluer l'efficacité des techniques non dichotomiques sans imposer de seuil arbitraire.*

***Mots clés MeSH :*** *decision support techniques; predictive value of tests; risk assessment/methods*

*© J Can Dent Assoc 2004; 70(5):313–6*
*Cet article a été révisé par des pairs.*

Part 1 of this series[1] introduced some of the basic concepts used in assessing diagnostic accuracy: reliability, validity, sensitivity and specificity. This article examines 2 additional concepts: predictive values and receiver operating characteristic (ROC) analysis.

### Predictive Values

By quantifying the sensitivity of a diagnostic procedure (see Part 1 of this series[1]) it is possible to determine one operating characteristic of that procedure to establish if a patient has the disease in question. Determining the specificity allows assessment of another operating characteristic of the procedure to determine if the patient does not have the disease. Sensitivity and specificity are relatively independent of the prevalence of a disease (the pretest probability that an individual patient has the disease), and therefore these parameters are generally stable for the same procedure administered in different study populations. In other words, sensitivity and specificity are inherent properties of the test. They are useful for comparing procedures and for deciding which test to use in a particular clinical setting. However, these values are not of great assistance to the clinician who wants an answer to one of the following questions: "I have a positive test result for this patient. How likely is it that the patient actually has the disease?"

***Table 1*** **A 2 × 2 contingency table illustrating the outcomes of a comparison between a diagnostic procedure and a gold standard and the use of these values to calculate negative and positive predictive values**

| | | Gold standard result | | |
|---|---|---|---|---|
| | | **Positive** | **Negative** | **Total** |
| **Procedure result** | **Positive** | True positive (TP) | False positive (FP) | TP + FP |
| | **Negative** | False negative (FN) | True negative (TN) | FN + TN |
| | **Total** | TP + FN | FP + TN | FN + TN + FP + TP |

Sensitivity = TP / (TP + FN)
Specificity = TN / (FP + TN)
Positive predictive value = TP / (TP + FP)
Negative predictive value = TN / (FN + TN)

Alternatively, "I have a negative test result for this patient. How likely is it that the patient is healthy?" Sensitivity and specificity do not aid in interpreting the result of a particular procedure for an individual patient; they do not help in ruling in or ruling out the disease once the results of the test are known, and so they have no predictive value. To answer these more practical questions, the predictive values of the diagnostic procedure must be determined.

The predictive values are easily derived from the contingency table described in **Table 1**. The positive predictive value (PPV) is the likelihood that the patient actually has the disease, given a positive test result.[2] The negative predictive value (NPV) is the likelihood that the patient does not have the disease should the procedure result be negative. Whereas the values for sensitivity and specificity depend only on the operating characteristics of the procedure itself, the PPV and NPV vary according to the prevalence of the disease. Thus, predictive values cannot be quoted without prior knowledge of disease prevalence in the population from which the estimates are being derived. In other words, PPV and NPV are not qualities of the procedure itself; rather, they are functions of both the characteristics of the procedure and the environment in which it is being used. Classic examples of the effect of prevalence on PPV and NPV have occurred where screening has been performed in "nontarget" populations, e.g., HIV tests in the general population. In this example, the prevalence of HIV infection was so low in the general population that the accuracy of PPV and NPV values was lower than random designation of individuals as infected or not infected. However, when the same screening procedures were applied to high-risk populations, they were highly effective in identifying those with the infection.

Sensitivity and specificity describe the results of a procedure in a dichotomous way.[3] For example, should a given tooth be extracted or not? Should this restoration be placed or not? However, many clinical procedures are not dichotomous, such as probing of periodontal pockets and assessment of radiographs for caries; with these procedures, a range of features must be examined to produce a degree of certainty regarding the presence or absence of disease. It is possible to assess the effectiveness of these tests, without simply imposing an arbitrary threshold, by using a technique known as receiver operating characteristic (ROC) analysis.

## Receiver Operating Characteristic Analysis

The use of ROC analysis has increased rapidly over the past 30 years, in particular following the publication of a landmark textbook by Swets and Pickett.[4] Early in its development, ROC analysis was conceived as an extension of signal theory, used by radar operators to appraise the strength of signals identified. Many of the early medical applications of ROC analysis were in the field of radiology, where subjective results are recorded on a rating scale. Today, the expansion of ROC analysis into the evaluation of a wide variety of diagnostic procedures yielding numeric results indicates its acceptance and its many applications.

ROC analysis is based on a graphic representation of the reciprocal relation between sensitivity and specificity, calculated for all possible threshold values. When sensitivity and specificity are analyzed jointly, a threshold score or cut-off must be set to divide patients into 2 categories: those presumed to have the disease and those presumed not to have the disease. A test scored on a continuous scale (i.e., not dichotomous) does not have just one value for the combination of sensitivity and specificity; rather, it has a range of values, with various possible cut-off points. Because reporting only one sensitivity–specificity pair may give an oversimplified picture of the performance of the diagnostic procedure, it is more useful to describe the entire range of values; plotting each pair of scores on an ROC plot is a good way to do this.

The true-positive probability (sensitivity) is plotted as a function of the false-positive probability (1 – specificity), for the entire range of cut-off points. The resulting ROC curve provides a graphic summary of the range of decision thresholds for the test. As the curve approaches the upper left corner of the plot, the true-positive fraction (TPF) approaches 1 (perfect sensitivity) and the false-positive fraction (FPF) approaches zero (perfect specificity); the closer the curve to the corner, the greater the overall accuracy of the test. The ROC plot also allows the results of 2 or more different tests to be graphed together, allowing a visual comparison of the performance of the tests. An example of an ROC analysis is shown in **Figs. 1a** and **1b**. Each of the numbered threshold values shown in **Fig. 1a** corresponds to an operating point on the ROC curve of **Fig. 1b**. When a high diagnostic threshold is used (point 1), all patients are determined to be nondiseased, which results in a zero value for both TPF and FPF. This situation connotes perfect specificity (100%) and is exemplified by the operating point in the lower left-hand corner of the ROC curve (**Fig. 1b**). When a very low diagnostic threshold is used (point 5), all patients are determined to be diseased, both

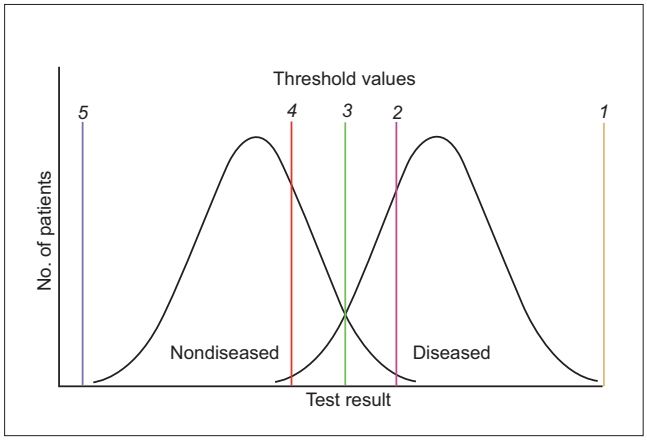


***Figure 1a:*** *Results of a diagnostic test performed in a patient population. Each numbered, coloured line (1 through 5) represents a threshold value that could be used as a diagnostic cut-off.*

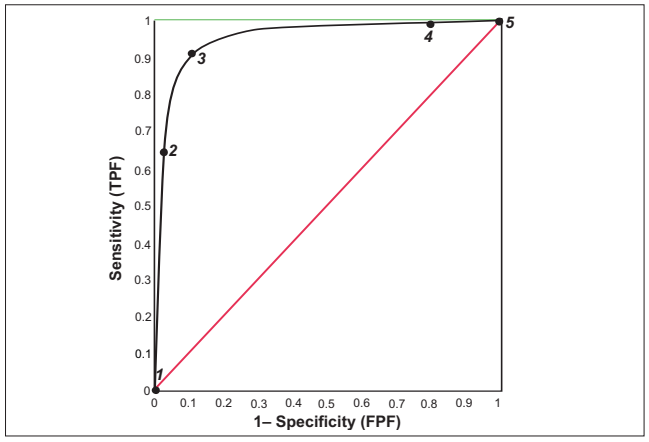


***Figure 1b:*** *Receiver operating characteristic (ROC) curve for these data. Each numbered point on the ROC curve relates to the corresponding threshold shown in Fig. 1a. The horizontal green line represents a perfect procedure, with no overlap between the distributions of diseased and healthy patients. The diagonal red line represents a procedure for which the diseased and nondiseased distributions overlap completely, a procedure that would have no discriminative value. TPF = true-positive fraction, FPF = false-positive fraction.*

TPF and FPF are 1, specificity is 0%, and the operating point appears in the upper right-hand corner of the curve (**Fig. 1b**). The other threshold values represent intermediate points of specificity and sensitivity between these 2 extremes.

An ROC curve represents the relation between sensitivity and specificity (and hence is a test to determine these values) when clinicians are allowed a degree of uncertainty in their decision-making not afforded in dichotomous decisions.[5] The method is equivalent to repeatedly asking clinicians to make simple, dichotomous decisions but with different treatment attitudes or thresholds. An example of this situation was presented when dentists were asked to assess caries in 2 groups of patients, one group who would return for re-evaluation in 6 months and a second group who would not return for a clinical exam until 2 years later.[6] In these 2 situations, a different decision might be made on the basis of the same clinical picture. Dentists may be more aggressive in their treatment of a hypothetical patient with poor attendance for follow-up than for a patient whom they can monitor regularly.

The discriminative ability of a procedure is defined by the distributions of diseased and nondiseased patients. The overlap of these groups determines the shape and position of the ROC curve. A straight line from the lower-left corner to the upper-right corner (shown in red in **Fig. 1b**) describes a procedure in which the diseased and nondiseased distributions overlap completely and the TPF and FPF are equal at any threshold. This procedure has no discriminative value and is worthless. A perfect procedure has no overlap between the distributions of diseased and healthy patients and would result in the straight line shown in green in **Fig. 1b**.

## Area under the Curve

In addition to the relative simplicity of this visual representation of test accuracy, it is possible to perform quantitative analysis yielding summary indices of the discriminatory accuracy of the test. The most common summary index is the area under the curve (AUC), that is, the area under the ROC curve. The AUC is a measure of the accuracy of a diagnostic procedure and is frequently used for comparisons between procedures or observers.[5,7] With statistical software, it is easy to compute and test multiple AUCs for significant differences by means of *z*-scores (univariate).[8] ROC curves can be generated for each observer in a study, the corresponding AUC values calculated and the results compared. It is also possible to pool data from various observers and produce a single ROC curve. If there are different groups of examiners, the AUCs can be compared to identify differences between groups, typically by means of a paired *t*-test. Some authors have stated that pooling results to create ROC curves can be misleading, in that it ignores the effect of case sample variation,[9] but this issue has been addressed by ensuring that each examiner assesses the same cases.

In the example illustrated in **Fig. 1b**, the AUC for the procedure that yields no discriminative value (represented by the red diagonal line) has a value of 0.5 or 50%. It is no better than random assignment of positive and negative results (e.g., by flipping a coin). The ROC line for a perfect procedure, represented in green, has an AUC of 1.0 or 100%. The results from diagnostic procedures used in real life fall between these 2 extremes. The closer the AUC value is to 1.0 or 100%, the more accurate the procedure.

## Conclusions

In the first 2 articles of this series examining diagnostic procedures and their operating characteristics in dental practice, the statistical methods and models for determining the accuracy of procedures have been described, along with their use for dichotomous, continuous and multiple-threshold data. Armed with knowledge of these procedures and the applications that will be outlined in the next 2 articles of the

series, readers will have a better understanding of diagnostic tests and the weight that can be afforded to the results of those tests. In particular, the third article will describe dental diagnostic procedures that have been assessed with ROC analyses; examples of such procedures or equipment include conventional[10] and digital[11] radiography, electronic caries monitors,[12] optical caries detectors,[13] plaque detection,[14] periodontal diagnosis[15] and sialography.[16] The use of ROC analysis may lead to a reduction in the use of some procedures and perhaps an increase in the use of others.

The final 2 articles in the series will describe novel techniques that may be introduced to dental practice in the future and will attempt to gauge whether such innovations are likely to represent any improvement over existing clinical approaches. ◆

*Le **Dr Pretty** est chargé de cours en prosthodontie, Université de Manchester, Manchester, R.-U.*

*Le **Dr Maupomé** est chercheur, Centre de recherche en santé, Portland (Oregon); professeur adjoint, Université de la Californie à San Francisco, San Francisco (Californie); et professeur en clinique, Université de la Californie, Vancouver.*

***Écrire au :** Dr Iain A. Pretty, Unit of Prosthodontics, Department of Restorative Dentistry, University Dental Hospital of Manchester, Higher Cambridge St., Manchester, M15 6FH, England. Courriel : iain.pretty@man.ac.uk.*

*Les auteurs n'ont aucun intérêt financier déclaré.*

## Références

1. Pretty IA, Maupomé G. A closer look at diagnosis in clinical dental practice: Part 1. Reliability, validity, specificity and sensitivity of diagnostic procedures. *J Can Dent Assoc* 2004; 70(4):251–5. Available from: URL: http://www.cda-adc.ca/jcda/vol-70/issue-4/251.html. Part 1 contains a glossary of epidemiology terms.

2. Everitt BS. Statistical methods for medical investigators. London: Edward Arnold; 1989.

3. van Erkel AR, Pattynama PM. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *Eur J Radiol* 1998; 27(2):88–94.

4. Swets JA, Pickett RM. Evaluation of diagnostic systems: methods from signal detection theory. New York (NY): Academic Press; 1982.

5. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143(1):29–36.

6. Kay EJ, Knill-Jones R. Variation in restorative treatment decisions: application of receiver operating characteristic curve (ROC) analysis. *Community Dent Oral Epidemiol* 1992; 20(3):113–7.

7. Hanley JA, McNeil BJ. A method of comparing receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148(3):839–43.

8. Metz CE KH. Statistical significance tests for binomial ROC curves. *J Mathematical Psychology* 1980; 22:218–43.

9. Swaving M, van Houwelingen H, Ottes FP, Steerneman T. Statistical comparison of ROC curves from multiple readers. *Med Decis Mak* 1996; 16(2):143–52.

10. Vaarkamp J, ten Bosch JJ, Verdonschot EH, Bronkhoorst EM. The real performance of bitewing radiography and fiber-optic transillumination in approximal caries diagnosis. *J Dent Res* 2000; 79(10):1747–51.

11. Ramesh A, Tyndall DA, Ludlow JB. Evaluation of a new digital panoramic system: a comparison with film. *Dentomaxillofac Radiol* 2001; 30(2):98–100.

12. Ashley PF, Blinkhorn AS, Davies RM. Occlusal caries diagnosis: an in vitro histological validation of the Electronic Caries Monitor (ECM) and other methods. *J Dent* 1998; 26(2):83–8.

13. al-Ismaily M, Chestnutt IG, al-Khussaiby A, Stephen KW, al-Riyami A, Abbas M, and others. Prevalence of dental caries in Omani 6-year-old children. *Community Dent Health* 1997; 14(2):171–4.

14. Sagel PA, Lapujade PG, Miller JM, Sunberg RJ. Objective quantification of plaque using digital image analysis. *Monogr Oral Sci* 2000; 17:130–43.

15. Hildebolt CF, Vannier MW, Shrout MK, Pilgram TK. ROC analysis of observer-response subjective rating data — application to periodontal radiograph assessment. *Am J Phys Anthropol* 1991; 84(3):351–61.

16. Yoshiura K, Kanda S. Analysis of the diagnostic process in sialography. *Dentomaxillofac Radiol* 1990; 19(4):149–56.