ELSEVIER

Journal of Clinical Epidemiology 61 (2008) 866–874

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

The agreement chart as an alternative to the receiver-operating characteristic curve for diagnostic tests

Shrikant I. Bangdiwala[a,*], Ana S. Haedo[b], Marcela L. Natal[c], Andrés Villaveces[a,d]

[a]*University of North Carolina, Chapel Hill, NC, USA*
[b]*Universidad de Buenos Aires, Buenos Aires, Argentina*
[c]*Universidad Nacional de Mar del Plata, Mar del Plata, Argentina*
[d]*CISALVA Institute, Universidad del Valle, Cali, Colombia*

Accepted 26 April 2008

Abstract

Objective: For diagnostic tests, the most common graphical representation of the information is the receiver-operating characteristic (ROC) curve. The ''agreement chart'' displays the information of two observers independently classifying the same *n* items into the same *k* categories, and can be used if one considers one of the ''observers'' as the diagnostic test and the other as the known outcome. This study compares the two charts and their ability to visually portray the various relevant summary statistics that assess how good a diagnostic test may be, such as sensitivity, specificity, predictive values, and likelihood ratios.

Study Design and Setting: The geometric relationships displayed in the charts are first described. The relationship between the two graphical representations and various summary statistics is illustrated using data from three common epidemiologically relevant health issues: coronary heart disease, screening for breast cancer, and screening for tuberculosis.

Results: Whereas the ROC curve incorporates information on sensitivity and specificity, the agreement chart includes information on the positive and negative predictive values of the diagnostic test.

Conclusion: The agreement chart should be considered as an alternative visual representation to the ROC for diagnostic tests.

Keywords: Predictive values; Sensitivity; Specificity; Likelihood ratios; Area under the curve; *B*-statistic

1. Introduction

In diagnostic tests, one is interested in examining how well a test performs in correctly identifying true positives and true negatives. In studying the performance of a test in a given population, one is also interested in knowing the predictive values of the diagnostic test. For diagnostic tests, the most common graphical representation of the test information is the receiver-operating characteristic (ROC) curve. Various quantifications are available for the test information, such as the test sensitivity, specificity, predictive values, likelihood ratios, the area under the ROC curve (AUC), and Youden's index [1]. The ROC is a plot in the unit square of the test sensitivity (true-positive proportion) against the test 1 − specificity (false-positive proportion) (see Fig. 1). The term ''curve'' arises when the diagnostic test result is on a continuous scale, and various cutpoints along the scale are selected, giving rise to different (1 − specificity, sensitivity) points within the unit square.

The ''agreement chart'' displays the information contained in the square contingency table of two observers independently classifying the same *n* items into the same *k* categories [2]. When the number of categories *k* is equal to two, the 2×2 contingency table contains the information from a diagnostic test if one considers one of the ''observers'' as the diagnostic test outcome and the other as the known or true outcome (see Table 1).

The agreement chart that corresponds to the information in Table 1 is presented in Fig. 2, and provides a visual assessment of agreement by comparing areas based on the cell frequencies from the contingency table. The row and column marginal totals determine rectangles within the larger square determined by the sample size *n*. The frequencies in the diagonal cells from the contingency table determine darkened areas of perfect agreement within the rectangles. The unshaded areas within the rectangles represent the off-diagonal cell entries of disagreement. The larger the darkened area within the rectangles, the larger

* Corresponding author. Department of Biostatistics, University of North Carolina, 137 E. Franklin Street, Suite 203, Chapel Hill, NC 27514, USA. Tel.: +1-919-962-3266; fax: +1-919-962-3265.
E-mail address: kant@unc.edu (S.I. Bangdiwala).

0895-4356/08/$ − see front matter
doi: 10.1016/j.jclinepi.2008.04.002

What is new?

Key findings

1. The agreement chart serves as a tool for summarizing the characteristics of diagnostic tests and measures of agreement in a single visual representation.
2. The agreement chart can be applied to multiple diagnostic tests and enhances the visualization of diagnostic test characteristics.

What this adds to what was known

3. Alternative visual tools for diagnostic tests, such as the agreement chart, can provide more comprehensive evaluations useful for clinicians.

What is the implication, what should change now?

4. The agreement chart should be considered as an alternative visual representation to the receiver-operating characteristic (ROC) curve for diagnostic tests.
5. The agreement chart is to be preferred over the ROC curve if one is interested in displaying the diagnostic tests' predictive values.

degree of agreement there is between the two observers, or in the case of the diagnostic test, between the diagnostic test and the correct diagnosis.

The corresponding *B*-statistic for the agreement chart quantifies the amount of agreement between the two observers as a proportion, adjusting the observed area of agreement determined by the cell entries in the diagonal of the contingency table with that expected owing to chance as constrained by the marginal totals of the same contingency table. The *kappa*-statistic [3] is the most commonly used measure for quantifying agreement, and it has been extended for multiple observers and partial agreement. It is also used for 2×2 tables, although it has been criticized [4]. The comparative behavior of *B* and *kappa* has been reported by Muñoz and Bangdiwala [5] for 3×3 and 4×4 tables, but not for 2×2 tables.

This article examines the relationship between the two graphical representations available for describing a diagnostic test and among the summary statistics for diagnostic tests. Specifically, we study if there is a geometric relationship between the AUC and the *B*-statistic.

2. Diagnostic tests

2.1. ROC curve

In evaluating a diagnostic test when the outcome is either binary or dichotomized into two well-defined groups such as ''diseased'' (positive) and ''nondiseased'' (negative or ''normal'') subjects, the accuracy of the test is judged using the two basic characteristics of sensitivity (Se) and specificity (Sp). Sensitivity is the probability that a diseased person has a ''positive'' result (true positive) from the test, whereas specificity is the probability that a ''normal'' person gets a ''negative'' result (true negative) from the test. These probabilities are estimated from the contingency table by the proportions:

$$\text{Sensitivity} = a_{11}/(a_{11} + a_{21}),$$
$$\text{Specificity} = a_{22}/(a_{12} + a_{22}).$$

Ideally, both these proportions should be close to 1 for a diagnostic test to be judged accurate. Biggerstaff [6] proposed the ''likelihood ratios graph,'' essentially using the (sensitivity, 1 − specificity) point of a binary diagnostic test and comparing several such tests based on their likelihood ratios. Youden's index *c* [1] is the difference between the true-positive rate and the false-positive rate:

$$c = \text{Se} - (1 - \text{Sp}) = \text{Se} + \text{Sp} - 1.$$

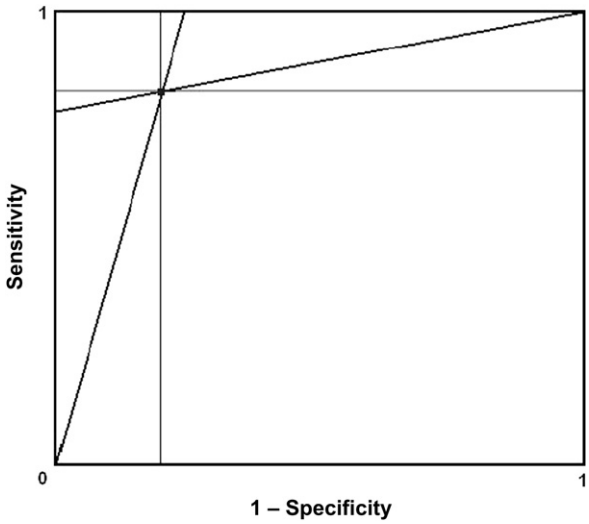

Fig. 1. Receiver-operating characteristic (ROC) curve for a binary outcome diagnostic test; hypothetical data (Biggestaff's ''likelihood ratio'' graph).

Table 1
A 2×2 contingency table for a diagnostic test

		Correct outcome: Positive	Correct outcome: Negative	Total
Diagnostic test Result	Positive	a_{11}	a_{12}	$a_{11} + a_{12}$
	Negative	a_{21}	a_{22}	$a_{21} + a_{22}$
	Total	$a_{11} + a_{21}$	$a_{12} + a_{22}$	n