ROC Curve Analysis: An Example Showing the Relationships among Serum Lipid and Apolipoprotein Concentrations in Identifying Patients with Coronary Artery Disease

Mark H. Zweig,1 Steven K. Broste,2 and Richard A. Reinhart3

Clinical accuracy, defined as the ability to discriminate between states of health, is the fundamental property of any diagnostic test or system. It is readily expressed as clinical sensitivity and specificity, and elegantly represented by the receiver operating characteristic (ROC) curve. To demonstrate the use of ROC curves, we reexamine a study of the ability of serum lipid and apolipoprotein measures to discriminate among degrees of coronary artery disease in patients undergoing coronary angiography. ROC curve analysis reveals that none of these indexes is highly accurate, but demonstrates a modest increase in the accuracy of apolipoprotein over lipid indexes.

What does it mean to say that a clinical laboratory test is accurate? What is “good” performance? Although a decision about the management of patient care usually is not made on the basis of one test result alone, each test result should address a particular clinical issue or question. Does this patient have a myocardial infarct or not? Is this prostate enlargement benign or malignant? Does this patient with Cushing’s syndrome have Cushing’s disease or is there a nonpituitary etiology? Has this patient’s malignancy responded to chemotherapy or not? Is this patient’s anemia a result of iron deficiency or something else?

As laboratorians, we frequently consider, discuss, and attempt to assess test performance. We are regularly presented with data on test performance and must make decisions about what to offer in our laboratory. It is necessary, then, to have in mind a concept of performance when operating a clinical laboratory. Swets (1) discusses a clear and simple concept of accuracy: the ability of the test or test system to discriminate between state A (e.g., hypercalcemia of malignancy) and state B (e.g., some other cause of hypercalcemia).

Accuracy addresses how good the information is rather than how useful it is. A test’s usefulness, or practical value, in managing patient care is another important issue, which involves many more considerations. Accuracy, however, is simple, straightforward information about the quality of the information contained in the test result. Accuracy, unlike other commonly used indexes such as predictive value and efficiency, is strictly a property of the test itself and independent of disease prevalence. Therefore, evaluating accuracy is an appropriate initial step in assessing test performance.

Swets (1) recommends receiver operating characteristic (ROC) curves as “a precise and valid measure of diagnostic accuracy.”4 Laboratorians (2–6) and others (7, 8) have discussed and used ROC curves as well. In this paper, we use serum lipid and apolipoprotein tests to illustrate this concept of accuracy and to demonstrate its simplicity, elegance, and power. This paper does not make judgments or reach conclusions about the clinical usefulness of any specific test.

A current public health goal is to identify persons with coronary artery disease (CAD) that is advanced enough to warrant dietary modification, exercise, medication, coronary angioplasty, or coronary bypass surgery. An earlier report examined the use of blood lipid and lipoprotein concentrations to discriminate between subjects with and without significant CAD, as defined by angiography, using multiple logistic regression analysis (9). It concluded that apolipoprotein (apo) A-I and B gave more information for predicting CAD than did standard lipid determinations such as total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, or the ratio of HDL to total cholesterol.

We report a new analysis of the data, which is different in several ways from the report already published. We use ROC curve analysis to assess the diagnostic accuracy of each test. In addition to examining the power of each index to discriminate between no CAD and any CAD, we assess their use in discriminating between mild CAD and severe CAD. Finally, we divide the subjects by sex to see whether there are differences in accuracy. These analyses are not meant to suggest a particular clinical application or management strategy but to illustrate the use of ROC curve analysis in evaluating diagnostic performance at its most elementary level.

Materials and Methods

The 394 adult patients (272 males, 122 females) were evaluated by cardiac catheterization for symptoms of CAD or for valvular or congenital heart disease. Arteriograms were classified as follows: category 1, normal,

---

1 Clinical Pathology Department, Warren G. Magnuson Clinical Center, National Institutes of Health, Bethesda, MD 20892.
2 Department of Epidemiology and Biostatistics, Marshfield Medical Research Foundation, 1000 North Oak Avenue, Marshfield, WI 54449.
3 Department of Cardiology, Marshfield Clinic, 1000 North Oak Avenue, Marshfield, WI 54449.
4 Nonstandard abbreviations: ROC, receiver operating characteristic; CAD, coronary artery disease; apo, apolipoprotein; HDL, high-density lipoprotein; and LDL, low-density lipoprotein.
5 Received February 19, 1992; accepted April 6, 1992.
Table 1. Areas under ROC Curves for Patients Undergoing Coronary Angiography

<table>
<thead>
<tr>
<th>Index</th>
<th>None vs. any CAD, all patients</th>
<th>None vs. any CAD, males only</th>
<th>None vs. any CAD, females only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td>0.65</td>
<td>0.69</td>
<td>0.57</td>
</tr>
<tr>
<td>apo B</td>
<td>0.68</td>
<td>0.74</td>
<td>0.62</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>0.62</td>
<td>0.7</td>
<td>0.55</td>
</tr>
<tr>
<td>apo A-I</td>
<td>0.60</td>
<td>0.58</td>
<td>0.61</td>
</tr>
<tr>
<td>apo A-I/ apo B</td>
<td>0.71</td>
<td>0.73</td>
<td>0.66</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>0.59</td>
<td>0.54</td>
<td>0.62</td>
</tr>
<tr>
<td>HDL/total cholesterol</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
</tbody>
</table>

no evidence of atherosclerotic CAD; 2, mild, evidence of plaque and (or) narrowing of the coronary artery to <70% of the luminal diameter; 3, moderate, ≥70% narrowing of the luminal diameter of one or two major coronary branches or primary branches; and 4, severe, ≥50% obstruction of the luminal diameter of the left main coronary artery and (or) ≥70% narrowing of the diameter of all three major coronary arteries and (or) a major branch of all these vessels. Among the 394 subjects, 58 were classified as normal, and 336 had CAD in categories 2, 3, or 4 (any CAD). Among the 272 males, 33 were considered normal, and 239 had CAD. Of the 122 women, 25 were normal and 97 had CAD. Overall, there were 44 subjects with mild CAD and 111 with severe CAD. Among males only, there were 27 classified as mild and 86 classified as severe. The original study consisted of 504 subjects. However, those with any missing data were omitted from this study.

Serum total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, apo A-I, and apo B were determined as previously reported (9). The interassay precision was 1.7% at 6.6 mmol/L for total cholesterol, 2.7% at 1.2 mmol/L for HDL cholesterol, 4% at 3.0 mmol/L for triglycerides, 5.0% at 98 mg/dL for apo A-I, and 3.5% at 113 mg/dL for apo B. Three logistic models were studied. Model 1 used total cholesterol, the ratio of HDL to total cholesterol, age, smoking history, and sex. Model 2 used apo B, the ratio of apo B to apo A-I, age, smoking history, and sex. Model 3 used all four lipid indexes and the three covariates.

Discriminatory power (diagnostic accuracy) was assessed in two ways: the ability to distinguish no CAD from any CAD, and the ability to distinguish mild CAD from severe CAD. ROC curve analyses were performed by using a beta version of Rulemaker™ (Digital Medicine, Inc., Lebanon, NH), a program for the Macintosh personal computer. The ROC curve represents the full spectrum of possible sensitivity-specificity pairs (corresponding to all of the possible decision levels) for a test in a particular clinical application. The closer the ROC curve approaches the upper left corner, reflecting high sensitivity and a low false-positive rate (high specificity), the greater its accuracy. One useful quantitative measure of accuracy is the area under the curve (1). The area of interest varies from 1.0, which corresponds to perfect discrimination, to 0.5, where no discrimination exists. Swets (1) suggests the following guidelines for interpretation of this area: 0.5–0.7, rather low accuracy; 0.7–0.9, accuracies useful for some purposes; and >0.9, rather high accuracy.

Results

ROC Areas

For discriminating no CAD from any CAD among all subjects, the areas under the ROC curves ranged from 0.59 to 0.77 (Table 1), moderate accuracy by Swets's criteria. Model 3 has the highest area, as expected (because models 2 and 1 are subsets of model 3), but this is not statistically higher than the area for model 2. Models 2 and 3 have significantly larger areas (P <0.04 and P <0.02, respectively) than does model 1, reflecting the superior discriminating ability of apolipoproteins over lipids. Furthermore, the area under the curve for apo B, the major protein constituent of LDL, is greater than that for either LDL cholesterol (P <0.02) or total cholesterol (P <0.02). The area for apo A-I, a major constituent of HDL, is slightly larger than the area for HDL cholesterol (not significant) and lower than the areas for the HDL:total cholesterol ratio, total cholesterol, or LDL cholesterol. On the other hand, the area under the curve for the ratio of apo A-I to apo B is greater than that for total cholesterol (P <0.04), HDL cholesterol (P <0.02), LDL cholesterol (P <0.04), or the ratio of HDL to total cholesterol (P <0.02). This supports the conclusion that apolipoproteins offer some increase in discriminating ability over traditional lipid measures.

The database was further subdivided by sex to examine the accuracy of these indexes when applied solely to one sex. In most cases, areas for one sex are similar to or lower than, and only occasionally slightly higher than, areas for both sexes taken together (Table 1).

We also were interested in whether subjects with mild CAD had different lipid or apolipoprotein concentrations than did subjects with severe CAD. We generated ROC curves based on these two groups alone (excluding subjects with no or moderate CAD). The null hypothesis would be as follows: Blood lipid or apolipoprotein concentrations in individuals with severe CAD are no different than they are in individuals with mild CAD. In fact, wo
found that all indexes had rather poor accuracy in distinguishing subjects with mild CAD from those with severe CAD. The areas ranged from 0.46 to 0.57. None was significantly different from 0.50, which is the area corresponding to no discrimination. We separated males and looked again at mild vs. severe CAD. The areas were still not significantly different from 0.50. Thus, it appears that subjects with mild and with severe CAD share the same distributions for the indexes we examined. The null hypothesis was not contradicted by the data.

ROC Curves

Examples of curves are shown in Figures 1 and 2. Figure 1 (left panel) compares apo B and LDL cholesterol. The areas are significantly different, although the absolute difference is small (0.058). The apo B curve is superior in general, but at the decision levels (cutoff values) corresponding to 60% sensitivity, the two tests have identical specificity (false-positive rates). At that decision level, the two tests seem to perform the same. This illustrates the limitation of displaying test performance at one decision level, such as in a 2 × 2 table, in contrast to the ROC curve, which displays performance at all decision levels.

Figure 1 (center panel), which compares the ratio of apo A-I to apo B with the ratio of HDL to total cholesterol, illustrates the additional discrimination provided by the apolipoprotein ratio over the lipid indexes. Although the ratio of apo A-I to apo B is significantly more accurate, neither achieves high true-positive rates (high sensitivity) at the same time that it has low false-positive rates (high specificity).

Figure 1 (right panel) plots model 2 (which includes apolipoprotein indexes) and model 1 (which includes only lipid indexes). Model 2 has a significantly larger area than does model 1, again reflecting the increased discriminating power when apolipoprotein data are included.

Figure 2 shows the lack of separation between subjects with severe CAD and those with mild CAD by the ratio of apo A-I to apo B. This curve varies back and forth along the 45° line, which corresponds to an area of 0.5, indicating no discrimination.

Discussion

ROC analysis can graphically represent pure, inherent diagnostic accuracy. It provides the raw data needed both to make the first-level assessment of test performance and to incorporate the test into the process of decision analysis, whereby an optimal overall strategy for patient care can be derived. One can readily assess the ability of tests (or variables such as age or history) to discriminate between no CAD and any CAD, between mild and severe CAD, or between none plus mild and moderate plus severe disease. One could also explore the accuracy of each test in selected groups of patients. ROC curves could have been generated for any index by using any group that had an adequate number of subjects.

ROC analysis of these data indicates that all of these indexes (including the regression models that combine indexes) have low to moderate accuracy in identifying clinically important degrees of CAD in patients referred for angiography. Accuracy, as reflected in the areas under the ROC curves, was generally higher for apolipoproteins than for lipids. Note the values for apo B and the ratio of apo A-I to apo B in Table 1. The ROC curves for all the parameters show that, even in the best cases, high sensitivity is always accompanied by low specific-
Likewise, high specificity (a low false-positive rate) is always accompanied by low sensitivity. This should come as no surprise because single measurements of lipids would not adequately summarize the lifetime exposure that can contribute to overt CAD.

When compared with conventional lipid measurements, apolipoprotein determinations seem to yield better ROC curves and larger areas under the curves. The differences are small, however, and diagnostic accuracy is moderate for all the indexes we examined. ROC curve analysis provides a convenient and elegantly simple qualitative (visual) and quantitative representation of pure diagnostic accuracy.

References