

# Comparison of Exercise Test Scores and Physician Estimation in Determining Disease Probability

Michael Lipinski, BS; Dat Do, MD; Victor Froelicher, MD; Lars Osterberg, MD; Barry Franklin, PhD; Jeff West, MD; Eddie Atwood, MD

**Background:** The recent American College of Cardiology/American Heart Association exercise testing guidelines provided equations to calculate treadmill scores and recommended their use to improve the predictive accuracy of the standard exercise test. However, if physicians can estimate the probability of coronary artery disease as well as the scores can, there would be no reason to add this complexity to test interpretation. To compare the exercise test scores with physician's estimation of disease probability, we used clinical, exercise test, and coronary angiographic data to compute the recommended scores and print patient summaries and treadmill reports.

**Objective:** To determine whether exercise test scores can be as effective as expert cardiologists in diagnosing coronary disease.

**Methods:** Five hundred ninety-nine consecutive male patients without previous myocardial infarction with a mean  $\pm$  SD age of  $59 \pm 11$  years were considered for this analysis. With angiographic disease defined as any coronary lumen occlusion of 50% or more, 58% had disease. The clinical/treadmill test reports were sent to expert cardiologists and to 2 other groups, including randomly selected cardiologists and internists, who classified the patients as having high, low, or intermediate probability of

disease and estimated a numerical probability from 0% to 100%.

**Results:** Forty-five expert cardiologists returned estimates on 336 patients, 37 randomly chosen practicing cardiologists returned estimates on 129 patients, 29 randomly chosen practicing internists returned estimates on 106 patients, 13 academic cardiologists returned estimates on 102 patients, and 27 academic internists returned estimates on 174 patients. When probability estimates were compared, the scores were superior to all physician groups (0.76 area under the receiver operating characteristic curve to 0.70 for experts [ $P = .046$ ], 0.73 to 0.58 for cardiologists [ $P = .003$ ], and 0.76 to 0.61 for internists [ $P = .006$ ]). Using a probability cut point of greater than 70% for abnormal, predictive accuracy was 69% for scores compared with 64% for experts, 63% to 62% for cardiologists, and 70% to 57% for internists.

**Conclusion:** Although most similar to the disease estimates of the presence of clinically significant angiographic coronary artery disease provided by the expert cardiologists, the scores outperformed the nonexpert physicians.

*Arch Intern Med.* 2001;161:2239-2244

From the Stanford University Cardiology Department at Palo Alto Veterans Affairs Health Care Center, Palo Alto, Calif (Mr Lipinski and Drs Do, Froelicher, Osterberg, West, and Atwood); the Cardiac Rehabilitation and Exercise Laboratories, William Beaumont Hospital, Royal Oak, Mich (Dr Franklin); and the Department of Physiology, Wayne State University, School of Medicine, Detroit, Mich (Dr Franklin).

**C**LINICAL and exercise test scores can enhance the decision-making process regarding whether a patient with symptoms possibly due to coronary artery disease should undergo coronary angiography.<sup>1,2</sup> The scores help make it less likely that individuals without coronary artery disease will be referred for unnecessary cardiac catheterization. By providing a second opinion, scores also help nonspecialists make more appropriate decisions regarding referral to a specialist. Thus, exercise test scores can provide a means to see that expensive technology is properly used and that access to specialized therapy is ensured.

Besides providing increased predictive accuracy, scores eliminate physician bias and lessen the variability of decision mak-

ing.<sup>3,4</sup> Physicians do not always follow a totally rational decision-making process, but often make clinical decisions based on personal experience and heuristics.<sup>5</sup> By eliminating the intuitive aspect of decision making, scores can provide an unbiased evaluation. Although the value of exercise testing scores to help physicians has been documented,<sup>6</sup> many physicians remain skeptical of the accuracy of scores and prefer to rely on the results of more expensive tests in making their decisions. This skepticism remains despite data demonstrating that scores improve the diagnostic characteristics of exercise tests and predict the presence of coronary disease as well as or better than echocardiographic or nuclear tests.<sup>7</sup> To resolve this skepticism, we performed an analysis to compare the diagnostic accuracy of exercise scores with that of cardiologists and generalists.

## MATERIALS AND METHODS

Patients were selected from a database of the last 2000 consecutive male patients who underwent clinical evaluation, exercise testing, and coronary angiography at the Long Beach and Palo Alto Veteran Affairs medical centers in Long Beach and Palo Alto, Calif, respectively. Patients with previous cardiac surgery or interventions, valvular heart disease, left bundle branch block, more than 1 mm depression, or Wolff-Parkinson-White syndrome on their resting electrocardiogram were excluded from the study. Previous cardiac surgery was the predominant reason for exclusion of patients. We then selected all patients who were referred to evaluate chest pain possibly due to coronary disease and who had complete data and coronary angiography within 4 months of the exercise treadmill test. As is the case for clinical observational studies such as this, there was no attempt to remove workup bias. To avoid falsely increasing the accuracy of the exercise treadmill test, we excluded patients with a previous myocardial infarction by history or diagnostic Q wave, leaving a target population of 599 patients.

Physicians who also examined the patients recorded with the use of computerized forms a thorough clinical history, including medications and risk factors, prospectively at the time of exercise treadmill testing.<sup>8,9</sup>

### EXERCISE TESTING

Patients underwent symptom-limited treadmill testing with the US Air Force School of Aerospace Medicine<sup>10</sup> or an individualized ramp treadmill protocol.<sup>11</sup> Before ramp testing, the patients were given a questionnaire to estimate the patient's exercise capacity before the test. This allowed most patients to reach maximal exercise at approximately 10 minutes.<sup>12</sup> Visual ST-segment depression was measured at the J junction and corrected for pre-exercise ST-segment depression while standing; ST slope was measured during the following 60 milliseconds and classified as up-sloping, horizontal, or down-sloping. Slope was coded as 1 for horizontal, 2 for down-sloping, and 0 for normal slope (ie, up-sloping or ST-segment depression of less than 0.5 mm). The ST response considered was the most horizontal or down-sloping ST-segment depression in any lead, except aVR during exercise or recovery. An abnormal response was defined as 1 mm or more of horizontal or down-sloping ST-segment depression.

No test result was classified as indeterminate,<sup>13</sup> medications were not withheld, and a maximal heart rate target was not used as an end point.<sup>14</sup> The exercise tests were performed, analyzed, and reported per standard protocol with a computerized database (EXTRA; Mosby Publishers, Chicago, Ill).<sup>15</sup> Decisions for cardiac catheterization were consistent with clinical practice.

### CORONARY ANGIOGRAPHY

Coronary artery narrowing was visually estimated and expressed as percentage of lumen diameter stenosis. Patients with 50% or greater narrowing in 1 or more of the following were considered to have significant angiographic coronary

artery disease: the left anterior descending, left circumflex, or right coronary arteries or their major branches or the left main coronary artery. The 50% criterion was chosen to be consistent with the cooperative trialist's choice.<sup>16</sup>

### PATIENT DATA SHEET

Reports of the patient information and treadmill test were then generated from the database. The results of the coronary angiography were excluded from the data sheet to blind the physician interpreter. The patient data sheet provided the information traditionally used by physicians to assess whether a patient presenting with possible coronary artery disease should undergo coronary angiography.

The studies were randomly divided into 78 groups of 12 studies. Each reviewer was sent the data sheets, a return envelope, and a cover letter, which explained the goals of the experiment and guidelines on assigning a patient to the high, intermediate, or low probability group for any coronary disease. We selected 110 cardiologists who were considered experts on the basis of their authorship of exercise testing or angiographic studies. The experts were sent 12 studies each. A 40% response rate resulted in a total of 336 studies filled out by 45 expert cardiologists.

A similar approach was taken with random cardiologists. The random cardiologists were nonacademic practicing cardiologists selected at random from a current membership directory of the American College of Cardiology. The cardiologists were selected as random cardiologists if they were not associated with a university or hospital and were not fellows in training, to distinguish them from the expert cardiologists. To increase the rate of participation in the study, only 6 data sheets of the group of 12 were sent to each random cardiologist. Approximately 400 random cardiologists were sent a packet of studies; 37 cardiologists responded, for a return rate of approximately 10% with 129 studies returned.

A group of random internists were also included for comparison. They were nonacademic practicing internists selected from the 1997 and 1998 official American Board of Medical Specialists directory of board-certified medical specialists. Those associated with a university or hospital were excluded. The randomly selected internists were then sent the same group of studies that were sent to the random cardiologists. Approximately 400 random internists were sent a packet of studies; 29 internists responded, for a return rate of 8% with 106 studies returned.

The final 2 groups of physicians were the local academic cardiologists and the local academic internists. We recruited colleagues at the Palo Alto Veterans Affairs Health Care Center, Palo Alto, Calif, and William Beaumont Hospital, Royal Oak, Mich, to participate and asked them to complete 12 studies each. Thirteen cardiologists returned data sheets on 102 patients, and 27 internists returned data sheets on 174 patients. The response rates were 40% and 75%, respectively.

### PHYSICIAN PARTICIPATION INSTRUCTIONS

Physicians were asked to classify the patient as having a high probability, intermediate probability, or low probability of

having clinically significant coronary disease. The physicians were requested to make this evaluation based on the following criteria for stratification: low probability: patient is reassured that symptoms are most likely not due to coronary disease; intermediate probability: other tests, even possibly angiography, indicated to clarify diagnosis and antianginal medications tried; and high probability: antianginal treatment indicated and angiography may be required if severe disease is likely and an intervention is clinically warranted. Physicians were also requested to provide a numerical percentage as their estimate of the probability that the patient had clinically significant coronary disease.

## SCORES

The clinical and exercise test data were input into the following equation to generate 3 probability estimates:

$$\text{Probability} = 1 / [1 + e^{-(a + bx + cy \dots)}],$$

where  $e$  is the natural log,  $a$  is the intercept,  $b$  and  $c$  are  $\beta$  coefficients, and  $x$  and  $y$  are variable values.

The appropriate coefficients and variables were from the 3 equations included in the American College of Cardiology/American Heart Association exercise testing guidelines.<sup>1</sup> The variables included age, symptoms, risk factors, and exercise test responses, which were derived from the Veterans Affairs (VA) Score, the Simplified VA Score, the Detrano Score, or the Morise Score.

### VA Score

The pre-exercise test equation,<sup>17</sup> including the chosen variables and their coefficients and the constant, is as follows:

$$\begin{aligned} & -2.1 + (0.03 \times \text{Age}) - (0.4 \times \text{Symptoms}) \\ & + (0.8 \times \text{Diabetes}) + (0.4 \times \text{Hypercholesterolemia}) \\ & + (0.01 \times \text{Pack-years}) \\ & + (0.7 \times \text{Resting ST Depression in Millimeters}). \end{aligned}$$

The postexercise test equation, including the chosen variables, their coefficients, and the constant, is as follows:

$$\begin{aligned} & -1.2 + (3.3 \times \text{Pretest}) \\ & + (0.5 \times \text{Exercise ST Depression in Millimeters}) \\ & + (0.6 \times \text{ST Slope}) \\ & - (0.16 \times \text{Metabolic Equivalents}) \\ & - (0.5 \times \text{Exercise Angina}), \end{aligned}$$

where pretest is a number between 0 and 1 generated by the pretest equation.

### Simplified VA Score

The simplified VA score<sup>18</sup> is as follows:

$$\begin{aligned} & (6 \times \text{Maximal Heart Rate}) + (5 \times \text{ST Depression Code}) \\ & + (4 \times \text{Age Code}) + (\text{Angina Pectoris Code}) \\ & + (\text{Hypercholesterolemia}) + (\text{Diabetes}) \\ & + (\text{Treadmill Angina Index}). \end{aligned}$$

### Detrano Score

Detrano et al<sup>19</sup> included 3549 patients from 8 institutions in the United States and Europe who underwent exercise testing and angiography between 1978 and 1989. Disease was defined as greater than 50% diameter narrowing in at least 1 major coronary arterial branch, and the prevalence

of disease according to this criterion was 64%. The selected Detrano equation components are as follows:

$$\begin{aligned} & 1.9 + (0.025 \times \text{Age}) - (0.6 \times \text{Sex}) - (0.1 \times \text{Symptoms}) \\ & - (0.05 \times \text{Metabolic Equivalents}) \\ & - (0.02 \times \text{Maximal Heart Rate}) \\ & + (0.36 \times \text{Exercise-Induced Angina}) \\ & + (0.6 \times \text{ST Depression in Millimeters}). \end{aligned}$$

Sex was coded as 1 for female and -1 for male. Symptoms were classified into the 4 categories of typical, atypical, nonanginal pain, and no pain and coded with the values 1, 2, 3, and 4, respectively. Exercise angina was coded as 1 for presence and -1 for absence.

### Morise Score

Morise et al<sup>20</sup> studied 915 consecutive patients without a history of previous myocardial infarction or coronary artery bypass surgery who were referred to the exercise laboratory at West Virginia University Medical Center, Morgantown, between June 1981 and December 1994 for evaluation of coronary disease. All patients had coronary angiography within 3 months of the exercise test. The patients were classified as having disease if there was at least a 50% lumen diameter narrowing in 1 or more vessels, and using this criterion, the prevalence of disease in their population was 41%. Morise et al generated both pre-exercise and postexercise logistic regression equations. The Morise pre-exercise test intercept and variables are as follows:

$$\begin{aligned} & -3.6 + (0.08 \times \text{Age}) - (1.3 \times \text{Sex}) + (0.6 \times \text{Symptoms}) \\ & + (0.7 \times \text{Diabetes}) + (0.3 \times \text{Smoking}) \\ & - (1.5 \times \text{Body Surface Area}) + (0.50 \times \text{Estrogen}) \\ & + (0.3 \times \text{Number of Risk Factors}) \\ & - (0.40 \times \text{Resting Electrocardiogram}). \end{aligned}$$

Sex was coded as 1 for female and 0 for male. Symptoms were classified into the 4 categories of typical, atypical, nonanginal pain, and no pain and coded with the values 4, 3, 2, and 1, respectively. Diabetes was coded as 1 if present and 0 if absent. Smoking was coded as 2 for current smoking, 1 for any previous smoking, and 0 for never smoked. Estrogen was coded as 0 for men and, for women, 1 for estrogen-negative (postmenopausal and no estrogen) and -1 for estrogen-positive (premenopausal or taking estrogen). Risk factors included history of hypertension, hypercholesterolemia, and obesity (body mass index [calculated as weight in kilograms divided by the square of height in meters],  $\geq 27$ ). Resting electrocardiogram was coded as 0 if normal and 1 if there were QRS or ST-T wave abnormalities.

The Morise posttest equation is as follows:

$$\begin{aligned} & -0.12 + (4.5 \times \text{Pretest}) \\ & + (0.37 \times \text{ST Depression in Millimeters}) \\ & + (1.0 \times \text{ST Slope}) \\ & - (0.4 \times \text{Negative ST}) \\ & - (0.016 \times \text{Maximal Heart Rate}). \end{aligned}$$

Pretest is the pretest probability (0 to 1) derived from the pretest equation. ST depression in millimeters was coded as 0 for women. ST slope was coded as 1 for down-sloping

Continued on next page

and 0 for up-sloping or horizontal. Negative ST was coded as 1 if ST depression was less than 1 mm of depression horizontal or down-sloping or if ST depression was less than 1.5 mm of up-sloping.

#### Consensus of Scores

Our group previously validated a means to make predictive equations more portable and self-calibrating by requiring a consensus for patient classification as to risk of coronary disease.<sup>18-20</sup> These studies used 2 thresholds of each of the computer-generated probability scores to separate the population into 3 groups: low probability (prevalence of coronary disease, <5%), intermediate probability (prevalence of coronary disease, 5%-70%), and high probability (prevalence of coronary disease, >70%). Patients were classified as having low or high probability if at least 2 of the 3 equations agreed, ie, there was a consensus by majority. This approach avoids difficulties due to differences in variable collection, test method, missing data, and disease prevalence.

#### DATA ANALYSIS

The numerical probability must have a cut point to separate normal and abnormal results to calculate test diagnostic characteristics. A 70% probability was established as the cut point for the physician estimates, with predictions greater than or equal to 70% indicating that the patient has coronary disease and predictions less than 70% indicating that the patient does not have coronary disease.

The probability predictions for the 5 groups of physicians were entered into the database and compared against the angiographic results of the patients. Using the 70% cut point, the diagnostic accuracy of predicting angiographic coronary disease was established for each of the 5 groups of physicians. The predictive accuracy of a group of physicians is determined by adding the number of true-positives results and true-negative results and then dividing the sum by the total number of patients. A true-positive result in the probability percentage section results from the physician giving a probability greater than or equal to 70% for a patient with coronary disease. A true-negative result is when the physician gives a probability less than 70% for a patient without coronary disease. A false-positive result is when the physician gives a probability greater than or equal to 70% for a patient without coronary disease. A false-negative result is when the physician gives a probability less than 70% for a patient with coronary disease. The same was done with the average of the probability estimates generated by the 3 equations. The probabilities computed from the scores and their averages were also used to construct receiver operating characteristic (ROC) curves for comparison of their diagnostic (discriminatory) characteristics. *P* values were calculated from the standard errors generated for each area under the ROC curve.

**Table 1. Clinical Characteristics of the 599 Study Patients**

Variable	Results
Patient age, mean ± SD (range), y	59 ± 10.6 (29-92)
Body mass index, mean ± SD, kg/m <sup>2</sup>	28.0 ± 4.6
Hypercholesterolemia, %	37
Diabetes, %	18
Currently smoking, %	32
Typical angina pectoris, %	32
Atypical angina pectoris, %	56
Maximum heart rate, mean ± SD, beats/min	129 ± 24
Maximum systolic blood pressure, mean ± SD, mm Hg	168 ± 29
Metabolic equivalents, mean ± SD	7.5 ± 3.8
Abnormal ST depression response to the exercise test, %	35
Angiographic coronary disease (>50), %	58

## RESULTS

**Table 1** describes our patient population (N=599). No significant differences were found in the 5 different physician data sheet samples.

### PREDICTIVE ACCURACY

**Table 2** shows the predictive accuracy using the 70% probability cut point for the 5 groups of physicians compared with the average of the 3 scores in the same patients reviewed by each physician group.

Consensus provided a significantly higher predictive accuracy than did the physicians (69% vs 63%, *P* < .01). These data, however, do not allow comparison among the different groups, because each group of physicians provided probability percentages for different patients.

### PROBABILITY CLASSIFICATION

The next area of concern is the determination of the probability stratification of each patient. Per the definition provided to the reviewers, this demonstrates the accuracy with which a physician determines whether a patient should undergo further testing. The categories in which a patient could have been placed are high probability, intermediate probability, and low probability. Because a patient classified in the intermediate-probability group will undergo additional diagnostic tests to determine the presence of coronary disease, it is difficult to draw any conclusions from the data for patients in this category. However, high-probability and low-probability categories provide critical data. Patients who are in the low-probability group probably do not have cardiac catheterization or restriction of their activities. If the patient has coronary disease and was in the low-probability group, an incorrect assessment of the likelihood of disease may result in a cardiac event that could have been avoided. Also, if a physician considers a patient without coronary disease to have high probability of disease, he/she may undergo needless costly procedures that are not without risk. **Table 3** provides a comparison among the 5 groups of physicians and scores in the number of patients with coronary disease who were con-

**Table 2. The Predictive Accuracy of the Physician Groups Compared With the Scores**

	Expert Cardiologists (n = 45)	Random Cardiologists (n = 37)	Random Internists (n = 29)	Academic Cardiologists (n = 13)	Academic Internists (n = 27)
Predictive accuracy, % (No.)	64 (215/336)	62 (80/129)	57 (60/106)	67 (68/102)	66 (115/174)
Predictive accuracy of the scores in the same patients, %	69	63	70	71	71

**Table 3. Comparison of the 5 Groups of Physicians and Scores for the Classification of Patients With Coronary Artery Disease (CAD) as Having Low Probability of Having Coronary Disease**

Physician Group	Patients With CAD Considered Low Probability, % (No.)	Scores, % (No.)
Expert cardiologists	11 (23/204)	6 (12/204)
Random cardiologists	15 (12/79)	8 (6/79)
Random internists	12 (7/57)	11 (6/57)
Academic cardiologists	5 (3/61)	8 (5/61)
Academic internists	12 (13/113)	10 (11/113)

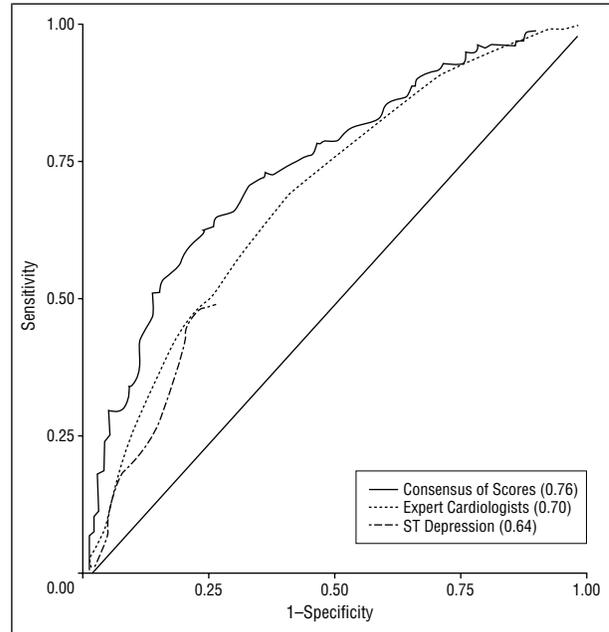
sidered to have low probability of disease divided by the total number of patients with coronary disease. The results show that scores performed better than did the physicians, with a sensitivity of 90% or more. A similar table was constructed for patients who did not have coronary disease but who were considered to have high probability of developing the disease, and the scores outperformed the physicians, with a specificity of approximately 90%. In addition, the scores performed better than did all 5 groups of physicians by putting a higher percentage of patients with coronary disease into the high-probability category.

### ROC CURVES

The ROC curves were plotted for the physician groups and the amount of horizontal ST depression, and the average of the 3 scores was used for consensus (**Figure**). The area under the curve of the scores was significantly greater than that for the other methods for diagnosis. This can also be seen in **Table 4**.

### COMMENT

Although scores based on exercise testing data have been advocated for years, only 3 previous studies have compared them with physician estimates of disease. Detrano et al<sup>21</sup> performed one of the first such studies. They derived a score for estimating probabilities of significant and severe coronary disease, then they validated and compared it with the assessments of cardiologists. The score performed at least as well as did the physicians when the latter knew the identity of the patients. The clinicians were more accurate when they did not know the identity of the subjects but worked from tabulated objective data. Detrano and colleagues concluded that the application of scores or consultation with cardiologists not directly involved with patient management might result in more rational assessments and decision making.



Receiver operating characteristic curves for the expert cardiologists estimates, the amount of horizontal ST depression, and the average of the 3 scores used for consensus. Straight line represents no discriminating value.

Hlatky et al<sup>22</sup> validated 2 scores by comparing their diagnostic accuracy with that of cardiologists. Ninety-one cardiologists participated in the study; each evaluated the clinical summaries of 8 randomly selected patients who had complete evaluations, including coronary angiography. The scores outperformed these cardiologists. A third study<sup>23</sup> considered scores for prognosis (rather than diagnosis), with clinical summaries for 100 patients sent to 5 senior cardiologists at 1 center. Again, the scores outperformed these cardiologists. Our study was larger and included different groups of physicians, validating the results of these earlier studies that showed that scores can predict angiographic results as well as physicians can.

We are not advocating that these scores replace physician judgment. The scores can be thought of as providing a readily available second opinion or consultation. The scores can also reassure nonspecialists that their decision to either manage the treatment of a patient themselves or refer that patient to a cardiologist is consistent with medical knowledge. Certainly, if a patient's symptoms reoccur or are not manageable, reassessment by the physician will lead to referral. In addition, by using the scores to objectively stratify probability, a management strategy becomes available that is more practical than the traditional interpretation of abnormal or normal exercise test results. Because the ability of scores based on any test to detect coro-

**Table 4. Comparison of Area Under the Receiver Operating Characteristic Curves for Physician Groups With Treadmill Scores**

	Random Experts	Random Cardiologists	Random Internists	Academic Cardiologists	Academic Internists
Area under the curve					
ST depression	0.64	0.67	0.57	0.71	0.71
Physician	0.70	0.58	0.61	0.73	0.71
Score	0.76	0.73	0.76	0.71	0.80
P value	.046	.003	.006	.34	.02

nary disease remains imperfect, physicians will always be paramount in the decision process. However, decision making has been greatly improved in a wide range of endeavors through the use of scores.<sup>2</sup>

Although numerous studies have shown that scores enhance the discriminatory power of the standard, inexpensive, and widely available exercise test, scores have not been widely applied. There has always been a tendency to replace the old with the new in medicine despite evidence of only marginal improvement.<sup>24</sup> Although imaging has certain advantages in patients with certain electrocardiogram abnormalities and for localization of ischemia, it is better applied in patients with chest pain when these advantages are the issues at hand. Certainly, most initial testing is best accomplished with the less expensive technology. We hope now that we have validated scores by comparing them with the interpretive abilities of a large number of expert cardiologists, physicians and other health care practitioners have the evidence to apply the strategy we have suggested. Although not designed as a cost-efficacy analysis, simple mathematical modeling allows demonstration of considerable savings. Also, our approach follows current health care mandates to empower nonspecialists yet ensure access to specialty care.

Legal liability and related issues are possible reasons that nonspecialists may refer patients to cardiologists, who may then recommend catheterization for those patients who have a low to intermediate probability of coronary disease. The decisions generated by the scores are not influenced by such liability or recent personal experiences with a rare bad outcome. Since the scores are objective and cited in the guidelines, they can shield physicians from such concerns.

Limitations to our study include its retrospective design, workup bias, lack of outcomes data, lack of women, and no formal cost-efficacy component. Another consideration is that the physicians did not interview or examine the patients themselves. However, we hope that our findings will encourage investigators to perform studies that overcome these limitations. At first, the low response rates seem to be a limitation, but if anything, they favor the physician groups, since those physicians with experience and confidence in their ability to diagnose coronary disease would be more likely to respond. Although the consensus of scores we applied could be considered complicated, a single score, such as our simple score,<sup>18</sup> validated in a clinical setting, would be just as effective.

In our study, scores did as well as or better than physicians in estimating the probability of clinically significant angiographic coronary artery disease.

Accepted for publication January 18, 2001.

Presented as an abstract at the Scientific Sessions of the American Heart Association, Atlanta, Ga, November 8, 1999.

Corresponding author and reprints: Victor Froelicher, MD, Cardiology Division (111C), Veterans Affairs Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304 (e-mail: vicmd@aol.com and Web site <http://www.cardiology.org>).

## REFERENCES

- Gibbons RJ, Balady GJ, Beasley JW, et al. ACC/AHA guidelines for exercise testing. *J Am Coll Cardiol.* 1997;30:260-311.
- Swets JA, Dawes RM, Monahan J. Better decisions through science. *Sci Am.* October 2000:82-87.
- Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS, for the Evidence-Based Medicine in Critical Care Group. Understanding articles describing clinical prediction tools. *Crit Care Med.* 1998;26:1603-1612.
- Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. XV: how to use an article about disease probability for differential diagnosis. *JAMA.* 1999;281:1214-1219.
- Hlatky MA, Califf RM, Harrell FE Jr, et al. Clinical judgment and therapeutic decision making. *J Am Coll Cardiol.* 1990;15:1-14.
- Yamada H, Do D, Morise A, Froelicher V. Review of studies utilizing multi-variable analysis of clinical and exercise test data to predict angiographic coronary artery disease. *Prog Cardiovasc Dis.* 1997;39:457-481.
- Froelicher VF, Fearon W, Ferguson C, West J, Hedenreich P. Lessons learned from studies of the standard exercise test? *Chest.* 1999;116:1442-1451.
- Ustin J, Umann T, Froelicher V. Data management: a better approach. *Phys Comput.* 1994;12:30-33.
- Froelicher V, Shiu P. Exercise test interpretation system. *Phys Comput.* 1996;14:40-44.
- Wolthuis R, Froelicher VF, Fischer J, Longo M, Triebwasser J. New practical treadmill protocol for clinical use. *Am J Cardiol.* 1977;39:697-700.
- Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Froelicher VF. A comparison of the ramp versus standard exercise protocols. *J Am Coll Cardiol.* 1991;17:1334-1342.
- Myers J, Do D, Herbert W, Ribisl P, Froelicher VF. A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data. *Am J Cardiol.* 1994;73:591-596.
- Reid M, Lachs M, Feinstein A. Use of methodological standards in diagnostic test research. *JAMA.* 1995;274:645-651.
- Fletcher GF, Froelicher VF, Hartley LH, Haskell WL, Pollock ML. Exercise standards: a statement for health professionals from the American Heart Association [revised in *Circulation.* 1995;91:580-632]. *Circulation.* 1990;82:2286-2321.
- Shue P, Froelicher V. Extra: an expert system for exercise reporting. *J Noninvasive Testing.* 1998;2(4):21-27.
- Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomized trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet.* 1994;344:563-570.
- Do D, West J, Morise A, Froelicher V. A consensus approach to diagnosing coronary artery disease based on clinical and exercise test data. *Chest.* 1997;111:1742-1749.
- Raxwal V, Shetler K, Do D, et al. Simple treadmill score to diagnose coronary disease. *Chest.* 2001;119:1933-1940.
- Detrano R, Bobbio M, Olson H, et al. Computer probability estimates of angiographic coronary artery disease: transportability and comparison with cardiologists' estimates. *Comput Biomed Res.* 1992;25:468-485.
- Morise AP, Detrano R, Bobbio M, Diamond GA. Development and validation of a logistic regression-derived algorithm for estimating the incremental probability of coronary artery disease before and after exercise testing. *J Am Coll Cardiol.* 1992;20:1187-1196.
- Detrano R, Bobbio M, Olson H, et al. Computer probability estimates of angiographic coronary artery disease: transportability and comparison with cardiologists' estimates. *Comput Biomed Res.* 1992;25:468-485.
- Hlatky M, Botvinick E, Brundage B. Diagnostic accuracy of cardiologists compared with probability calculations using Bayes' rule. *Am J Cardiol.* 1982;49:1927-1931.
- Lee KL, Pryor DB, Harrell FE Jr, et al. Predicting outcome in coronary disease: statistical models versus expert clinicians. *Am J Med.* 1986;80:553-560.
- McDonald CJ, Overhage JM, Dexter P, Takesue BY, Dwyer DM. A framework for capturing clinical data sets from computerized sources. *Ann Intern Med.* 1997;127(8 pt 2):675-682.