Comparison of treadmill scores with physician estimates of diagnosis and prognosis in patients with coronary artery disease

Michael Lipinski, BS,[a] Victor Froelicher, MD,[a] Eddie Atwood, MD,[a] Anna Tseitlin, BA,[a] Barry Franklin, PhD,[b] Lars Osterberg, MD,[a] Dat Do, MD,[a] and Jonathan Myers, PhD[a] ***Palo Alto, Calif, and Royal Oak and Detroit, Mich***

Objective Our purpose was to compare exercise test scores and ST measurements with a physician's estimation of the probability of the presence and severity of angiographic disease and the risk of death. The American College of Cardiology/American Heart Association exercise testing guidelines provide equations to calculate treadmill scores and recommend their use to improve the predictive accuracy of the standard exercise test. However, if physicians can estimate the probability of coronary artery disease and prognosis as well as the scores, there is no reason to add this complexity to test interpretation.

Methods A clinical exercise test was performed and an angiographic database was used to print patient summaries and treadmill reports. The clinical/treadmill test reports were sent to expert cardiologists and to 2 other groups, including randomly selected cardiologists and internists. They classified the patients summarized in the reports as having a high, low, or intermediate probability for the presence of any severe angiographic disease and estimated a numerical probability from 0% to 100%. The Social Security Death Index was used to determine survival status of the patients.

Results Twenty-six percent of the patients had severe angiographic disease, and the annual mortality rate for the population was 2%. Forty-five expert cardiologists returned estimates on 473 patients, 37 randomly chosen practicing cardiologists returned estimates on 202 patients, 29 randomly chosen practicing internists returned estimates on 162 patients, 13 academic cardiologists returned estimates on 145 patients, and 27 academic internists returned estimates on 272 patients. When probability estimates for presence and severity of angiographic disease were compared, in general, the treadmill scores were superior to physicians' and ST analysis at predicting severe angiographic disease. When prognosis was estimated, treadmill prognostic scores did as well as expert cardiologists and better than most other physician groups.

Conclusion Estimates of the presence of clinically significant and severe angiographic coronary artery disease provided by scores were superior to physician estimates and ST analysis alone. Estimates of prognosis provided by scores were similar to the estimates made by expert cardiologists and more accurate than the estimates made by most other physician groups. (Am Heart J 2002;143:650-8.)

In 1997, the American College of Cardiology/American Heart Association (ACC/AHA) guidelines first recommended the use of treadmill scores to improve the diagnostic and prognostic characteristics of the exercise test.[1] However, physicians remain uncertain as to which scores provide the greatest diagnostic and prognostic accuracy. Scores can provide physicians with a second opinion, which could decrease the number of patients with coronary artery disease (CAD) who are missed and reduce the number of patients without CAD who undergo costly diagnostic tests. Treadmill scores can also assist physicians in estimating prognosis and in formulating an appropriate course of action for managing patients.

Besides improving diagnostic and prognostic accuracy, scores eliminate physician bias and lessen the variability of decision making.[2,3] Physicians often make clinical decisions on the basis of personal experience and heuristics[4] rather than a rational decision-making process. Still, physicians tend to rely more on the results of expensive tests such as exercise nuclear imaging or echocardiography. Treadmill scores have been shown to perform as well as or better than these tests,[5,6] but

From the [a]Department of Cardiology, Stanford University at Palo Alto Veterans Affairs Health Care Center, Palo Alto, Calif, and the [b]Cardiac Rehabilitation and Exercise Laboratories, William Beaumont Hospital, Royal Oak, and the Department of Physiology, Wayne State University, School of Medicine, Detroit, Mich.
Submitted June 29, 2001; accepted October 10, 2001.
Based on abstracts presented at the Scientific Sessions of the American Heart Association, New Orleans, La, November 2000, and Scientific Sessions of the American College of Cardiology, Orlando, Fla, March 2001.
Reprint requests: Victor Froelicher, MD, Cardiology Division (111C), VA Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304.
E-mail: vicmd@aol.com
4/1/120967
doi:10.1067/mhj.2002.120967

the complex nature of the scores has deterred physicians from incorporating them into their decision-making process. The purpose of this study was to evaluate the relative accuracy of diagnostic and prognostic treadmill scores in comparison with physician estimates, with specific reference to internists and cardiologists. We previously reported on how a consensus of scores performed compared with physician groups,[7] but in this report we also consider ST analysis alone, our simple score,[8] the Duke Treadmill Score, and we add severe CAD and death as end points for prediction.

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. Patients with prior cardiac surgery or interventions, valvular heart disease, left bundle branch block, >1 mm ST-segment depression, or Wolff-Parkinson-White syndrome on their resting 12-lead electrocardiograms were excluded from the study. Previous cardiac surgery was the predominant reason for the exclusion of patients. We then selected all patients with complete data who were evaluated for chest pain that was possibly due to coronary disease and who had coronary angiography within 4 months of the exercise treadmill test. A thorough clinical history, medications, and coronary risk factors were recorded prospectively using computerized forms at the time of exercise treadmill testing.[9]

We generated data sheets with clinical and treadmill results from these patients and sent them to physicians. Completed responses were returned on 686 patients and provided the data for this study. For the diagnosis of any significant angiographic disease, patients with previous myocardial infarction (MI), by history or by diagnostic Q wave, were excluded because their diagnosis was established. A target population of 599 patients remained for estimating diagnosis of significant disease. All 686 patients were considered for evaluation of severe angiographic disease and prognosis.

Exercise testing

Patients underwent symptom-limited treadmill testing.[10-12] Visual ST-segment depression was measured at the J junction; ST slope was measured over the following 60 ms and classified as upsloping, horizontal, or downsloping if there was ≥0.5 mm depression. The ST response considered was the most horizontal or downsloping ST-segment depression in any of the 12 leads except aVR during exercise or recovery. An abnormal response was defined as ≥1 mm of horizontal or downsloping ST-segment depression. However, for the receiver operating characteristic (ROC) curve analysis, ST-segment depression was used as a continuous variable and was defined as any measurement >0.5 mm depression.

No test was classified as indeterminate,[13] medications were not withheld, and a maximum heart rate target was not used as an end point.[14] The exercise tests were performed, analyzed, and reported by use of a computerized database that generates the report and stores the responses of the physicians (EXTRA, Mosby, Chicago, Ill).[15]

Coronary angiography

Decisions for cardiac catheterization were consistent with clinical practice. Coronary artery narrowing was estimated visually and expressed as percent lumen diameter stenosis. Significant angiographic CAD was defined as ≥50% luminal occlusion in one or more of the following: left anterior descending artery, left circumflex artery, right coronary artery, or their major branches, or in the left main coronary artery. Severe angiographic CAD was defined as ≥50% luminal occlusion in 3 of the previously mentioned arteries, 2 vessels when 1 was the proximal left anterior descending, or when there was left main disease. The 50% and severe disease criteria are consistent with the findings of the cooperative trialists.[16]

Patient outcome

To compare the ability of physicians and treadmill scores with estimate prognosis, the Social Security Death Index was used to determine whether the 686 patients were dead or alive. All-cause mortality was used as the end point, and the time from the treadmill test until death was calculated in months for patients who died. We did not have access to the cause of death or the occurrences of coronary events.

Patient data sheet

Patient information and treadmill test reports were 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 with possible CAD 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 that explained the goals of the experiment and guidelines on assigning a patient to high, intermediate, or low probability for coronary disease. We selected 110 "expert" cardiologists (defined on the basis of their authorship of exercise testing/angiographic studies). The experts were sent 12 studies each. A 40% response rate resulted in a total of 473 studies completed by 45 expert cardiologists, of which only 336 were used for any significant disease prediction because of 143 patients having a prior MI or diagnostic Q waves, or both.

A similar approach was taken with "random" cardiologists. The random cardiologists were nonacademic practicing cardiologists selected at random from a current membership directory for the American College of Cardiology. To distinguish them from the experts, cardiologists were selected as random if they were not associated with a university or hospital and were not fellows in training. 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. A total of 37 cardiologists responded for a return rate of approximately 10%, with 202 studies returned.

A group of "random internists" was also included for comparison. They were nonacademic practicing internists selected from the 1997 and 1998 Official Advisory 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 6 studies that were sent to the random cardiologists. Approximately 400 random internists were sent a

Table I. Summary of numeric distribution of the physician groups and the number of patient forms they interpreted and returned

	Expert cardiologists	Random cardiologists	Random internists	Academic cardiologists	Academic internists
No. of mailed reports	110	400	400	28	32
No. who returned reports	45	37	29	13	27
No. of returned patient forms analyzed	473	202	162	145	272
Return rate (%)	40	10	8	46	75

packet of studies. A total of 29 internists responded for a return rate of 8%, with 162 patient evaluations 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 145 patients and 27 internists returned data sheets on 272 patients. The response rates were 40% and 75%, respectively. The distribution and return rates of patient reports are summarized in Table I.

Physician participation instructions

Physicians were asked to classify the patient as high probability, intermediate probability, or low probability of having clinically significant coronary disease and severe coronary disease. We requested that the physicians make this evaluation on the basis of the following criteria for significant disease stratification:

1. Low probability, patient is reassured that symptoms are most likely not due to coronary disease
2. Intermediate probability, other tests indicated to clarify diagnosis, antianginal medications tried
3. High probability, antianginal treatment indicated, angiography may be required if severe disease is likely and an intervention is clinically warranted.

The following served as criteria for severe disease stratification:

1. Low probability, patient is reassured that symptoms are most likely not due to severe coronary disease and, therefore, there is no need for additional procedures to improve prognosis
2. Intermediate probability, other tests indicated to clarify severity
3. High probability, if clinically appropriate, the patient should be informed of increased probability and advised to undergo angiography in consideration for an intervention.

We also requested that physicians provide numeric percentages as their estimate of the probability that the patient had clinically significant and severe coronary disease. It was assumed that a high probability of severe disease was equivalent to an increased likelihood of mortality.

The Duke Treadmill Score

Using the Cox proportional hazard analysis, Mark et al[17,18] originally developed the Duke Treadmill Score (DTS) as a prognostic score. However, they later validated the DTS as a diagnostic score.[19] The DTS uses 3 variables to generate the score: exercise capacity, amount of ST depression, and the angina that occurs during the exercise treadmill test. The DTS is calculated as:

Exercise time – (5 × ST depression) – (4 × Treadmill angina index)

Exercise time is measured in minutes of the Bruce protocol, which can be derived from metabolic equivalent of the tasks; ST depression is measured in millimeters at the J point; and the treadmill angina is coded from 0 to 2. Zero on the treadmill angina code is no angina during the treadmill test, 1 is for nonlimiting angina during the test, and 2 is for termination of the test because of angina. Using the established cut points, patients with a DTS ≤–11 were categorized as high probability; patients with a DTS ≥5 were categorized as low probability; and the patients between 5 and –11 were categorized as intermediate probability. This stratification was used for both prognosis and diagnosis.

Consensus diagnostic score

The clinical and exercise test data were put into equations included in the ACC/AHA exercise testing guidelines to generate 3 probability estimates. Variables included age, symptoms, coronary risk factors, and exercise test responses.[20-22] We averaged the 3 computer-generated probability scores to separate the population into 3 groups: low probability patients had a score of <30%, intermediate probability (30%-69.9%), and high probability ≥70%.

Simple diagnostic score

The simple score was generated at the Palo Alto Veteran Affairs Hospital for use in a male population and validated at the University of West Virginia. The simple score is calculated as shown in Figure 1. Patients were then categorized into the high probability of CAD group if they had a score >60, low probability if the score was <40, or intermediate probability if the score was between 40 and 60. A modification was made to predict severe disease by adding 5 points for history of MI or diagnostic Q waves, or both.

Simple prognostic score

A simple prognostic score was previously developed on the basis of exercise test results and clinical data from the database of 6000 patients.[23] Using the Cox hazard function analysis, we found that the following 4 variables were independently and significantly associated with time to all-cause death: age, history of congestive heart failure (CHF), history of MI or presence of a diagnostic Q wave, and METs. Considering their coefficients, the following prognostic score was calculated as:

Figure 1

Variable	Circle response	Sum
Maximal Heart Rate	Less than 100 bpm = 30	
	100 to 129 bpm = 24	
	130 to 159 bpm =18	
	160 to 189 bpm =12	
	190 to 220 bpm =6	
Exercise ST Depression	1-2mm =15	
	> 2mm =25	
Age	>55 yrs =20	
	40 to 55 yrs = 12	
Angina History	Definite/Typical = 5	
	Probable/atypical =3	
	Non-cardiac pain =1	
Hypercholesterolemia?	Yes=5	
Diabetes?	Yes=5	
Exercise test induced	Occurred =3	
Angina	Reason for stopping =5	
	Total Score:	

Males

Choose only one per group

40=low prob

40-60= intermediate probability

>60=high probability

The simple score for estimating the probability of angiographic CAD. For estimating severe disease, 5 points were added each for history of MI and the presence of diagnostic Q waves.

METs <5 (1 = yes, 0 = no), Age >65 (1 = yes, 0 = no) + History of CHF (1 = yes, 0 = no) + History of MI or Q wave on electrocardiogram (1 = yes, 0 = no)

Patients with a score of 0 were considered to have a low risk of death, patients with a score >1 were considered to have a high risk of death, and patients with a score of 1 were considered to have an intermediate risk of death.

Statistical analysis

The physician estimates and the treadmill scores on the percentage probability of both significant and severe disease were compared with the patient's angiographic results. ROC curves were used to analyze the predictive accuracy of the scores and the physician estimates by comparing the area under the ROC curve for the physician estimates and the treadmill scores with the same sample of patients. Because physician groups evaluated different patients, comparisons could only be made between the physicians and the treadmill scores for the same set of patients, not between physician groups. The higher the value for the area under the ROC curve, the better the score or physician group was at discriminating between patients with and patients without CAD.

CIs for the ROC curves were calculated to determine statistical significance between the ST analysis, physicians, and treadmill scores. The formula used for calculating the confidence interval is

$$CI = SE \times Z \text{ score}.$$

Using a Z score to give a *P* value <.05 and the SEs, the CIs for all the area under the ROC curves were calculated to determine whether the differences between the area under the curves (AUCs) were statistically significant.

This same format was also used for the calculation of the predictive accuracy (PA) for physician groups, treadmill scores, and ST analysis. PA was calculated by taking cut points adjusted to match the specificity of 1 mm of ST depression. The cut points used were 70 for consensus, 50 for the simple score for any disease, 75 for severe disease, 1 for the DTS, and 1 mm for ST depression. PA is calculated by adding the number of true positives and true negatives and then dividing by the total number of patients. We then used cross tabulation with the clinical cut points and the angiographic data to calculate the PA for significant and severe disease. The difference in the PA between 2 methods (physician estimates vs ST analysis) is the difference in the number of patients correctly classified out of 100

Table II. Clinical characteristics for the populations of the any significant and severe disease analyses

Patient variables	Any significant disease group (n = 599)	Severe disease group (n = 686)
Age (y)	59 ± 10.6	59 ± 10.4
BMI (kg/m^2)	28.0 ± 4.6	28 ± 4.4
Hypercholesterolemia (%)	37	41
Diabetes (%)	18	19
Currently smoking (%)	32	36
Typical angina pectoris (%)	32	36
Atypical angina pectoris (%)	56	51
Maximum HR (beats/min)	129 ± 24	124 ± 24
Maximum SBP (mm Hg)	168 ± 29	164 ± 30
METs	7.5 ± 3.8	7.0 ± 3.0
Abnormal exercise induced ST depression (%)	35	41
Percent with any significant or severe angiographic coronary disease	58	25

(ie, the more or less percent of patients correctly classified compared with ST analysis alone).

We compared the prognostic accuracy of physicians with the treadmill score by use of Kaplan-Meier survival analysis and Cox hazard function. With all-cause mortality as an end point, we compared the percent annual mortality of the high, low, and intermediate probability groups for the physician groups with the treadmill scores. Patients evaluated by the expert cardiologists were compared with the scores for prognosis because of the larger sample size. In addition, ROC curves were constructed for the scores and the physician's estimates.

Results

Table II describes our total patient population as well as the diagnostic subgroup. Of the entire population of 686 patients, 25% had severe angiographic coronary disease, and in the diagnostic subgroup of 599, 58% had some type of clinically significant disease. No significant differences were found in the 5 different samples sent to the physician groups.

ROC curve comparisons

ROC curves were plotted for the physician group estimates of probability, the amount of horizontal ST-segment depression, the DTS, the consensus of scores, and the simple score for predicting clinically significant and severe coronary disease. Tables III and IV display the results comparing area under the ROC curve with significant and severe disease prediction data. These results show that the simple score and the consensus of scores perform similarly while discriminating significantly better than physician groups and abnormal ST depression for the prediction of significant and severe angiographic CAD. The consensus of scores and the simple score performed as well as the DTS for prediction of severe disease but performed better than the DTS for prediction of any significant disease. The DTS performed as well as or better than the physician groups and performed better than ST analysis for both significant and severe disease prediction.

Predictive accuracy

The disease prevalence did not differ significantly among groups for both significant and severe disease. Therefore, prevalence was not adjusted to calculate the predictive accuracy.[24] The data for predictive accuracy of physician groups, treadmill scores, and ST analysis for significant and severe disease are analyzed in Table V. In the table, the predictive accuracy of the ST-segment analysis was subtracted from the predictive accuracy of each score or physician group estimate. This provided the number of patients out of 100 who were correctly classified compared with the ST-segment analysis. The data reveal that scores performed as well as or better than the physician groups. However, the consensus of scores did poorly for classifying patients with severe disease.

Probability classification

Patients placed at low probability are important to correctly classify because they do not require cardiac catheterization or restriction of their activities. If the patient has coronary disease and is considered low probability, an incorrect assessment of their likelihood of disease may result in a cardiac event that could have been avoided. Table VI provides a comparison among the 5 groups of physicians and the treadmill scores in the number of patients with significant and severe coronary disease who were considered low probability, divided by the total number of patients with either significant or severe coronary disease. The results reveal that treadmill scores performed better than the physicians, missing fewer patients in the low probabil-

Table III. Comparison of area under the ROC curve for physician group estimates with treadmill scores and ST analysis for any significant disease (95% CI)

	Expert cardiologists	Random cardiologists	Random internists	Academic cardiologists	Academic internists
Physician	0.70 (0.65-0.75)	0.58 (0.50-0.66)	0.61 (0.52-0.70)	0.73 (0.65-0.81)	0.71 (0.65-0.77)
ST segment	0.64 (0.59-0.69)	0.67 (0.59-0.75)	0.57 (0.48-0.66)	0.71 (0.63-0.79)	0.71 (0.65-0.77)
Duke score	0.69 (0.64-0.74)	0.65 (0.57-0.73)	0.63 (0.54-0.72)	0.72 (0.64-0.80)	0.72 (0.66-0.78)
Consensus	0.76 (0.71-0.81)	0.73 (0.65-0.81)	0.76 (0.67-0.85)	0.71 (0.63-0.79)	0.80 (0.74-0.86)
Simple score	0.73 (0.68-0.78)	0.71 (0.63-0.79)	0.72 (0.63-0.81)	0.78 (0.70-0.86)	0.78 (0.72-0.84)

The greater the AUC, the better the score or the physician group was at discriminating between patients with and without disease. Comparisons should only be made within each column.

Table IV. Comparison of area under the ROC curve for physician group estimates with treadmill scores and ST analysis for severe disease (95% CI)

	Expert cardiologists	Random cardiologists	Random internists	Academic cardiologists	Academic internists
Physician	0.67 (0.62-0.72)	0.60 (0.54-0.66)	0.69 (0.61-0.77)	0.71 (0.63-0.79)	0.66 (0.60-0.72)
ST segment	0.64 (0.59-0.69)	0.70 (0.64-0.76)	0.71 (0.63-0.79)	0.67 (0.59-0.75)	0.68 (0.62-0.74)
Duke score	0.71 (0.66-0.76)	0.71 (0.65-0.77)	0.77 (0.69-0.85)	0.76 (0.68-0.84)	0.70 (0.64-0.76)
Consensus	0.72 (0.67-0.77)	0.76 (0.70-0.82)	0.76 (0.68-0.84)	0.78 (0.70-0.86)	0.73 (0.67-0.79)
Simple score	0.73 (0.68-0.78)	0.76 (0.70-0.82)	0.77 (0.69-0.85)	0.75 (0.67-0.83)	0.73 (0.67-0.79)

The greater the AUC, the better the score or the physician group was at discriminating between patients with and without disease. Comparisons should only be made within each column.

Table V. The number of patients out of 100 that can be correctly classified relative to using ST-segment analysis alone for both any significant and severe disease (any/severe) by physician estimates or scores

	Expert cardiologists	Random cardiologists	Random internists	Academic cardiologists	Academic internists
Physicians	4 more (any)/ 8 more (severe)	3 more (any)/ 2 more (severe)	2 less (any)/ 3 more (severe)	8 more (any)/ 4 more (severe)	7 more (any)/ 1 more (severe)
Duke score	3 more (any)/ 1 more (severe)	0 more (any)/ 2 less (severe)	1 less (any)/ 4 more (severe)	6 more (any)/ 6 more (severe)	9 more (any)/ 0 more (severe)
Consensus	10 more (any)/ 8 less (severe)	4 more (any)/ 3 less (severe)	11 more (any)/ 2 less (severe)	12 more (any)/ 6 less (severe)	12 more (any)/ 2 less (severe)
Simple score	8 more (any)/ 11 more (severe)	3 more (any)/ 5 more (severe)	6 more (any)/ 9 more (severe)	14 more (any)/ 4 more (severe)	13 more (any)/ 6 more (severe)

The number of patients more or less (out of 100) classified correctly compared with ST analysis alone is selected on the basis of using predictive accuracy calculations. Predictive accuracy is the percent of true calls (TP + TN) out of those tested. Comparisons should only be made within each column between any versus any and severe versus severe. The greater the number, the more patients correctly classified compared with ST analysis alone. *any*, Any CAD analysis; *severe*, severe CAD analysis.

ity classification compared with ST analysis alone. The exception is the DTS, which tended to miss more patients with any significant disease.

Prognostic results

The annual all-cause mortality rate for the population was 2%. ROC curves were plotted and the AUC was calculated for the prognostic scores and the physician group estimates of severe disease. The AUC of the scores and the expert cardiologists were not significantly different. All of the other physician groups (probability stratification not tabulated) had AUCs consistent with no discrimination (0.46 for the academic cardiologists, 0.53 for the academic internists, and 0.45 for the random cardiologists, except for the small group of random internists with an AUC of 0.66 [0.48-0.77]). Because of these poor results and small sample size, further analysis was concentrated on the expert cardiologists.

The patients classified into high, low, and intermediate probability groups by the expert cardiologists and the prognostic scores along with all-cause mortality

Table VI. Comparison of percent of patients with any significant and severe coronary disease (any/severe) categorized as low probability by physician groups and treadmill scores

	Expert cardiologists (n = 473)	Random cardiologists (n = 202)	Random internists (n = 162)	Academic cardiologists (n = 145)	Academic internists (n = 272)
Percent of patients with CAD missed by physicians by being called low probability	11% (any)/ 28% (severe)	15% (any)/ 22% (severe)	12% (any)/ 15% (severe)	5% (any)/ 29% (severe)	12% (any)/ 26% (severe)
Percent of patients with CAD missed by Duke score	21% (any)/ 11% (severe)	22% (any)/ 8% (severe)	26% (any)/ 9% (severe)	26% (any)/ 7% (severe)	26% (any)/ 11% (severe)
Percent of patients with CAD missed by consensus	6% (any)/ 2% (severe)	8% (any)/ 2% (severe)	11% (any)/no patients with severe CAD	8% (any)/no patients with severe CAD	10% (any)/ 1% (severe)
Percent of patients with CAD missed by Simple score	9% (any)/ 6% (severe)	14% (any)/ 3% (severe)	16% (any)/ 3% (severe)	7% (any)/no patients with severe CAD	7% (any)/ 3% (severe)

Comparisons should only be made within each column between any vs any and severe vs severe. *n*, Number of patients whose reports were returned for analysis; *any*, any CAD; *severe*, severe CAD.

Table VII. Comparison of annual all-cause mortality data for low, intermediate, and high probability groups for expert cardiologists, DTS, and the simple prognostic score

	Expert cardiologists	DTS	Simple prognostic score
Low probability of death	1% (n = 231)	1% (n = 158)	1% (n = 159)
Intermediate probability of death	3% (n = 151)	2% (n = 277)	2% (n = 158)
High probability of death	5% (n = 91)	6% (n = 23)	4% (n = 141)
AUC	0.65 (0.62-0.72)	0.64 (0.61-0.71)	0.66 (0.63-0.72)

All of the other physician groups (probability stratification not tabulated) had AUC of 0.50 (0.41-0.59) except for the small group of random noncardiologists with an AUC of 0.66 (0.48-0.77). *n*, Number of patients classified by probability stratification out of the total of 473 patient forms returned by the expert cardiologists.

data were analyzed with Kaplan-Meier survival analysis. Table VII compares the mortality rates of the different probability groups and the ROC analysis for the prognostic scores and expert cardiologists.

Discussion

The ability of scores to perform as well or better than other stress tests[5,25,26] brings into question whether reliance on these tests is necessary if a treadmill score can more accurately diagnose patients. The ACC/AHA guidelines recommend the exercise treadmill test as the first diagnostic procedure to be performed on patients with suspected CAD.[1] The inclusion of treadmill scores into common medical practice could decrease the number of patients with CAD who are missed and reduce the number of patients without disease who are referred for more costly diagnostic tests.

Scores should not replace physician judgment, but they should be used as an aid to the physician. By providing a second opinion, scores can help physicians in managing patients and decrease the number of patients with coronary disease who do not receive proper medical care.[27] It has been suggested that the limited use of treadmill scores in the clinical setting may be due to the complex nature of the multivariate equations that scores usually require. Most of the scores we have compared in this study are easy to calculate and rely on fewer variables than the equations recommended by the ACC/AHA guidelines. In fact, these simple scores did as well or better than the consensus of scores, which requires computerization.

Only 3 other studies have compared scores with physician estimates of disease. Detrano et al[27] performed one of the first such studies. They derived a score for estimating probabilities of significant and severe coronary disease, then validated and compared it with the assessments of cardiologists. The score performed at least as well as the clinicians 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. They concluded that the application of scores or consultation with cardiologists not directly involved with patient management might assist in more rational assessments and decision making. Hlatky et al[28] validated 2 scores by comparing their diagnostic accuracy with those 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[29] considered scores for prognosis (rather than diagnosis) with 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 these earlier studies that concluded scores can predict angiographic results and prognosis as well as or better than physicians. Our study is the first to present data on both several treadmill scores and different physicians comparing their ability to predict angiographic coronary disease and mortality.

This study compared physician groups with a prognostic treadmill score based on standard predictive variables. The DTS was compared even though it was derived using infarct free survival as the outcome and certainly is the ideal score for this outcome. However, it did as well as both a Veterans Affairs simple prognostic score and expert cardiologists, and better than other physician groups. Because the cut points for risk of the scores were derived in different populations, they did not equally or similarly divide our population. Nonetheless, the results reveal that the scores performed similarly to the expert cardiologists.

One limitation of this study is that the analysis was performed on an entirely male VA population. Treadmill scores have been validated in female populations,[30,31] but a comparison of scores and physicians with a female population should be pursued. Another limitation is that the DTS was derived on a population in which 75% luminal occlusion was considered the criteria for disease. The recommended cut points for classifying patients into high, low, and intermediate probability groups were based on the 75% occlusion criteria, which would result in lower disease prevalence. This explains the higher number of false-negative results (missed patients) when the DTS stratified patients into probability groups. Modification of the DTS cut points would result in improved classification. Another limitation of this study is the use of all-cause mortality instead of coronary events and cardiac death, as well as no data on intervention or nonfatal events during follow-up.

The low return rate for the random cardiologists and random internists can be considered a limitation of this study. However, it seems likely that physicians who are confident in their ability would take the time to fill out the patient report evaluations. This should therefore favor the physician groups. The final limitations of this study are its retrospective design and workup bias.

In conclusion, treadmill scores perform as well as or better than physicians in predicting angiographic results and mortality and should be calculated as the part of the interpretation of every exercise test performed.[32]

References

1. Gibbons RJ, Balady GJ, Beasely JW, et al. ACC/AHA guidelines for exercise testing. J Am Coll Cardiol 1997;30:260-315.
2. Randolph AG, Guyatt GH, Calvin JE, et al. Understanding articles describing clinical prediction tools: evidence based medicine in critical care group. Crit Care Med 1998;26:1603-12.
3. Richardson WS, Wilson MC, Guyatt GH, et al. User's guides to the medical literature; XV: how to use an article about disease probability for different diagnosis. Evidence-Based Medicine Working Group. JAMA 1999;281:1214-9.
4. Hlatky MA, Califf RM, Harrell FE, et al. Clinical judgment and therapeutic decision making. J Am Coll Cardiol 1990;15:1-14.
5. Froelicher VF, Fearon W, Ferguson C, et al. Lessons learned from studies of the standard exercise test. Chest 1999;116:1442-51.
6. Yamada H, Do D, Morise A, et al. Review of studies utilizing multivariable analysis of clinical and exercise test data to predict angiographic coronary artery disease. Prog Cardiovasc Dis 1997;39:457-81.
7. Lipinski M, Do D, Atwood JE, et al. Can physicians diagnosis coronary disease as well as scores? Arch Intern Med 2001;161:2239-44.
8. Raxwal V, Shetler K, Morise A, et al. A simple treadmill score. Chest 2001;119:1933-40.
9. Froelicher V, Shiu P. Exercise test interpretation system. Physicians Computers 1996;14:40-4.
10. Woltjuis R, Froelicher VF, Fischer J, et al. New practical treadmill protocol for clinical use. Am J Cardiol 1977;39:697-700.
11. Myers J, Buchanan N, Walsh D, et al. A comparison of the ramp versus standard exercise protocols. J Am Coll Cardiol 1991;17:1334-42.
12. Myers J, Do D, Herbert W, et al. A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data. Am J Cardiol 1994;73:591-6.
13. Reid M, Lachs M, Feinstein A. Use of methodological standards in diagnostic test research. JAMA 1995;274:645-51.
14. Fletcher GF, Froelicher VF, Hartley LH, et al. Exercise standards: a statement for health professionals from the American Heart Association. Circulation 1995;91:580-632.
15. Shue P, Froelicher V. Extra: an expert system for exercise reporting. J Noninvas Test 1998;II-4:21-7.
16. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet 1994;344:563-70.
17. Mark DB, Hlatky MA, Harrell FE Jr, et al. Exercise treadmill score for predicting prognosis in coronary artery diesase. Ann Intern Med 1987;106:793-800.
18. Mark DB, Shaw L, Harrell FE Jr, et al. Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. New Engl J Med 1991;325:849-53.
19. Shaw LJ, Peterson ED, Shaw LK, et al. Use of a prognostic treadmill score in identifying diagnostic coronary disease subgroups. Circulation 1998;98:1622-30.
20. Do D, West J, Morise A, et al. A consensus approach to diagnosing coronary artery disease based on clinical and exercise test data. Chest 1997;111:1742-9.
21. Detrano R, Bobbio M, Olson H, et al. Computer probability estimates of angiographic coronary artery disease: transportability and comparison with cardiologist's estimates. Comput Biol Res 1992;25:468-85.
22. Morise AP, Detrano R, Bobbio M, et al. Development and valida-

tion 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;20:1187-96.
23. Prakash M, Myers J, Froelicher VF, et al. Clincial and exercise test predictors of all-cause mortality: results from more than 6000 consecutive referred male patients. Chest 2001;120:1003-13.
24. Morise AP, Diamond GA, Detrano R, et al. The effect of disease-prevalence adjustments on the accuracy of a logistic prediction rule. Med Decis Making 1996;16:133-42.
25. Fleischmann KE, Hunink MG, Kuntz M, et al. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA 1998;280:913-20.
26. O'Rourke RA, Brundage BH, Froelicher V, et al. American College of Cardiology/American Heart Association expert consensus document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. Circulation 2001;102:126-40.
27. 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;25:468-85.
28. Hlatky M, Botvinick E, Brundage B. Diagnostic accuracy of cardiologists compared with probability calculations using Bayes' rule. Am J Cardiol 1982;49:1927-31.
29. Lee KL, Pryor DB, Harrell FE, et al. Predicting outcome in coronary disease. Statistical models versus expert clinicians. Am J Med 1986;80:553-60.
30. Morise A, Froelicher V. A new simplified exercise test score for the diagnosis of coronary artery disease in women: development using gender specific variables [abstract]. Circulation 2000;2:738.
31. Alexander KP, Shaw LJ, Shaw LK, et al. Value of exercise treadmill testing in women. J Am Coll Cardiol 1998;32:1657-64.
32. Swets JA, Dawes RM, Monahan J. Better decisions through science. Sci Am 2000;4:82-7.

Exclusive Mosby Online Exclusive Mosby Online

The following article is an AHJ Online Exclusive.
Full text of this article is available at no charge at our Web site:
www.mosby.com/ahj.

Morphologic changes in the microcirculation induced by chronic smoking habit: A videocapillaroscopic study on the human labial mucosa

Raffaele Molino Lova, MD,[a] Benedetta Miniati, MD,[a] Claudio Macchi, MD,[a] Massimo Gulisano, MD,[b] Gherardo Gheri, MD,[b] Claudio Catini, MD,[b] Andrea A. Conti, MD,[a,c] and Gian Franco Gensini, MD[a,c] ***Florence, Italy***

Background Cigarette smoking has been identified as a major risk factor for cardiovascular and pulmonary diseases. Although an impressive literature does exist on the subject, no evidence is available on morphologic changes induced with chronic smoking habit in the human microcirculation.

Subjects and Methods One hundred healthy subjects, 50 smokers and 50 nonsmokers, underwent videocapillaroscopy of the labial mucosa. For each subject, the caliber of capillary loops, the number of visible capillary loops, the background optical transmittance, the tortuousness of capillary loops, the presence of microaneurysms, the presence of microhemorrhages, and the cumulative smoking habit (pack-year index) were noted.

Results Smokers had a lower caliber of capillary loops ($P < .001$), with a higher number of visible capillary loops ($P < .001$), a lower background optical transmittance ($P < .001$), and a more marked tortuousness of capillary loops ($P < .001$). Microaneurysms and microhemorrhages were absent in nonsmokers, and 1 smoker of 3 had microaneurysms alone, and 1 smoker of 3 had both microaneurysms and microhemorrhages. A significant correlation was found between cumulative smoking habit and tortuousness of capillary loops ($P < .001$) and between cumulative smoking habit and total score (tortuousness score + microaneurysm score + microhemorrhage score; $P < .005$).

Conclusion Chronic smoking habit does induce significant morphologic changes in the microcirculation of the human labial mucosa, and these changes can be easily and comfortably recorded with videocapillaroscopy. (Am Heart J 2002;143:e2.)

4/90/121461