Better Decisions Through Science: Exercise Testing Scores

Victor Froelicher, Katerina Shetler, and Euan Ashley*

Statistical tools can be used to create scores for assisting in the diagnosis of coronary artery disease and assessing prognosis. General practitioners and internists frequently function as gatekeepers, deciding which patients must be referred to the cardiologist. Therefore, they need to use the basic tools they have available (ie, history, physical examination and the exercise test) in an optimal fashion. Scores derived from multivariable statistical techniques considering clinical and exercise data have demonstrated superior discriminating power compared with diagnosis only using the ST segment response. In addition, by stratifying patients as to probability of disease and prognosis, they provide a more practical management strategy than a response of normal or abnormal. Although computers, as part of information management systems, can calculate complicated equations and derive these scores, physicians are reluctant to trust them. However, when represented as nomograms or simple additive discrete pieces of information, scores are more readily accepted. The scores have been compared with physician judgment and have been found to estimate the presence of coronary disease and prognosis as well as expert cardiologists and often better than nonspecialists. However, the discriminating power of specific variables from the medical history and exercise test remains unclear because of inadequate study design and differences in study populations. Should expired gases be substituted for estimated METs? Should ST/heart rate index be used instead of putting ST depression and heart rate separately into the models? Should right-sided chest leads and heart rate in recovery be considered? There is a need for further evaluation of these easily obtained variables to improve the accuracy of prediction algorithms, especially in women. The portability and reliability of scores must be ensured because access to specialized care must be safeguarded. Assessment of the clinical and exercise test data and application of the newer scores can empower the clinician to assure the cardiac patient access to appropriate and cost-effective cardiologic care.

Coronary artery disease (CAD) continues to be the leading cause of morbidity and mortality in the United States. Although the incidence of coronary artery disease has been decreasing over the last 2 decades, the prevalence is expected to increase given the increasing proportion of the population that is elderly.[1-3] In addition, despite efforts to control costs, first with DRGs and more recently with health maintenance organizations (HMOs), health care costs had the greatest increase in the decade last year. With half of the cost increases owing to highly effective cardiovascular pharmaceuticals that decrease heart disease interventions and events, the next target of cost containment must be expensive diagnostics and interventions. It is important to implement clinically cost-effective strategies that direct the appropriate patients to the optimal procedures through clinical risk prediction. There are statistical tools available that can make this possible,[4] and these authors and others have attempted to apply them to patients presenting with chest pain using clinical and exercise test variables.

The goal of clinical risk prediction through statistical methods is to provide the clinician and the

From the Cardiology Division, Veterans Affairs Palo Alto Health Care System, Stanford University, Palo Alto, CA; and Radcliff Infirmary, Oxford University, Oxford, UK.*

Address reprint requests to Victor Froelicher, MD, Cardiology Division (111C), VA Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304; vicmd@aol.com/www.cardiology.org

0033-0620/02/4405-0001$35.00/**22/1/122693**

doi:10.1053/pcad.2002.122693

Table 1. Five Common Mistakes Made When Evaluating Diagnostic Tests

1. Using a target population that consists of only normal subjects and only those with severe disease (limited challenge)
2. Failure to limit work-up bias
3. Using heart rate targets to exclude patients
4. Inclusion of patients with myocardial infarction
5. Use of surrogates instead of appropriate measurements or endpoints

patient with a logical estimate as to the likelihood of the occurrence of important deleterious clinical events. The most important outcome is death. However, the future risk of nonfatal clinical outcomes, particularly myocardial infarction, stroke, symptomatic heart failure, hospitalization for unstable angina, as well as changes in functional capacity and overall quality of life, is an important element of risk evaluation. Besides providing increased predictive accuracy, these statistical methods eliminate physician bias and lessen the variability of decision making.[5,6] Physicians do not always follow a totally rational decision-making process but often make clinical decisions based on personal experience and heuristics.[7] By eliminating the intuitive aspect of decision making, statistics can provide an unbiased evaluation of patients.

Criteria for Evaluation of Studies of Diagnostic Techniques

Studies describing the value of diagnostic tests including scores must be evaluated by standardized rules. Biostatiticians have presented these rules so that diagnostic technologies can be properly evaluated before they are adopted for practice.[8-11] Critical to fulfilling the rules are that only consecutive patients presenting with the symptoms or signs of the disease being diagnosed are used to evaluate the test or score and that work-up bias is reduced. Table 1 lists the common mistakes that have been made by researchers attempting to determine the diagnostic characteristics of a test.

Consideration of More Than the Exercise Electrocardiogram

Since the seminal work of Ellestad et al[12] demonstrating that combining other clinical and exercise parameters along with the ST responses could improve the accuracy of the test, many clinical investigators have published studies proposing multivariable equations to enhance the accuracy of the standard exercise test.[13] In much the same way that clinicians take account of all the clinical information on a patient before making a decision, diagnostic and prognostic predictive accuracy increases when multiple pieces of information from the patient's clinical history and the treadmill test are integrated,[14-18] but issues remain about their portability.[19,20] The American College of Cardiology/American Heart Association (ACC/AHA) guidelines for exercise testing strongly advocated the use of scores.[21]

Statistical Techniques for Diagnosis

Multivariable Discriminate Statistical Techniques

When developing a prediction rule, investigators consider variables that they believe may predict the occurrence of the outcome. The variables found to have discriminating power (ie, clinical information and treadmill responses) are combined to form an algorithm for estimating the probability of CAD. Many mathematical techniques are available for demonstrating what variables are predictive as well as their relative predictive power. Regression analysis methods are especially attractive because they make it possible to derive complex regression functions directly from a database. Logistic regression has been preferred because it models the relationship to a sigmoid curve, (which often is the mathematical relationship between a probability variable and an outcome), and its output is between 0 and 1 representing the probability of disease being present (ie, from 0 to 100% probability of the predicted outcome).[22] In addition, in the logistic regression analysis, dichotomous (ie, yes, no) and continuous variables (a number such as heart rate; ie, 110 bpm) can be considered together. The general linear logistic regression model used takes the following form:

$$\text{Probability (0 to 1)} = 1/(1 + e^{-(a+bx+cy....)}) \quad (1)$$

Where *a* is the intercept, *b* and *c* are coefficients, and *x* and *y* are variable values. Usually forward selection of the variables is used with entry at a significance level less than .05. The model separates patients with and without a given outcome (ie, CAD). In comparison, the output of a discriminate function is a unitless numerical score while a logistic regression provides an actual probability.

Rather than require computation of equations to estimate probability or hazard, later we will describe an approach that results in a very simple linear score in which the health care provider merely compiles the variables in the score, multiplies by the appropriate number, and then adds up the products. Surprisingly, these simple linear scores have the same receiver operating characteristic (ROC) areas as the more complicated equations.

The ability of any score or measurement to diagnose a disease (ie, CAD) depends on how much the score differs among those with and without the disease. These measurements could be ST segment depression, calcium score using electron beam computed tomography, perfusion scan values, or echocardiographic wall motion estimates. Figure 1 consists of actual data from over 1000 male veterans who underwent both exercise testing and coronary angiography. Unfortunately, as shown in the figure, the values for the score or measurement usually greatly overlap. The better the test, the further apart the curves and the less they overlap. The cut-point we chose of 50 is a practical choice for the treadmill score we use so that those above 50 are considered to have disease and those below are considered to be free of CAD. However, as can be seen, this is not really the case. Figure 2 shows how the two curves separately considered with the cut-point result in the 4 classifications (true positives [TP], true negative [TN], false positive [FP], false negatives [FN]) that permit the calculations of the standard assessments of test performance (sensitivity [bottom curve of population distribution] and specificity [top curve]) and shows their inverse relationship.

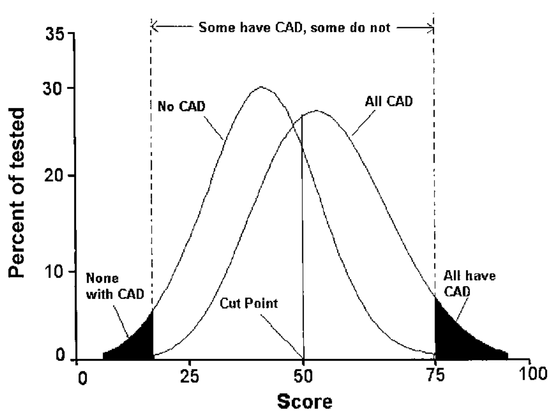

Fig 1. Range of characteristics (ROC) plots for the simple treadmill score for those with and those without angiographic coronary artery disease (CAD).

Score Evaluation (ROC Curves)

The accuracy of the model to separate is assessed by means of the area under a ROC curve. ROC curve analysis is based on the plotting of sensitivity and specificity for a range of cut-points (criteria for abnormal) for a test measurement or the value of a score. The area ranges from 0 to 1, with 0.5 corresponding to no discrimination (ie, random performance), 1.0 to perfect discrimination, and values less than 0.5 to worse-than-random performance. Most prediction rules, like other diagnostic modalities, have a range of possible results. Several possible cut-off criteria could be used to separate results into "positive" and "negative" groups. For each criterion chosen, the rule will have a different sensitivity and specificity. An ROC curve is a plot of the sensitivity versus specificity for the full range of the score. The shape of the curve shows the trade-offs between sensitivity and specificity produced at different criteria with specificity and sensitivity being inversely related. The area under the ROC curve is a measure of the ability of the rule to discriminate between presence and absence of disease, independent both of cut-off criteria and of disease prevalence.[23] Figure 3 is a ROC plot of the authors' simple treadmill score ranging from 0 to 100 illustrating two other cut-points, 40 and 60. These cut-points could be appropriate for particular purposes of the test; ie, screening well people where a high specificity is needed or for ruling out ischemia after presentation to an emergency department for chest pain where high sensitivity is required. Figure 4 shows comparison of the diagnostic characteristics of the Morise pre-test clin-

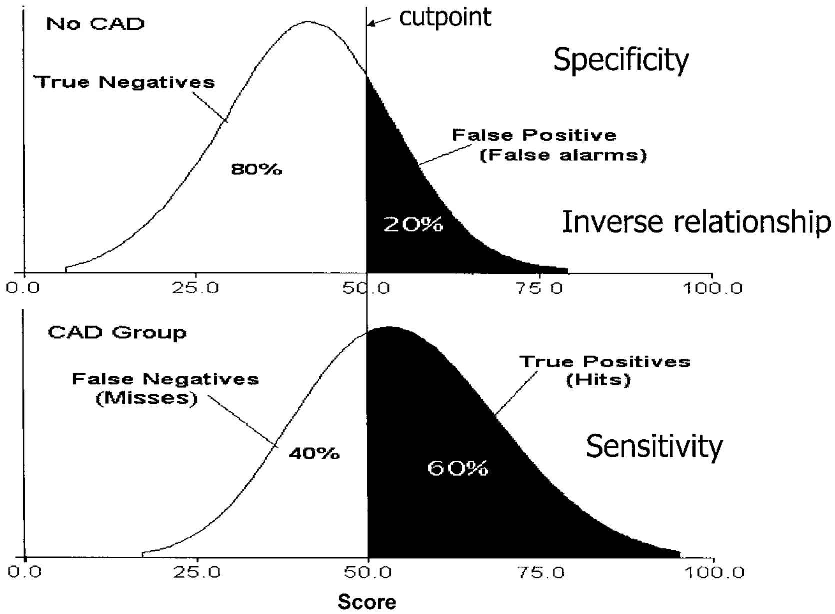

Fig 2. Separate frequency plots indicating the 4 test responses that enable calculation of test characteristics (true-positives, true-negatives, false-positives, false-negatives). Sensitivity = TP/TP + FN; Specificity = TN/TN + FP.

ical score, ST analysis alone, and the authors' simple treadmill score. The four curves allow for a comparison of the diagnostic value of these techniques. The treadmill test clearly adds to the discriminatory value of clinical data. Surprisingly, every attempt to use a computer to improve on visual ST analysis of the exercise test failed to improve the diagnostic value of the test although a score clearly is an improvement over ST analysis alone.

Although a score's ability to separate those with and without disease persists in another group of patients, it must be confirmed that the score's calibration is the same. That is, does the value of the score for 70% probability of disease as determined

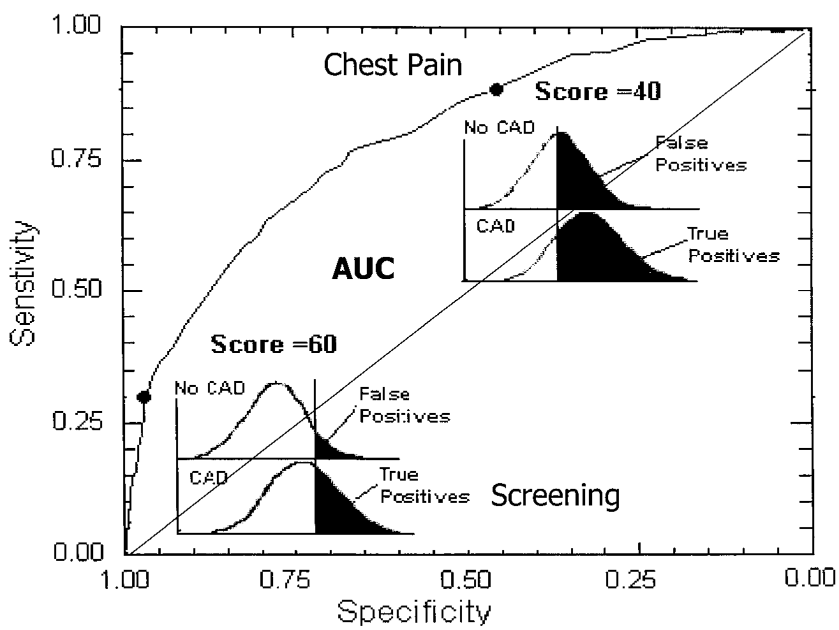

Fig 3. ROC plot of the simple treadmill score showing how different cut-points can be chosen according to the specific use of the test.

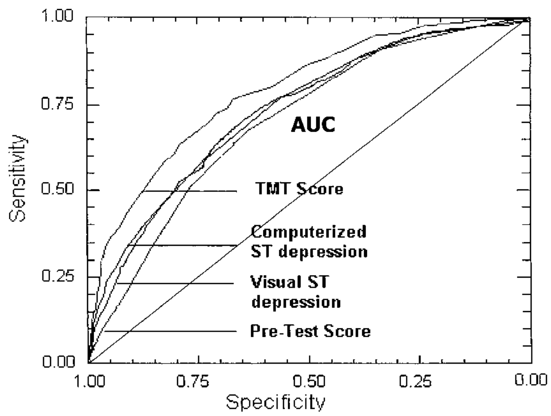

Fig 4. ROC plots comparing the discriminating power of a pre-test score, ST measurements, and a simple treadmill test score.

in one group, for instance, still represent 70% instead of another probability in another group?

Pretest Scores

In the guidelines, the exercise test is the recommended test for diagnosing CAD in patients at intermediate probability for CAD. The classification of pretest probability is enabled through a table considering age, gender, and chest pain characteristics using the Diamond-Forrester tabular method[21] (Table 2). The intermediate pretest probability category was assigned a class I indication, whereas the low and high pretest probability were assigned class IIb indication for exercise testing.

Morise et al[24] proposed a pretest score for categorizing patients with suspected coronary disease and normal resting electrocardiograms (ECGs) that possibly is superior to the method advocated by the guidelines. The Morise score is calculated as follows:

$$\begin{aligned}&\text{Age code} + \text{Angina pectoris code}\\&+ \text{diabetes} \times 2 + \text{hypertension}\\&+ \text{smoking now} + \text{hypercholesterolemia}\\&+ \text{family history of CAD} + \text{obesity}. \qquad (2)\end{aligned}$$

Where age less than 40 = 3, age between 40 to 55 = 6, and age more than 55 = 9 for men and less than 50 = 3, 50 to 65 = 6, and greater than 65 = 9 for women. For estrogen status, 3 points were subtracted for positive and 3 points were added for estrogen negative status. Typical chest pain = 5, atypical chest pain = 3, nonanginal chest pain = 1, and no chest pain = 0. For diabetes mellitus, 2 points were added and 1 point was added for each of the other 5 risk factors (hypertension, present smoking, hypercholestrolemia, family history CAD and obesity) (Fig 5). In a subsequent paper, Morise et al proposed that this score was superior to the guidelines method of stratifying patients and compared these two methods of determining patient's pre-test probability of disease.[25]

Exercise Test Diagnostic Scores

What Variables Should be in Diagnostic Scores?

Numerous exercise test and clinical variables have been proposed as predictors of angiographic dis-

Table 2. Estimates of Pre-Test Probability of Coronary Disease by Symptoms, Gender, and Age as Recommended by the AHA/ACC Exercise Test Guidelines[21]

Age	Gender	Typical/Definite Angina Pectoris	Atypical/Probable Angina Pectoris	Nonanginal Chest Pain	Asymptomatic
30-39	Men	Intermediate	Intermediate	Low (<10%)	Very low (<5%)
	Women	Intermediate	Very low (<5%)	Very low	Very low
40-49	Men	High	Intermediate	Intermediate	Low
	Women	Intermediate	Low	Very low	Very low
50-59	Men	High (>90%)	Intermediate	Intermediate	Low
	Women	Intermediate	Intermediate	Low	Very low
60-69	Men	High	Intermediate	Intermediate	Low
	Women	High	Intermediate	Intermediate	Low
High = >90%		Intermediate = 10 to 90%	Low = <10%		Very low = <5%

There is no data for patients younger than 30 or older than 69, but it can be assumed that coronary artery disease prevalence increases with age.

Variable	Circle response	Sum
Age	Men<40, Women<50 = 3	
	Men40-55,Wom50-65=6	
	Men>55,Women>65 =9	
Estrogen Status	Positive=-3	
	Negative=+3	
Diabetes	Yes = 2	
Obesity?	Yes = 1	
Family History?	Yes = 1	
Hypercholesterolemia?	Yes = 1	
HBP?	Yes = 1	
Smoking?	Yes = 1	
	Total Score	

Choose only one per group

<=8 low prob

9-15 = intermediate probability

>=16 high probability

Fig 5. Calculation of the simple clinical score for angiographic coronary disease.

ease. This article reviews 24 studies attempting to predict presence of any angiographic disease and lists the 30 equations created.[13] In Table 3, the number of equations that chose the specific variable to be a candidate in the model is the numerator and the total number of studies that considered it is the denominator. The variables chosen in more than half of the studies are marked with an asterisk.

Validated Diagnostic Scores

Detrano et al were among the first investigators to use modern statistical techniques to derive a score: they included 3,549 patients from 8 institutions who underwent exercise testing and angiography between 1978 and 1989.[26] Disease was defined as greater than 50% diameter narrowing in at least one major coronary arterial branch, and the prevalence of disease according to this criterion was 64%. They considered a total of 15 clinical and exercise variables, which contributed significant and independent information to disease probability and had been judged clinically relevant by a panel of cardiologists as candidates for logistic regression. In another seminal study, Morise et al studied a total of 915 consecutive patients without a history of prior myocardial infarction or coronary artery bypass surgery who were referred to the exercise laboratory at West Virginia University Medical Center between 1981 and 1994 for evaluation of coronary disease.[27] Disease was defined as greater than 50% diameter narrowing in at least one major coronary arterial branch, and the prevalence of disease according to this criterion was 41%. Their logistic regression

Table 3. Results From Meta-Analysis of Studies With Angiographic Findings as theGold Standard for Any Significant Coronary Disease

Variables	Fraction Time a Variable Is Selected as a Significant Predictor When the Variable Was Considered, %	
Gender*	20/20	100
Chest pain symptoms*	17/18	94
Age*	19/27	70
Elevated cholesterol*	8/13	62
Diabetes mellitus	6/14	43
Smoking history	4/12	33
Abnormal resting ECG	4/17	24
Hypertension	1/8	13
Family history of CAD	0/7	0
ST segment slope*	14/22	64
ST segment depression*	17/28	61
Maximal heart rate*	16/28	57
Exercise capacity	11/24	46
Exercise-induced angina	11/26	42
Double product	2/13	15
Maximal systolic BP	1/12	8

*Chosen more than half the time.
ECG, electrocardiography; CAD, coronary artery disease; BP, blood pressure.
Adapted from Yamada et al.[13]

equations are listed in the exercise testing guidelines.[21]

Consensus of Scores

In an attempt to make the scores more portable (valid when applied to other populations), the authors proposed a consensus approach.[28] Consideration that NASA calculates spacecraft trajectories using several equations and then plots the trajectories using the equations that agree, led the authors to do the same with coronary disease. They used the Detrano and Morise equations along with their own equation. A probability score was calculated for each patient using the Long Beach-Palo Alto VAMC equation and Detrano and Morise validated equations. Thresholds were set in each equation; if a patient was high probability in at least 2 of the 3, he or she was considered high probability; similarly, if low in at least 2 of the equations, he was low risk. All others would be intermediate. Because the patients in the intermediate group would be sent for further testing and would eventually be correctly classified, the sensitivity of the consensus approach was 94% and specificity was 92%. The percent of correct diagnosis increased from the 67% for standard exercise ECG analysis and from the 77% for multivariable predictive equations alone to greater than 90% correct diagnoses for the consensus approach. The consensus approach avoids the need to calibrate the equations for disease prevalence, and it avoids some of the problems associated with missing data, differences in the definition of collected variables, and even angiographic interpretation and criteria. Because of the complexity of the equations involved, the consensus approach is only practically applied using a computer program.[29]

"Simplified" Score Derivation

Simplified scores that only require physicians to only add points have been developed for pre-test estimates of disease and for prognosis. To develop a simple score for diagnosis, the authors analyzed data from 2 Veterans Administration Medical Centers.[30] All 1,276 male patients had coronary angiography within 4 months of their treadmill test. The score derived was then validated in 476 men from another institution. METs were estimated in this diagnostic model by maximal heart rate, whereas the converse is the situation for predicting prognosis. However, maximal heart rate and METs were highly correlated and METs would take the place of maximal heart rate in the equation if heart rate was not considered.

Three steps were used to derive the new treadmill score. Initially, the authors validated the pre-test score of Morise by comparing it with a pre-test equation derived in their population (ROC AUC = 0.71 *v* 0.73, nsd).[24] Second, the authors derived an equation considering only the hemodynamic variables (ie, METs, systolic blood pressure, maximal heart rate, and treadmill angina index) in a logistic regression model (ROC AUC = 0.68). Third, the authors entered the Morise pre-test score, the hemodynamic equation, and exercise-induced ST depression into a logistic model. The resulting equation exhibited a ROC AUC of approximately 0.79. The variables previously chosen were reconsidered in a logistic model that eliminated some variables.

To decrease the complexity of the predictive equations, the authors converted the variables chosen in logistic regression into a simple linear score. They first coded all variables with the same number of intervals so that the coefficients would be proportional. Then they coded the bin with the larger value to be associated with higher probability of disease. For instance, if 5 is the chosen interval, dichotomous variables are 0 if not present and 5 if present and continuous variables like age and heart rate are coded in 5 bins by appropriate ranges. All codes then would be directly related to probability (ie, a heart rate code of 5 would be a low heart rate, whereas age code of 5 would be for the oldest individuals) and the smallest coefficient is associated with the least important variable. The multiplier of this least important variable was reduced to unity and the other coefficients into their proportional weight or importance by dividing each coefficient by the smallest coefficient. This makes the relative importance of the selected variables very obvious. This approach results in a very simple linear score in which the health care provider merely compiles the variables in the score, multiples by the appropriate number, and adds up the products. Surprisingly, these simple linear scores have the same ROC areas as the more complicated equations requiring the calculation of exponentials.

Men's Score:

6 × maximal heart rate
+ 5 × ST depression code
+ 4 × age code + angina pectoris code
+ hypercholesterolemia code + diabetes code
+ treadmill angina index code (3)

This diagnostic score did not perform well for women (AUC < .65), so Morise developed a specific score for women.

Patients were assigned into 3 groups. The group with low probability of CAD was defined as having a score of less than 40, intermediate a score between 40 and 60, and high probability as a score above 60. The prevalence of any CAD in the low probability group was 27%, 62% in the intermediate, and 92% in the high probability group that was comparable to the validation group with 22%, 58%, and 92% in low, intermediate, and high probability groups, respectively. The scores show good portability and require use of simple coding that can be carried on index cards (see Fig 6 for men and Fig 7 for women). Their diagnostic characteristics were unaffected by resting ST depression,[31] β-blockade or chronotropic incompetence,[32] or diabetes.[33]

Score-Directed Management Strategy

Scores can also provide a management strategy for patients with possible CAD rather than just classifying them as diseased or not diseased. This is done by placing patients into 3 categories of risk rather than just dichotomizing them as positive or negative.[34] It is important to determine the cut-off point or threshold of the post-test probability to accept the individual patient as being diseased. Kotler and Diamond[35] reported on the means to define the upper and lower thresholds of the intermediate probability group. The lower threshold level is the cut-point below which the number of false-positive responses exceeds the number of true-positive responses. An upper threshold level is the cut-point above which the number of false-negative responses exceeds the number of true-negative responses. In other cases, the lines are drawn arbitrarily.

Low-risk patients have an excellent prognosis and may be risk stratified by the treadmill test. This patient cohort can be managed safely with watchful waiting as well as symptomatic medical therapy without further testing. High-risk patients should be considered candidates for more aggressive management that may include cardiac catheterization. In patients with an intermediate-

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	Occurred =3	
induced Angina	Reason for stopping =5	
	Total Score:	

Males

Choose only one per group

<40=low prob

40-60= intermediate probability

>60=high probability

Fig 6. Calculation of the simple score for angiographic coronary disease in men.

Variable	Circle response	Sum
Maximal Heart Rate	Less than 100 bpm = 20	
	100 to 129 bpm = 16	
	130 to 159 bpm =12	
	160 to 189 bpm =8	
	190 to 220 bpm =4	
Exercise ST Depression	1-2mm =6	
	> 2mm =10	
Age	>65 yrs =25	
	50 to 65 yrs = 15	
Angina History	Definite/Typical = 10	
	Probable/atypical =6	
	Non-cardiac pain =2	
Smoking?	Yes=10	
Diabetes?	Yes=10	
Exercise test	Occurred =9	
induced Angina	Reason for stopping =15	
Estrogen Status	Positive=-5, Negative=5	
	Total Score	

Women

Choose only one per group

<37=low prob

37-57= intermediate probability

>57=high probability

Fig 7. Calculation of the simple score for angiographic coronary disease in women.

probability treadmill score, myocardial perfusion imaging and other tests (Table 4) appear to be of value for further risk stratification.[36-40]

Should the ST/HR Index be Included in Scores?

The ST heart rate index is attractive theoretically, but it has not been validated[41] probably for the following reasons. A ratio requires that both numerator and denominator have equal weights for prediction. This clearly is not the case for ST depression and heart rate. First, the ranges and units are totally different, and secondly, when heart rate and ST depression are entered into logistic regression (or any multivariable model) they have totally different weights. The only way to deal with the potential variables for prediction of angiographic disease is to put them into a mathematical model and allow them to be selected or not and to be weighted as to their predictive power if they are chosen. The studies supportive of the ST/HR index have not followed this approach and have "broken" some of the rules for evaluating diagnostic procedures.[9] Morise's report[42] of 1,358 individuals exercise tested including only 152 with catheterization data and Okin's report[43] considering heart rate reserve had 238 control subjects and 337 patients with CAD. The Helsinki study considered the maximum value of the ST segment depression/heart rate (ST/HR) hysteresis over a different number of leads for the detection of CAD.[44] The study population consisted of 127 patients with CAD and 220 patients with a low likelihood of the disease referred for an exercise test. Unfortunately, these studies did not consider consecutive patients who reported chest pain. Those tested were a limited challenge because extremes of pre-test probabilities were chosen. Limited challenge favors ST/HR index since the

Table 4. Paradigm for the Clinical Reaction to the Estimated Probability of CAD

Probability for clinically Significant CAD	
Low probability	Patient reassured symptoms most likely not caused by CAD
Intermediate probability	Require other tests, such as stress echo, nuclear, or angiography to clarify diagnosis; anti-anginal medications tried
High probability	Anti-anginal treatment indicated; intervention clinically appropriate; angiography may be required

CAD, coronary artery disease.

healthy individuals have relatively high heart rates and the sick have low heart rates. The only study that evaluated ST/HR index and included consecutive patients with chest pain was QUEXTA.[45] This large study was multi-center also followed a protocol to reduce workup bias and was analyzed by independent statisticians. ST/HR slope or index had no superiority to simple measurement of the ST segment. The studies in healthy people[46,47] using ST/heart rate index have little relevance to diagnostics because Baysian concepts limit any of the available tests for screening asymptomatic, low-prevalence populations.[48]

How About Right-Sided Chest Leads?

The local effect of right ventricular ischemia could cause ST elevation over the right-sided leads, but even perfusion scanning does not have the resolution to validate ischemia in the thin right ventricle. The norms for the right-sided ST response need to be determined by appropriate studies, and the findings of the Athens group[49] must be validated before right-sided leads are used clinically to diagnose ischemia.[50]

Statistical Techniques for Prognosis

Other statistical techniques are required to predict death or cardiac events to develop prognostic scores. All variables should be explored by means of Kaplan-Meier survival curves for univariate comparisons and the Cox model for multivariate analysis. A Cox proportional hazards model should be used to determine the effect of a given independent variable on time to death. Many of the variables univariately predictive of death are likely to have overlapping prognostic significance. A multivariate stepwise Cox regression analysis can be used. The Cox model assumes that the hazard that equals the instantaneous death rate is given by the formula:

$$\begin{aligned} h_i(t) &= h(t)\ C_i, \text{ where } C_i \\ &= \exp\,(B_1X_{1i} + B_2X_{2i} + \ldots \\ &\quad + B_pX_{\pi}) \end{aligned} \tag{4}$$

The model assumes that the hazard (h) of death for patient j at the time t (h_j (t)) equals the hazard of death for an "average patient" at the same time (h(t)) multiplied by a factor (C_i) that is a function of the prognostic profile of patient i; this is the proportional hazards assumption that gives the model its name. The proportional hazard coefficient for patient i (C_i) is, in turn, a function of the values for that patient of a set of prognostic factors ($X_{1i}, \ldots, X_{pi}$), multiplied by a corresponding set of regression coefficients ($B_1, \ldots, B_p$) that measures the strength of the association between the prognostic factor and outcomes of large numbers of patients with the same disorder.[51] The Cox model also assumes that the effect of a prognostic factor on outcome is linear. Variables of prognostic significance may be discrete or they may be continuous. Many studies analyze the strength of a continuous prognostic factor by setting an arbitrary "cut-point" and dividing the patients into subgroups with values above and below the cutpoint. Although this technique is helpful to illustrate findings and to facilitate drawing survival curves, it discards valuable prognostic information and may weaken the apparent prognostic significance of a continuous variable.[51]

Endpoints and Censoring

The relative prognostic importance of the ischemic variables can be minimized by not censoring on interventions for ischemia (ie, removal of intervened patients from observation when the intervention occurs in followup) because the intervention stops patients from dying. Consideration of all-cause mortality instead of cardiovascular mortality can have the same effect. This may explain why the ischemic variables included in the Duke score that clearly had diagnostic power[52] do not predict all-cause mortality. Although all-cause mortality has advantages over cardiovascular mortality as an endpoint,[53] the Duke score was generated using the endpoints of infarction and cardiovascular death.[54] Interventions such as bypass surgery or catheter procedures were censored in the Duke study (that is, subjects were removed from the survival analysis when interventions occurred). Such censoring should appropriately increase the association of ischemic variables with outcome by removing patients whose disease has been alleviated and thereby would not be as likely to experience the outcome. Often researchers do

Table 5. Frequency of Clinical and Exercise Test Variables Chosen as Significantly and Independently Associated With Time Until Death in 9 Previous Prognostic Studies

Variable	Out of 9 studies
Clinical	
Age	2
CHF	2
MI by history or Q waves	1
Resting ST depression	1
Exercise Responses	
Exercise capacity (METs)	7
Angina	5
ST depression	4
Maximal heart rate	3
Maximal SBP	2
ST elevation	1
PVCs	1
Maximal double product	1

From Exercise and the Heart, Froelicher and Myers.[64]
CHF, congestive heart failure; MI, myocardial infarction.

not censor patients if they had a cardiovascular procedure during follow-up because they do not have that information. From a previous study using a similar VA patient population with an annual all-cause mortality of 3%, the authors' group found that 75% of deaths were cardiovascular deaths, and that 6% of patients were censored in follow-up because of bypass surgery.[55] The contradictory results regarding the prognostic power of ischemic variables could also be owing to the more effective methods of treatment currently available for ischemic coronary disease compared with left ventricular dysfunction.

The use of coronary interventions as endpoints falsely strengthens the association of ischemic variables with endpoints because the ischemic responses clinically result in the intervention being performed. Although some investigators have justified their use by requiring a time period to expire after the test before using the intervention/procedure as an endpoint, this still influences the associations. Another problem is that variables predicting infarction can be different than those predicting death, creating a situation where one variable's contrasting effects with respect to two endpoints can cancel each other out. A recent editorial discusses the use of prognostic scores in clinical practice.[56]

Exercise Test Prognostic Scores

Previous Studies

Nine studies have considered multiple exercise and clinical variables in a proportionate Cox Hazard model to try to predict outcomes. Table 5 lists the number of times the major prognostic variables were chosen as significantly and independently predictive of time to death out of the times they were considered in the published prognostic studies.[57-64] Table 6 lists the most common mistakes made in prognostic studies—mistakes avoided in the following two studies.

The Long Beach VA Score

We previously created a prognostic VA score using 2,546 male veterans from Long Beach Veterans Affairs Hospital.[65,66] In contrast to the Duke score, which is strictly exercise test based, clinical data were included also. Using multivariable Cox regression analysis, 4 variables with the best predictive power were chosen: (1) history of congestive heart failure or digoxin use, (2) a score for the change in systolic blood pressure during exercise, (3) exercise capacity (METs), and (4) exercise-induced ST depression. The score that was derived from the survival analysis was:

$$\begin{aligned} &5 \times (\text{congestive heart failure or digoxin use} \\ &[\text{yes} = 1, \text{no} = 0]) \\ &+ \text{exercise-induced ST depression in millimeters} \\ &+ \text{change in systolic blood pressure score} \\ &- \text{METs}(5) \end{aligned} \tag{5}$$

Where systolic blood pressure score was equal to 0 for increase of systolic blood pressure greater than 40 mm Hg during exercise test, 1 for increase of 31 to 40 mmHg, 2 for increase of 21 to 30 mm Hg, 3 for increase of 11 to 20, 4 for increase of 0 to 11 mm Hg and 5 for a reduction below standing systolic pre-exercise blood pressure. ST depression is measured in millimeters. Three groups

Table 6. Four Common Mistakes made When Evaluating Prognostic Tests

1. Limited challenge and workup bias
2. Incomplete follow-up
3. Failure to censor
4. Use of misleading endpoints

were formed in which −2 indicated low risk, −2 to 2 indicated moderate risk, and greater than 2 indicated high risk. The annual cardiovascular mortality was less than 2% for low-risk (77% of population), 7% for moderate-risk (18% of cohort), and 15% for high-risk groups (6% of patients).

Duke Treadmill Score

The Duke Treadmill score was validated in the above VA population and can be used both for prognosis[54] and diagnosis.[52] Mark et al developed the Duke treadmill score using data collected from the 2,842 hospitalized patients with known or suspected CAD, all of whom had a catheterization.[67] The score was validated in 613 outpatients evaluated before the decision for cardiac catheterization. Infarct-free survival was the endpoint, censoring was performed on interventions, and a Cox hazard function model was appropriately applied to determine which exercise test variables were independently and significantly associated with time to cardiac event. The Duke score uses the 3 chosen variables and their coefficients: (1) the amount of ST depression, (2) exercise capacity, and whether angina occurred during the test or was the reason for stopping. This can be done using a nomogram (see the AHA/ACC guidelines[21]) or by a computer. (Some of the commercial treadmill systems automatically calculate it or use www.cardiology.org.) When using the nomogram, be sure that you have the version with unevenly spaced vertical lines. An initial printing of the guidelines included a improper version with the spacing equal that produces erroneous mortality estimates.

The final Duke treadmill score was calculated as follows:

$$\text{Exercise time} - (5 \times \text{ST depression}) - (4 \times \text{treadmill angina index}) \quad (6)$$

Exercise time is measured in minutes of the Bruce protocol (METs from other protocols requires transformation to equivalent in Bruce protocol); ST depression is measured in millimeters; and the treadmill angina index is coded from 0 to 2. A value of 0 was assigned if angina was absent, 1 if typical angina occurred during exercise, and 2 if angina was a reason the patient stopped exercising. The final score ranges from +15 or greater (corresponding to a patient who exercises through stage 5 without angina) or ST changes to −25 or less (corresponding to a patient who stops exercising at 3 minutes or less because of angina and who has 4 mm of ST depression). The high-risk group is defined by a score less or equal to −11, moderate-risk group with score ranging from −10 to +4, and low-risk with score more or equal to +5. Patients with a predicted average annual cardiac mortality rate less or equal 1% per year can be managed medically without the need for cardiac catheterization. Patients with predicted average annual cardiac mortality rate more or equal to 3% per year should be considered for cardiac catheterization. Patients with predicted average annual cardiac mortality rate of 1% to 3% per year, including those with suspected left ventricular dysfunction, should have either cardiac catheterization or an exercise imaging study.[68,69]

In addition to providing accurate prognostic estimates, the Duke Treadmill score also provides valuable information about the presence and severity of coronary disease. According to Shaw et al,[52] the Duke Treadmill score was diagnostic for significant and severe CAD. For low-risk patients, 60% had no coronary stenosis and 16% had single-vessel stenosis equal or greater than 75%. More than 80% of high-risk patients had severe coronary disease.[70]

Should Heart Rate in Recovery be Included in Prognostic Scores?

Recent studies have highlighted the prognostic value of heart rate recovery or the decrease in heart rate after an exercise test. While earlier physiological studies suggested a rapid heart rate recovery response to exercise to be a marker of physical fitness, only recently has its prognostic value been reported. The speed of heart rate return to baseline after exercise is theorized to be owing to high vagal tone associated with fitness and good health. While the prognostic value of heart rate recovery has recently been highlighted, its relative value compared with other treadmill responses and its diagnostic value remains uncertain. Table 7 shows comparisons of the studies across a number of important parameters. In the first study, Cole et al. looked at 2,428 adults who had been referred for exercise scintigraphy.[71] Cole

Table 7. Previously Published Prognostic Studies Relating to the Decrease of Heart Rate After Exercise

Study	Population	Sample Size (% Women)	Exclusion Criteria	F/U (y, mean)	Test Protocol/ Recovery Status	Minutes of Recovery/ Cut-Point	Mortality (All-Cause)	Sensitivity/ Specificity For Death	Other Variables Studied	β-Blocker Status
Cole et al[71]	Referral for exercise perfusion; 9% with known CAD	2,428 (37)	CABG, angiography, CHF/digoxin use, LBBB	6	Bruce, with 2-minute cool down, symptom-limited	1 min/12 bpm	213 (9%)	(cut-point-12 bpm) 56%/77% (cut-point-8 bpm) 33%/90%	METs, male sex, age, perfusion defects on scintigraphy, chronotropic incompetence	Used by 12% of study population, no association with abnormal test
Cole et al[72]	Participants in Lipid Research Clinics Prevalence Study, asymptomatic	5,234 (39)	β-blockers, other cardiac meds, h/o cardiovascular disease	12	Bruce, without cool down 85% age-predicted heart rate	2 min/42 bpm	325 (6.2%), 36% believed to be cardiovascular follow-up 100%	54%/69%	No comparison	Excluded
Nishime et al[73]	Referral for ETT; 8% prior CABG, 75% screening asymptomatic, 9% prior MI	9,454 (22)	CHF, LBBB, digoxin, valvular heart disease	5.2	Bruce, with 2-minute cool down, symptom-limited	1 min/12 bpm	312 (3%)	49%/81%	METs, maximal HR, Duke Treadmill score; TM AP score, and EI-ST depression not prognostic	Heart rate recovery not predictive of death in β-blocker group
Watanabe et al[73a]	Referral for ETT-echo	5,785 (37)	CHF, valve disease, afib, pacer	3	N/A, no cool down	1 min/18 bpm	190 (3.5%)	33%/87%	Data not available	N/A
Shetler et al[74]	Referral for standard ETT; 42% with prior MI	2,193 (all men)	CABG, angiography, LBBB, pacer	6.8	Ramp without cool down, symptom limited	2 min/22 bpm	413 (19%)	35%/83%	Age, METs, history of typical angina; TM AP score, and EI-ST depression not prognostic	Used by 34% of the study population, heart rate recovery equally predictive

Abbreviations: F/U, follow-up; CAD, coronary artery disease; CABG, coronary artery bypass grafting; CHF, congestive heart-failure; LBBB, left bundle branch block; ETT, exercise treadmill test; HR, heart rate; TM AP, treadmill test–induced angina pectoris; EI-ST, exercise-induced ST depression; MI, myocardial infarction.

found that using a decrease of 12 bpm or less as the definition of an abnormal response exhibited a relative risk of 4.0 for death, with the group having a value less than 12 having a mortality of 19%, and the group with a value higher than 12 having a mortality of 5% over the 6-year period. The study used the symptom-limited Bruce protocol with a 2-minute cool-down walk, and heart rate recovery was measured at 1 minute after peak exercise. Patients on β-blockers were included in the study, and no difference was seen in the ability of the test to discriminate between low- and high-risk patients in those patients on β-blockers. In this study, the investigators used all-cause mortality and performed survival analysis both with and without censoring of interventions (coronary artery bypass grafting [CABG] and percutaneous transluminal coronary angiography [PTCA]) and found no difference in results.

The same investigators then studied a different patient population.[72] Asymptomatic patients enrolled in the Lipid Research Clinics Prevalence Study underwent exercise testing using a Bruce protocol. The tests were stopped when 85% to 90% of peak heart rate was achieved, and no cool down period was allowed. Heart rate recovery was measured at 2 minutes of recovery. heart rate recovery continued to be a strong predictor of all-cause mortality; patients with an abnormal value had a mortality rate of 10% whereas patients with a normal value had a mortality rate of 4% at 12 years of follow-up.

To further elucidate the power of heart rate recovery in distinct populations, these same investigators then published another study using patients referred for standard treadmill testing.[73] Using the same methods as the original study, the investigators found similar results, although notably the cut-off value for an abnormal test was different. Patients with abnormal heart rate recovery had 8% mortality at 5.2 years, whereas patients with normal heart rate recovery had only 2% mortality. Neither this nor the previous study censored for CABG or PTCA, and this study had 8% patients with CABG enrolled along with 75% asymptomatic individuals. The investigators also compared the prognostic ability of heart rate recovery to that of the Duke Treadmill Score. Although ischemic components of the Duke score did not have prognostic power, METs did since the DTS produced similar survival curves to heart rate recovery. Patients with abnormal DTS and heart rate recovery survival were even further compromised.

Our group attempted to validate the use of heart rate recovery for prognosis in a male veteran population.[74] The mortality rate in this study was higher than in previous studies of heart rate recovery. Using similar statistical analysis, the authors found that heart rate recovery of less than 22 bpm at 2 minutes recovery identified a high-risk group of patients. They also found that β-blockers had no significant impact on the prognostic value of heart rate recovery. Through multivariate analysis, they evaluated the power of several other clinical and treadmill variables to see how they compared with heart rate recovery in their ability to predict poor outcome. Similar to Cole et al,[71] they found that low MET capacity was the most powerfully associated with outcome.

A distinct advantage over previous studies is that we selected a group who underwent coronary angiography. This made it possible to evaluate the diagnostic ability of heart rate recovery. Surprisingly, heart rate recovery was not selected among the standard variables to be included in a logistic model, and its ROC curve did not indicate any discriminatory value. Thus, although heart rate recovery has been validated as an important prognostic variable, it did not help diagnosing coronary disease in this study. Because, in general, these studies did not censor on events or consider event-free survival, heart rate recovery may well just be a surrogate for physical fitness/activity level predicting outcome along with medical therapy. The authors conclude that these studies support the health benefits of a lifestyle of physical activity rather than supporting the addition of heart rate recovery to scores designed to help direct patients to appropriate therapies. However, it is still an important addition to every exercise test performed.

Predicting Prognosis in Women

Clinical presentation, performance in diagnostic tests, and prevalence of CAD are different between men and women presenting with chest pain. To demonstrate the value of exercise testing in women, Duke University researchers analyzed data from 976 women referred for evaluation of chest pain who underwent exercise treadmill test and cardiac catheterization.[75] Women and men differed significantly in the Duke Treadmill score, disease prevalence (32% v 72% significant CAD), and 2-year mortality (1.9% for the study women compared with 4.9% for the men). Mortality increased for higher-risk Duke Treadmill score groups in both genders. Two-year mortality for women was 1.0%, 2.2%, and 3.6%, respectively, for low-, moderate-, high-risk groups. Two year mortality for men was 1.7%, 5.8% and 16.6%, respectively for low, moderate and high risk groups. Because of the differences in disease prevalence, women had better survival at all values of the Duke Treadmill score. Although, overall, women had better survival, the Duke score performed actually better in women than in men for excluding disease, with fewer low-risk women having mild or severe disease.

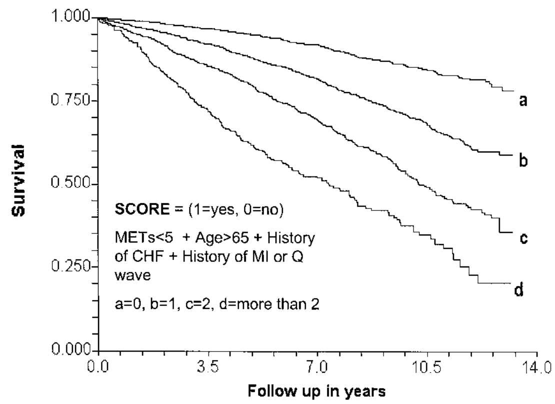

Fig 8. Kaplan-Meier survival curves for the "all-comers" prognostic score.

Predicting Prognosis in Patients With Resting ST Depression

Kwok et al demonstrated that the Duke Treadmill score can effectively risk stratify patients with ST-T abnormalities on the resting ECG.[76] When patients with ST-T abnormalities were classified into risk groups according to the Duke score, there were significant overall differences among the risk groups for all outcome endpoints. The 7-year event-free survival was 94%, 88%, and 69% for the low-, intermediate-, and high-risk groups, respectively. More patients with ST-T changes were classified as high risk (5% v 2%), and their 7-year survival was lower than that of the control population high-risk patients (76% v 93%).

Prognosis in "All-Comers" to the Exercise Laboratory

Previous prognostic studies focused on specific subsets of patients. Our group decided to analyze all patients referred for evaluation at their exercise laboratory between 1987 and 2000 to determine the prevalence of exercise test abnormalities.[77] There were 6,213 men (mean age 59 ± 11 years) who had standard exercise ECG treadmill tests over the study period with a mean 6-year follow-up. There were no complications of testing in this clinically referred population, 78% of whom were referred for chest pain, risk factors, or signs or symptoms of ischemic heart disease. Overlapping thirds had typical angina or history of myocardial infarction. A total of 579 had prior coronary artery bypass surgery, and 522 had a history of congestive heart failure. Indications for testing were in accordance with published guidelines. Twenty percent had died over the follow-up for an average annual mortality of 2.6%. Cox hazard function chose the following variables in rank order as independently and significantly associated with time to death: METs less than 5, age greater than 65, history of congestive heart failure, and history of myocardial infarction. A score based on simply adding these variables classified patients into low-, medium-and high-risk groups. The high-risk group (score of 3 or more) has a hazard ratio of 5 (4.7 to 5.3, 95% CI) and a 5-year mortality of 31% (Fig 8).

What do these findings mean to the clinician? First, it should be noted that all studies have population-specific attributes that may be difficult to define. Nevertheless, if the aim is to predict infarct-free survival, the Duke Treadmill score is preferred to the authors' because censoring was performed and infarct-free survival was predicted. If diagnosis is the issue, either the Duke score or other treadmill diagnostic scores are indicated. If diagnosis is known, prognostication using the Duke Treadmill score can help direct therapy. If diagnosis is not known and prognosis is guarded, then further diagnostic efforts may be indicated. If diagnosis is not determined and a patient is high risk by the score, then risk is likely to be improved by an exercise program and risk factor modification. If prognosis is favorable, perhaps diagnosis is not as important as alleviating symptoms. These findings strengthen the importance of exercise capacity, a reflection of the integrity of the cardiopulmonary system and a marker of physically active living, as an important predictor of survival along with or despite modern medical treatment.

The Elderly

The decline in function that accompanies aging is a consequence of age-related decrements in cardiovascular, pulmonary, and musculoskeletal structure. Ultimately, these result in impaired physical function in the elderly.[78] While the Duke Treadmill score was validated in patients in the age range when CAD first appears, data are limited in the elderly. To determine the prognostic value the treadmill test in the elderly, researchers from the Mayo Clinic and the Olmsted Medical Group compared the prognostic value of the test in patients less than 65 and older than 65 years of age.[79] Elderly (n = 514) and younger (n = 2,593) patients who underwent treadmill testing between 1987 and 1989 were identified retrospectively and followed up for 6 years. Compared with younger patients, elderly patients had more comorbid conditions, a higher prevalence of abnormal ST depression (28% *v* 9%), and achieved lower workloads (6.0 METs *v* 10.7 METs). A poor exercise capacity and angina during the exercise test were associated with future cardiac events. Exercise-induced ST depression did not carry significant value in the elderly and was associated with future cardiac events only in younger patients. An increase of 1 MET in the workload was associated with a 14% decrease in risk for a cardiac event in younger patients and with a 18% risk reduction among the elderly. After adjustment for clinical factors, there was a strong inverse association between exercise capacity and outcome. Workload achieved was the only treadmill exercise-testing variable that provided prognostic information for mortality and cardiac events. In the elderly, exercise capacity was also inversely associated with the likelihood of nursing home placement. In a similar but larger study in 2,079 male veterans, the authors validated their findings and showed that diagnostic scores predict angiographic disease in the elderly as well as in younger patients.[80]

Comparing Scores and Physicians

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 performed one of the first such studies.[81] 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 and et al validated 2 scores by comparing their diagnostic accuracy to that of cardiologists.[82] 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 considered scores for prognosis (rather than diagnosis) with 100 patients sent to 5 senior cardiologists at one center.[83] Again, the scores outperformed these cardiologists.

Our group performed a study that was larger and included different groups of physicians, validating these earlier studies that scores can predict angiographic results and prognosis as well as physicians.[84] A total of 599 consecutive male patients without prior myocardial infarction with a mean age of 59 ± 11 years were considered for this analysis. With angiographic disease defined as any coronary lumen occlusion equal or greater than 50%, 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 them as high, low, or intermediate probability of disease in addition to estimating a numerical probability from 0% to 100%. 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 109 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 (0.77 area under the ROC curve) to all the physician groups (0.69 for the experts, 0.65 for the cardiologists, and 0.66 for the internists: $P<.01$). Using a probability cut-point of greater than 70% for abnormal, predictive accuracy was 70% for the scores versus 64% for the experts, and 64% and 62.5% for the other physicians. In a subsequent analysis, the authors found the scores to predict prognosis as well or better than physicians.[85]

Conclusions

Physicians should not reduce their diagnostic assessments to blindly using and memorizing prediction rules. However, despite the methodological limitations of the available studies, the scores make possible better decisions. Statistical approaches cannot make counter-intuitive leaps of tangential thinking but excel at that which humans do not: *Considering vast quantities of information perfectly, then categorizing, analyzing it without bias and developing scores that make diagnoses*. Making use of statistics as described gives clinicians a powerful second opinion and allows them to concentrate on what the computer can never do: *Look after patients as individuals*. In particular, scores make available the experience of the specialist clinician to generalists. Generalists have to cover a wide range of specialties, and they cannot be equally up to date in each. The authors have shown that scores can, in certain cases, equal the diagnostic reasoning of specialist physicians. Making these "opinions" available to the generalist would allow resources to be concentrated on those who need it the most. Scores can help diagnose, thereby avoiding expensive unnecessary invasive investigations and their associated risk, and help with prediction of prognosis, allowing optimal use of secondary prevention measures. Since Laennec's invention of the stethoscope,[86] doctors have worked to develop tools to aid clinical assessment. In this technological age, clinical scores represent the natural extension of this historical tradition. The Duke prognostic score, the VA/West Virginia diagnostic scores, and the decrease of heart rate during recovery should be calculated as part of every exercise test.

References

1. Rosamond WD, Chambless LE, Folsom AR, et al: Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. N Engl J Med 339:861-887, 1998
2. Reeder GS, Gersh BJ: Modern management of acute myocardial infarction. Curr Prob Cardiol 21:590-662, 1996
3. Elveback LR, Connolly DC, Kurland LT: Coronary artery disease in residents of Rochester, Minnesota. II. Mortality, incidence, and survivorship, 1950-1975. Mayo Clin Proc 56:665-672, 1981
4. Swets JA, Dawes RM, Monahan J: Better decisions through science. Sci Am October:82-87, 2000
5. Randolph AG, Guyatt GH, Calvin JE et al: Understanding articles describing clinical prediction tools. Evidence based medicine in critical care group. Crit Care Med 26:1603-1612, 1998
6. Richardson WS, Wilson MC, Guyatt GH, et al: Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis. Evidence-Based Medicine Working Group. JAMA 281:1214-1219, 1999
7. Hlatky MA, Califf RM, Harrell FE Jr, et al: Clinical judgment and therapeutic decision making. J Am Coll Cardiol 15:1-14, 1990
8. Philbrick JT, Horowitz, Feinstein AR: Methodological problems of exercise testing for coronary artery disease: Groups, analysis and bias. Am J Cardiol 64: 1117-1122, 1989
9. Reid M, Lachs M, Feinstein A: Use of methodological standards in diagnostic test research. JAMA 274: 645-651, 1995
10. Guyatt GH: Readers' guide for articles evaluating diagnostic tests: What American College of Physicians Journal Club does for you and what you must do yourself. ACP J Club 115:A-16, 1991
11. Morise A, Duval R: Comparison of three Bayesian methods to estimate posttest probability in patients undergoing exercise stress testing. Am J Cardiol 64: 1117-1122, 1989
12. Ellestad MH, Savitz S, Bergdall D, et al: The false positive stress test. Multivariate analysis of 215 subjects with hemodynamic, angiographic and clinical data. Am J Cardiol; 40:681-687, 1977
13. Yamada H, Do D, Morise A, et al: Review of studies utilizing multi-variable analysis of clinical and exercise test data to predict angiographic coronary artery disease. Prog Cardiovasc Dis 39:457-481, 1997
14. Kansal S, Roitman D, Bradley EL, et al: Enhanced evaluation of treadmill tests by means of scoring based on multivariate analysis and its clinical application: A study of 608 patients. Am J Cardiol 52: 1155-1160, 1983
15. Fisher LD, Kennedy JW, Chaitman BR, et al: Diagnostic quantification of CASS (coronary artery surgery study) clinical and exercise test results in determining presence and extent of coronary artery disease. A multivariate approach. Circulation 63:987-1000, 1981
16. Hollenberg M, Budge WR, Wisneski JA, et al: Treadmill score quantifies electrocardiographic response to exercise and improves test accuracy and reproducibility. Circulation 61:276-285, 1980
17. Cohn K, Kamm B, Feteih N, et al: Use of treadmill score to quantify ischemic response and predict ex-

tent of coronary disease. Circulation 59:286-296, 1979

18. Berman JL, Wynne J, Cohn PF: A multivariate approach for interpreting treadmill exercise tests in coronary artery disease. Circulation 58:505-512, 1978
19. Laupacis A, Sekar N, Stiell IG: Clinical prediction rules: A review and suggested modifications of methodological standards. JAMA 277:488-494, 1997
20. Do D, Marcus R, Froelicher V, et al: Predicting severe angiographis coronary artery disease using computerization of clinical and exercise test data. Chest 114:1437-1445, 1998
21. Gibbons RJ, Balady GJ, Beasley JW, et al: ACC/AHA guidelines for exercise testing: Executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 260-311, 1997
22. Knottnerus A: Diagnostic prediction rules: Principles, requirements, and pitfalls. Prim Care 22:341-361, 1995
23. Poses RM, Cebul RD, Collins M, et al: The importance of disease prevalence in transporting clinical prediction rules. Ann Intern Med 105:586-591, 1986
24. Morise A, Haddad J, Beckner D: Development and validation of a clinical score to estimate the probability of coronary artery disease in men and women presenting with suspected coronary disease. Am J Med 102:350-356, 1997
25. Morise A: Comparison of the Diamond-Forrester method and a new score to estimate the pretest probability of coronary disease before exercise testing. Am Heart J 138:740-745, 1999
26. Detrano R, Bobbio M, Olson H, et al: Computer probability estimates of angiographic coronary artery disease: Transportability and comparison with cardiologist's estimates. Comp Bio Res 25:468-485, 1992
27. Morise AP, Detrano R, Bobbio M, et al: 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 20:1187-1196, 1992
28. Do D, West J, Morise A, et al: A consensus approach to diagnosing coronary artery disease based on clinical and exercise test data. Chest 11:1742-1749, 1997
29. Shiu P, Froelicher V: EXTRA: An expert system for exercise testing utilizing consensus to predict coronary disease. J Inv Card 2:21-26, 1998
30. Raxwal V, Shetler K, Morise A, F. Simple treadmill score to diagnose coronary disease. Chest 119:1933-1940, 2001
31. Fearon W, Lee D, Froelicher V: The effect of resting ST segment depression on the diagnostic characteristics of the exercise treadmill test. J Am Coll Cardiol 35:1206-1211, 2000
32. Gauri AJ, Raxwal VK, Roux L, et al: Effects of chronotropic incompetence and beta-blocker use on the exercise treadmill test in men. Am Heart J 142:136-141, 2001
33. Lee DP, Fearon WF, Froelicher VF: Clinical utility of the exercise ecg in patients with diabetes and chest pain. Chest 119:1576-1581, 2001
34. Gibbons R, Chatterjee K, Daley J, et al: ACC/AHA/ACP-ASIM Guidelines for the management of patients with chronic stable angina: Executive summary and recommendations. Circulation 99:2829-2848, 1999
35. Kotler TS, Diamond GA: Exercise thallium-201 scintigraphy in the diagnosis and prognosis of coronary artery disease. Ann Intern Med 113:684-702, 1990
36. Chatziioannou SN, Moore WH, Ford PV, et al: Prognostic value of myocardial perfusion imaging in patients with high exercise tolerance. Circulation 99:867-872, 1999
37. Hachamovitch R, Berman D, Kiat H, et al: Exercise myocardial perfusion SPECT in patients without known coronary artery disease. Incremental prognostic value and use in risk stratification. Circulation 93:905-914, 1996
38. Kuntz KM, Fleischmann KE, Hunink M, et al: Cost-effectiveness of diagnostic strategies for patients with chest pain. Ann Intern Med 130:709-718, 1999
39. Garber AM, Solomon NA: Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. Ann Intern Med 130:719-728, 1999
40. Kotler TS, Diamond GA: Exercise Thallium-201 scintigraphy in the diagnosis and prognosis of coronary artery disease. Ann Intern Med 113:684-702, 1990
41. Lachterman B, Lehmann KG, Neutel J, et al: Comparison of the ST/heart rate index to standard ST criteria for analysis of the exercise electrocardiogram. Circulation 82:44-50, 1990
42. Morise AP. Accuracy of heart rate-adjusted ST segments in populations with and without posttest referral bias. Am Heart J 134:647-655, 1997
43. Okin PM, Roman MJ, Schwartz JE, et al.: Relation of exercise-induced myocardial ischemia to cardiac and carotid structure. Hypertension 30:1382-1388, 1997
44. Viik J, Lehtinen K, Malmivuo J: Detection of coronary artery disease using maximum value of ST/HR hysteresis over different number of leads. J Electrcardiol 32(Suppl):70-75, 1999
45. Froelicher VF, Lehmann KG, Thomas R, et al: The electrocardiographic exercise test in a population with reduced workup bias: diagnostic performance, computerized interpretation, and multivariable prediction. QUEXTA. Ann Intern Med 128:965-974, 1998
46. Okin PM, Grandits G, Rautaharju PM, et al: Prognostic value of heart rate adjustment of exercise-induced ST segment depression in the multiple risk factor intervention trial. J Am Coll Cardiol 27:1437-1443, 1996
47. Okin PM, Anderson KM, Levy D, et al: Heart rate adjustment of exercise-induced ST segment depression. Improved risk stratification in the Framingham Offspring Study. Circulation 83:866-874, 1991
48. Livschitz S, Sharabi Y, Yushin J, et al: Limited clinical value of exercise stress test for the screening of coronary artery disease in young, asymptomatic adult men. Am J Cardiol 86:462-464, 2000
49. Michaelides A, Psomadaki Z, Dilaveris P, et al: Im-

proving detection of coronary artery disease by exercise electrocardiography with the use of right precordial leads. N Engl J Med 340:340-345, 1999
50. Bokhari S, Blood DK, Bergmann SR: Use of right precordial leads during exercise testing. N Engl J Med 343:968-969, 2000
51. Hlatky MA, Califf RM, Harrell FE, et al: Clinical judgment and therapeutic decision making. J Am Coll Cardiol 15:1-14, 1990
52. Shaw LJ, Peterson ED, Shaw LK, et al: Use of a prognostic treadmill score in identifying diagnostic coronary disease subgroups. Circulation 98:1622-1630, 1998
53. Lauer MS, Blackstone E, Young J, et al: Cause of death in clinical research: Time for a reassessment? J Am Coll Cardiol 34:618-620, 1999
54. Mark DB, Shaw L, Harrell FE Jr, et al: Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. N Engl J Med 325:849-853, 1991
55. Froelicher VF, Morrow K, Brown M, et al: Prediction of artherosclerotic cardiovascular death in men using a prognostic score. Am J Cardiol 73:133-138, 1994
56. Redelmeier DA, Lustig AJ: Prognostic indices in clinical practice. JAMA 285:3024-3025, 2001
57. Gohlke H, Samek L, Betz P, et al: Exercise testing provides additional prognostic information in angiographically defined subgroups of patients with coronary artery disease. Circulation 68:979-985, 1983
58. Brunelli C, Cristofani R, L'Abbate A, for the ODI Study Group: Long-term survival in medically treated patients with ischemic heart disease and prognostic importance of clinical and electrocardiographic data. (The Italian CNR Multicenter Prospective Study OD1). Eur Heart J 10:292-303, 1989
59. Weiner DA, Ryan T, McCabe CH, et al: Prognostic importance of a clinical profile and exercise test in medically treated patients with coronary artery disease. J Am Coll Cardiol 3:772-779, 1984
60. Peduzzi P, Hultgren H, Thomsen J, et al: Prognostic value of baseline exercise tests. Progress Cardiovasc Dis 28:285-292, 1986
61. Lerman J, Svetlize H, Capris T, et al: Follow-up of patients after exercise test and catheterization. Medicina (Buenos Aires) 46:201-211, 1986
62. Wyns W, Musschaert-Beauthier E, Van Domburg R, et al: Progostic value of symptom limited exercise testing in men with a high prevalence of coronary artery disease. Eur Heart J 6:939-945, 1985
63. Detry JM, Luwaert J, Melin J, et al: Non-invasive data provide independent prognostic information in patients with chest pain without previous myocardial infarction: Findings in male patients who have had cardiac catheterization. Eur Heart J 9:418-426, 1988
64. Froelicher VF, Myers J: Exercise and the Heart (ed 4). Philadelphia, Saunders-Mosby, 1999
65. Morrow K, Morris C, Froelicher V, et al: Prediction of cardiovascular death in men undergoing noninvasive evaluation for coronary artery disease. Ann Intern Med 118:689-695, 1993
66. Froelicher V, Morow K, Brown M, et al: Prediction of atherosclerotic cardiovascular death in men using a prognostic score. Am J Cardiol 73:133-138, 1994
67. Mark D, Hlatky M, Harrell F, et al: Exercise treadmill score for predicting prognosis in coronary artery disease. Ann Intern Med 106:793-800, 1987
68. Gibbons R, Chatterjee K, Daley J, et al: ACC/AHA/ACP-ASIM Guidelines for the management of patients with chronic stable angina: Executive summary and recommendations. Circulation 99:2829-2848, 1999
69. Gibbons RJ, Hodge DO, Berman DS, et al: Long-term outcome of patients with intermediate-risk exercise electrocardiograms who do not have myocardial perfusion defects on radionuclide imaging. Circulation 100:2140-2145, 1999
70. Iscandrian AS, Ghods M, Helfeld H, et al: The treadmill exercise score revised: Coronary arteriographic and thallium perfusion correlates. Am Heart J 124:1581-1586, 1992
71. Cole CR, Blackstone EH, Pashkow FJ, et al: Heart-rate recovery immediately after exercise as a predictor of mortality. N Engl J Med 341:1351-1357, 1999
72. Cole CR, Foody JM, Blackstone EH, et al: Heart rate recovery after submaximal exercise testing as a predictor of mortality in a cardiovascularly healthy cohort. Ann Intern Med 132:552-555, 2000
73. Nishime EO, Cole CR, Blackstone EH, et al.: Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise ECG. JAMA 284:1392-1398, 2000
73a. Watanabe J, Thamilarasan M, Blackstone EH, et al. Heart rate recovery immediately after treadmill exercise and left ventricular systolic dysfunction as predictors of mortality: the case of stress echocardiography. Circulation 104:1911-1916, 2001
74. Sheller K, Marcus R, Froelicher V, et al: Heart rate recovery validation and methodologic issues. J Am Coll Card 38:1980-7, 2001
75. Alexander K, Shaw L, Delong E, et al: Value of exercise treadmill testing in women. J Am Coll Cardiol 32:1657-1664, 1998
76. Kwok JM, Miller TD, Christian TF, et al: Prognostic value of a treadmill exercise score in symptomatic patients with nonspecific ST-T abnormalities on resting ECG. JAMA 282:1047-1053, 1999
77. Prakash M, Froelicher V, Myers J, et al: Clinical and exercise test predictors of all-cause mortality: Results from over 6,000 consecutive referred male patients Chest 120:1003-1013, 2001
78. Brechue W, Pollock M: Exercise training for coronary artery disease in the elderly. Clin Geriatr Med 12:207-229, 1996
79. Goraya T, Jacobsen S, Pellikka P, et al: Prognostic value of treadmill exercise testing in elderly persons. Ann Intern Med 132:862-870, 2000
80. Spin J, Prakash M, Froelicher V, et al: Exercise testing in elderly males: Prognostic and diagnostic value. Am J Med 112;121-126, 2002
81. 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 25:468-485, 1992

82. Hlatky M, Botvinick E, Brundage B: Diagnostic accuracy of cardiologists compared with probability calculations using Bayes' rule. Am J Cardiol 49:1927-1931, 1982

83. Lee KL, Pryor DB, Harrell FE Jr, et al: Predicting outcome in coronary disease. Statistical models versus expert clinicians. Am J Med 80:553-560, 1986

84. Lipinski M, Froelicher V, Franklin B, et al: Can scores predict disease as well as physicians? Arch Internal Medicine 2001, 161:2239-2244

85. Lipinski M, Froelicher V, Franklin B, et al: Can scores predict severe disease and prognosis as well as physicians? Am Heart J 2002, 143:58-64

86. Kandela P: The stethoscope. Lancet 352:997, 1998