Use of Treadmill Score to Quantify Ischemic Response and Predict Extent of Coronary Disease

KEITH COHN, M.D., BARBARA KAMM, M.D., NIZAR FETEIH, M.D., RICHARD BRAND, PH.D., AND NORA GOLDSCHLAGER, M.D.

SUMMARY. In this study we assessed whether various responses to exercise testing could be quantified in order to derive the probabilities of presence of coronary disease, and if present, to assess its severity. A treadmill score based on the exercise response was determined in 405 patients who had both treadmill tests and coronary angiograms. The score was derived using discriminant function analysis, by weighting and combining depth and configuration of ST depression (downsloping, horizontal or slowly upsloping), timing onset and duration of ischemia, grading ventricular arrhythmias, heart rate and blood pressure change, coexistence of exercise-induced chest pain and sex. The treadmill score was effective in detecting coronary disease (lesions with an \( \geq 50\% \) narrowing), with a predictive accuracy (PA) (probability that a subject manifesting a positive test has disease) of 87\%, a true negative rate (TNR) (probability of a subject with a negative test having no disease) of 80\%, and sensitivity of 94\%. The treadmill score also detected severe disease (triple-vessel, main left and/or \( > 90\% \) proximal occlusion of the left anterior descending artery), with a PA of 73\%, TNR of 79\% and sensitivity of 82\%.

We conclude that the exercise response, expressed numerically as a treadmill score, permits analysis of most of the relevant data from exercise testing, increases test accuracy by 10–15\% compared with standard criteria for treadmill test interpretation, and enables the derivation of probability statements for presence and severity of coronary disease. The validity of any prediction on the basis of exercise performance may thus be quantitatively judged.

UNTIL RECENTLY, INTERPRETATION of exercise tests was predominantly qualitative, with the aim of determining the presence or absence of myocardial ischemia, and, by inference, coronary artery disease (CAD). Recent studies have shown: 1) test insensitivity could be explained predominantly by extent of disease (patients with single-vessel involvement often having normal stress tests); and 2) treadmill responses reflecting more severe CAD include: ST segments depressed by at least 1 mm, and downsloping, early onset and prolonged duration of ischemic ST segments (longer than 8 minutes), marked depth of ischemic ST depression, hypotension during modest levels of exertion, inappropriately low heart rate increments during exercise ("chronotropic incompetence"), and malignant ventricular arrhythmias, especially at low heart rates or associated with electrocardiographic evidence of myocardial ischemia. Based on these observations, we conclude that severity of CAD is a principal determinant of severity of myocardial ischemia, and this in turn is reflected in the treadmill exercise response.

In this study we explored the quantification of the exercise test response. We used a treadmill score, encompassing all variables considered to reflect severity of ischemic response, to predict the presence and severity of CAD. We also derived a series of probability statements to estimate the likelihood of coronary disease, given a particular exercise test response.

Methods

Patient Sampling

Four hundred five patients evaluated consecutively (except for those excluded as described below) at Presbyterian Hospital of Pacific Medical Center with both treadmill exercise tests and selective coronary angiograms were studied. Two hundred ninety-one of the patients had either coronary disease or typical angina; and 114 were studied with the explicit purpose of proving or disproving the diagnosis of CAD, the latter subgroup with symptoms ranging from pain atypical of angina to pain which was at least partially typical. None of the subjects were taking digitalis or propranolol.

Coronary Angiography

 Coronary angiograms were made using standard Judkins or Sones techniques, and were interpreted by two independent observers; interobserver differences were resolved by an independent arbiter. Only narrowing \( \geq 50\% \) of the luminal diameter was considered. The validity of using this degree of narrowing is supported by observations that coronary angiograms often underestimate the severity of stenosis, and by reports documenting a correlation between this degree of coronary narrowing and life expectancy. One hundred fifteen patients (28\%) had absent or insignificant CAD, 65 (16\%) had single-, 92 (23\%) had double- and 131 (32\%) had triple-vessel disease.
Classification of Severity of Coronary Disease

The extent of CAD having been determined, each patient was placed in one or more of the following non-exclusive (overlapping) categories:
1) absent or insignificant CAD, with no lesion > 50% luminal diameter narrowing;
2) CAD — single-, double-, or triple-vessel involvement;
3) moderate to severe degree of CAD — either double- or triple-vessel, or main left involvement;
4) advanced CAD — triple-vessel and/or left main coronary disease, or triple-vessel, main left and/or critically narrowed (≥ 90%) proximal left anterior descending artery disease.

Only 23 patients had left main coronary disease, and 19 had coexistent double- or triple-vessel involvement.

Exercise Test Protocol

Using continuous electrocardiographic monitoring during and after the stress tests, patients were exercised on a motor-driven treadmill using a right shoulder to V5, positioned bipolar lead,3 with a modified Bruce protocol. Exercise was continued until 1 mm or more horizontal or downsloping ischemic ST-segment changes appeared, and for 1–3 minutes thereafter (to insure the evoked abnormalities were persistent and not artifactual); or until the patient achieved least 90% of the age-predicted maximum heart rate. Patients were excluded from the study if 1) they had ST-T abnormalities at rest or during a pre-exercise period of hyperventilation with either 0.5 mm ischemic-appearing ST depression or T inversion;12 2) they had been taking digitalis within a 10-day period; and 3) abnormalities of ventricular activation, such as left bundle-branch block, Wolff-Parkinson-White, or left ventricular hypertrophy were present. These conditions enhance the probability of false positive results and interfere with test interpretation.10 Patients were also excluded if 90% of the target heart rate was not achieved and the test remained negative, to eliminate false negative responses due to inadequate degrees of exercise stress.

Treadmill Response

The following variables recorded during and after the exercise test were used to develop a treadmill score:
1) maximum depth of ST depression at the J point (mm), with onset of the QRS complex as baseline;
2) depth of the ST segment 80 msec after the J point (mm);
3) configuration of ST-segment depression, graded by severity:
   1 = normal ST-segment response (including typical J depression with ST segments rapidly returning to baseline);
   2 = 1 mm or greater J point depression with slowly upsloping ST segments, as previously defined;4
   3 = 1 mm or greater J point depression with horizontal ST segments;
   4 = 1 mm or greater J point depression with downsloping ST segments;
When more than one type of ST response occurred, the assigned grade reflected the most advanced abnormality.5
4) timing of onset of ischemia (the duration of exercise in minutes at onset of ischemia);
5) timing of onset of ischemia (beats/min). (Both 4 and 5 were tabulated when horizontal or downsloping ST-segment changes first appeared; see appendix A);
6) duration of ischemic changes during the recovery phase (minutes);
7) ventricular arrhythmias, graded by presence and severity:
   0 = absence of arrhythmia;
   1 = unifocal ventricular ectopic beats, with frequency less than one per five QRS complexes;
   2 = unifocal ventricular premature beats with frequency greater than one per five QRS complexes;
   3 = multiform ventricular ectopic beats;
   4 = pairs of ventricular ectopic beats;
   5 = salvos or runs of ventricular tachycardia;
   6 = ventricular fibrillation.
8) systolic blood pressure change (mm Hg), representing the change between the last two stages of exercise; a fall in systolic pressure was recorded as a negative value.
9) heart rate change (beats/min/stage), recorded as the difference between the heart rate at rest and at termination of exercise divided by the number of stages of exercise completed, and representing an attempt to quantify the appropriateness of the heart rate increment for the exercise performed.
10) exercise-induced chest pain or discomfort:
   0 = absence of pain;
   1 = presence of pain;
   2 = concomitance of chest pain and ischemic ST depression.
   11) Sex:
      0 = male;
      1 = female.
   12) Age (years).

Derivation of Treadmill Score:
Multivariate Linear Discriminant Function Analysis

Multivariate linear discriminant function analysis is a statistical means of distinguishing two or more groups based on collection of discriminating variables.14–17 In our study, the groups were formed by the classification of severity of coronary disease (dependent variable). The discriminating (independent) variables consisted of the treadmill responses, and were combined to form a single treadmill score:

\[
\text{Treadmill Score} = C_1 V_1 + C_2 V_2 + \ldots + C_n V_n
\]
Where \( C_i = \text{Coefficient} \)
\( V_i = \text{Variable} \)

The independent variables were selected in a stepwise manner for entry into the analysis on the basis of their
**Figure 1. Use of treadmill score to predict presence or absence of coronary disease (one- to three-vessel involvement).** This graph is a traced computer plot of treadmill score against sensitivity, specificity, predictive accuracy, true negative rate, and posterior probability — all expressed as percentages. The prevalence of coronary disease in the population studied is 72%. The intercept of Z on the x-axis defines the treadmill score (−0.81, in this instance), which serves as the optimal value dividing normality (all subjects with a score falling at or below Z) and abnormality (subjects with greater than Z). Note the differences between the way posterior probability and the other measures of test performance (sensitivity (SENS), specificity (SPEC), predictive accuracy (PA), and true negative rate (TNR)) are used in this graph.

The posterior probability (PP) at any treadmill score defines the likelihood of disease being present; once an exercise test is run and a treadmill score defined, one simply extrapolates the score vertically to the PP curve, delineating the likelihood of that subject having disease. The greater the score, the greater the probability of disease being present. Thus, the PP changes with each subject, depending on the magnitude of the treadmill score. (The 50% PP point also specifies the optimum criterion, Z, which maximizes the number of correct diagnoses.)

In contrast, the SENS, SPEC, PA and TNR plots reflect statements of overall test accuracy, assuming a particular criterion, Z, is chosen to separate normal from abnormal responders. In this manner, all subjects with scores falling above Z are classified abnormal whether the score is slightly greater or considerably in excess of Z), and the test accuracy — defined in terms of SENS and PA for all such subjects are determined by the intercept of Z on their respective curves. Similarly, all patients with scores under Z are classified as normal, the TNR and SPEC for all such individuals being defined again by the Z intercept. Slight jaggedness was smoothed algebraically.

discriminating power, and the weighted coefficient, C, was calculated with established statistical methods. Standardized coefficients were also derived which normalize for differences in units of measurements; the absolute magnitude of the standardized coefficient provides an index of the relative strength of each variable in separating the patient groups. Although there is potential bias using the same observations to calculate a function and then evaluate it, Lachenbruch's data indicates that this poses little problem for moderate-to-large samples.

**Classification of Patients**

**Into Coronary Disease Categories**

The use of discriminant analysis as a classification technique begins after these initial computations. A treadmill score is calculated for each patient; these scores are then ranked from highest to lowest. A "cutpoint" (Z) is chosen as the criterion to divide the continuum of ranked patients in two. Patients with scores greater than Z are classified as having CAD; those with scores ≤ Z are classified as free of CAD. This criterion may be arbitrarily set anywhere between the highest (most abnormal) and lowest (most normal) treadmill score. The customary dividing point is that which minimizes the probability of misclassification.

The following test parameters may all be determined after a cut-point has been chosen:

- True and false positive treadmill diagnoses = TP and FP
- True and false negative treadmill diagnoses = TN and FN

**Sensitivity** — The percentage of patients with disease* having an abnormal treadmill test:

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

**Specificity** — The percentage of patients with no disease having a normal treadmill test:

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

**Predictive accuracy** (PA), also termed predictive

*The term "disease" is used variably in these definitions, depending on whether the treadmill score is being used to predict the presence or absence of CAD, moderate-to-severe CAD, or advanced CAD.
value of a positive test — The percent of patients with an abnormal treadmill test having disease:

\[
\frac{TP}{TP + FP}
\]

**True negative rate** (TNR), also known as PA or predictive value of a negative test — The percent of patients with a normal treadmill test having no disease:

\[
\frac{TN}{TN + FN}
\]

**Predictive error** — Probability that a patient with a normal test has disease.

**Correct Classification Rate** — The percent of subjects correctly classified:

\[
\frac{TP + TN}{NO.}
\]

(where NO. = total number of patients)

**Posterior Probability** (PP) — This may also be termed post-test probability. The probability that a person has disease, given his or her treadmill score17 (see Results).

**Risk and Odds Ratio** — Indices of the proportionate likelihood that a patient has disease, given an abnormal test, relative to the likelihood of a patient having disease, given a normal exercise test.

Risk Ratio: \( \frac{\text{Predictive Accuracy}}{\text{Predictive Error}} \)

Odds Ratio: \( \frac{TP \times TN}{FP \times FN} \)

See appendix B for extensions of these formulas.

The results were calculated according to the prevalence of coronary disease in our study population, as well as for an assumed prevalence of coronary disease in a standard population; for purposes of calculation, sensitivity and specificity were considered to be fixed, as yet an unproved assumption. The prevalence of disease in an asymptomatic, presumably healthy adult population was approximated to be 0.20 for coronary narrowing involving one to three vessels, 0.16 for two- to three-vessel disease and 0.12 for severe disease (appendix C).

Results

Figure 1 is a traced computer plot of sensitivity and specificity (y-axis), graphed against treadmill score (x-axis). The lower or more negative the treadmill score (the further to the left it falls), the more normal the exercise test response; and the higher and more positive the score, the more abnormal the test. A treadmill score, or dividing line, (Z), separating normal from abnormal responders is chosen before the test is performed. Patients who have a score above that level are considered positive (having coronary disease) and those with a score below are considered negative (free of disease). The magnitude of sensitivity and specificity depend on the criterion chosen (cut-point Z). The intercept of Z on the sensitivity/specificity plots provides quantitative statements as to percentage of correct and incorrect diagnoses. Computer plots of PA and TNR may be graphed against treadmill score in similar fashion, and may be superimposed upon the graphs expressing sensitivity and specificity (fig. 1).

**Predicting Presence of Coronary Disease** —

**One- to Three-Vessel Involvement**

In the case shown in figure 1, a treadmill score of −0.81 is the optimal dividing point. This criterion maximizes the correct classification rate (percentage of correct diagnoses) at 85%. The intercept of the vertical line Z provides a quantification of sensitivity and specificity of 94% and 65%, respectively. With the same criterion of abnormality (a treadmill score Z of −0.81), a PA of 87% is achieved, with a TNR of 80% (fig. 1) (table 1).

**Predicting Moderate or Advanced Degrees of Coronary Narrowing**

In predicting the presence of moderate-to-severe CAD (two- to three-vessel and/or main left disease), 85% sensitivity is achieved, 77% specificity, 83% PA, 79% TNR, and 81% correct classification rate. In assessing the likelihood of advanced degrees of CAD, triple-vessel disease, left main disease, or critical left anterior descending obstruction, sensitivity falls to 83%, with PA of 73%, TNR of 84%, and correct classification rate of 78% (figs. 2–4) (table 1).
Posterior Probability

The PP plot has three uses:

**Predicting the Likelihood of Disease**

The plot of PP against treadmill score is a direct statement of the likelihood that a given patient has disease, once the test has been run and a treadmill score determined. If a test score is reasonably high, e.g., $+2.0$, the probability that the patient has disease is over 90%; however if the test score is low, e.g., $-2.6$, the probability of the patient having coronary disease is only approximately 10% (fig. 1). The PP may be used in this manner when the extremes of test score are achieved, but interpretation is questionable when the PP is around 50% (range of 35–65%).

**Defining Test Criteria**

A PP of 50% defines the optimal test score ($Z$), for dividing normality from abnormality, before a test is run. This cut-point maximizes the yield of correct diagnoses, and minimizes false positive and false negative results.

**Cost/Benefit Analysis (appendix D)**

Different errors in classification may have different costs. It may be a more serious error to fail to detect an advanced lesion than to erroneously suspect a healthy patient of having it. If one makes a prejudgment to minimize failure of detection of patients with advanced disease (i.e., to have a false positive rate in excess of the false negative rate), the criterion which separates normality from abnormality may be liberalized ($Z'$, fig. 2). In such instances, a normal exercise test response (score below the newly defined cut-point $Z'$) is unlikely to represent a false negative
response. In figure 2, with the decision to lower the probability of failing to detect a patient with double- or triple-vessel or left main disease, the false positive rate was set at six times the false negative rate. Under this circumstance the TNR has been increased to 88% — only 12% false negative responses — with concomitant reduction in PA a positive test to 77%.

Potency of Each Exercise Test Variable in Predicting Disease

Most of the independent variables generated by the exercise test — depth and types of ST-segment depression, duration of ischemic changes, onset of ischemia, sex, and presence of chest pain — showed significant mean differences between disease vs non-disease categories (table 2). Only the magnitude of heart rate increment during exercise and the change in blood pressure did not distinguish presence or extent of CAD. Not all variables, however, are included in the treadmill score. The discriminant procedure selects variables in a stepwise fashion until the maximum information about group differences is obtained. As many of the exercise test variables are highly correlated, a reduced set of variables provides full discriminating power (table 3).

The absolute magnitude of the standardized coefficient is an indicator of the relative strength of each variable in discriminating between disease groups (table 3). Configuration and depth of ST depression and sex of the subject were the most influential variables in predicting presence and severity of disease. Timing of onset and duration of ischemic ST-segment changes, and the various grades of ventricular arrhythmias served as additional, but not as highly influential, predictors of severe degrees of coronary narrowing.

Moreover, a 10–15% increment in sensitivity, PA, and/or correct classification rate is achieved using the treadmill score compared with the results of exercise testing achieved from the standard criteria (1 mm or greater flat or downsloping ST depression), a statistically significant improvement (McNemer test, $p < 0.001$).

Influence of Disease Prevalence

The above results are based on patients evaluated at our hospital, that is, subjects referred to a cardiology center with proven, probable or suspected CAD. They constitute a population in which high prevalence of CAD exists relative to the general population. PA,
TABLE 2.  Treadmill Exercise Test Variables: Mean = sd and Two Sample t Test p values

<table>
<thead>
<tr>
<th>Grouping of coronary disease by angiogram</th>
<th>Any degree†</th>
<th>Moderate-Severe†</th>
<th>Severe†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Absent</td>
<td>p</td>
<td>Present</td>
</tr>
<tr>
<td><strong>Maximum ST (J) (mm)</strong></td>
<td>2.11</td>
<td>1.39</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±1.27)</td>
<td>(±0.99)</td>
<td></td>
<td>(±1.21)</td>
</tr>
<tr>
<td><strong>Maximum ST (80 msec) (mm)</strong></td>
<td>2.01</td>
<td>0.41</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±1.37)</td>
<td>(±0.80)</td>
<td></td>
<td>(±1.30)</td>
</tr>
<tr>
<td><strong>Configuration of ST (1-4)</strong></td>
<td>3.01</td>
<td>1.47</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±1.11)</td>
<td>(±0.90)</td>
<td></td>
<td>(±0.96)</td>
</tr>
<tr>
<td><strong>Onset of ischemia (min.)</strong></td>
<td>7.05</td>
<td>11.14</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±4.08)</td>
<td>(±2.32)</td>
<td></td>
<td>(±3.85)</td>
</tr>
<tr>
<td><strong>Onset of ischemia (HR)</strong></td>
<td>144.99</td>
<td>190.23</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±39.26)</td>
<td>(±22.69)</td>
<td></td>
<td>(±35.88)</td>
</tr>
<tr>
<td><strong>Duration of ischemia (min.)</strong></td>
<td>3.83</td>
<td>1.54</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±6.94)</td>
<td>(±9.50)</td>
<td></td>
<td>(±7.48)</td>
</tr>
<tr>
<td><strong>Ventricular arrhythmias (0-5)</strong></td>
<td>0.49</td>
<td>0.38</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±1.05)</td>
<td>(±0.82)</td>
<td></td>
<td>(±1.15)</td>
</tr>
<tr>
<td><strong>Chest pain (0-2)</strong></td>
<td>0.79</td>
<td>0.39</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±0.80)</td>
<td>(±0.59)</td>
<td></td>
<td>(±0.81)</td>
</tr>
<tr>
<td><strong>Systolic pressure change (mm Hg)</strong></td>
<td>11.82</td>
<td>10.73</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±15.58)</td>
<td>(±13.85)</td>
<td></td>
<td>(±16.43)</td>
</tr>
<tr>
<td><strong>Heart rate change (beats/min/stage)</strong></td>
<td>27.34</td>
<td>28.32</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±13.38)</td>
<td>(±10.10)</td>
<td></td>
<td>(±14.56)</td>
</tr>
<tr>
<td><strong>Sex (0, 1)</strong></td>
<td>0.08</td>
<td>0.40</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±0.28)</td>
<td>(±0.49)</td>
<td></td>
<td>(±0.28)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>51.39</td>
<td>48.08</td>
<td>0.000*</td>
</tr>
<tr>
<td>(±8.76)</td>
<td>(±8.85)</td>
<td></td>
<td>(±8.35)</td>
</tr>
</tbody>
</table>

*Indicates statistical significance.
†Group of coronary disease is identical with that given in table 1.

TNR and PP are governed by disease prevalence as well as by sensitivity and specificity of a test. In evaluating a group of patients with a lower prevalence of CAD, as would exist, for example, in healthy young to middle-aged adults, the magnitude of these three variables would have to be re-evaluated, based on the formulas described in appendix B. (Any calculation based on a disease prevalence different from that encountered in the sample population should be considered an approximation.) Figure 5 illustrates that in a population with disease prevalence of 20% (using the cut-point previously established in fig. 1), the TNR

TABLE 3. Multiple Discriminant Function Coefficients for Treadmill Score: Prediction of Presence and Severity of Coronary Disease

<table>
<thead>
<tr>
<th>Coronary disease* (One- to three-vessels)</th>
<th>Moderate-Severe disease (Two- or three-vessel and/or left main)</th>
<th>Severe disease (Three-vessel and/or left main)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>Unstandardized</td>
<td>Standardized</td>
</tr>
<tr>
<td>Maximum ST depression (J) (mm)</td>
<td>-0.452</td>
<td>-0.560</td>
</tr>
<tr>
<td>Maximum ST depression (80 msec) (mm)</td>
<td>0.605</td>
<td>0.865</td>
</tr>
<tr>
<td>Configuration of ST (1, 2, 3, 4)</td>
<td>0.340</td>
<td>0.428</td>
</tr>
<tr>
<td>Onset of ischemia (min.)</td>
<td>-0.021</td>
<td>-0.162</td>
</tr>
<tr>
<td>Duration of ischemia (min.)</td>
<td>-0.201</td>
<td>-0.162</td>
</tr>
<tr>
<td>Ventricular arrhythmias (0, 1, 2, 3, 4, 5)</td>
<td>-1.082</td>
<td>-0.412</td>
</tr>
</tbody>
</table>

*Each variable is multiplied by the “unstandardized” coefficient. The “standardized” coefficients represent values normalized for differences in units of measurement, and provide an index of the relative strength of each variable in separating the patient groups.
remains very high, over 90%; it is likely, therefore, that a normal test correctly represents nondisease. The PA, however, is considerably lower, and the likelihood that an abnormal test correctly identifies disease is 30–50% (fig. 5). Hence, in a population with lower disease prevalence, it is possible that an ischemic exercise response represented a false positive result.

Discussion

Quantitative information derived from treadmill testing may be used to assess the extent of myocardial ischemia and, indirectly, severity of CAD. Several types of exercise test abnormality have been associated with subsequent morbidity and mortality coronary events, and with angiographically determined extent of coronary artery narrowing. In this study we have used quantitative aspects of treadmill tests, weighting each of the test variables and combining them into a single treadmill score. The technique by which this is accomplished, multivariate linear discriminant function analysis, is a statistical method which maximizes the separation of patients into different categories based on treadmill response. This technique is analogous to decision-making in general. When several variables are involved in a clinical judgment, a physician attaches a subjective weight to each of the aspects of clinical information. The weight and sum of these clinical variables indicates the diagnostic probabilities. Discriminant analysis represents a formal analog to the human brain's thought processes. Variables that achieve significance in predicting the likelihood of disease are chosen, weighted and combined mathematically as a test score. The score is used to define the optimal criterion that divides normality from abnormality or to derive a probability statement of the likelihood that, given a particular score, disease of a given severity is present.

The benefits of such an approach include: 1) most or all of the relevant data from the exercise test response is incorporated in the test interpretation; 2) increased test sensitivity, PA and correct classification rate (of 10–15%) are achieved with treadmill score compared with standard criteria for treadmill test interpretation; 3) the response to exercise is presented as a spectrum, with the scores ranging from normal to most abnormal, in contrast to the traditional, qualitative interpretation of positive or negative based simply on a fixed criterion of abnormality (−1 mV or greater
"ischemic" ST depression); the quantitated exercise test response may then be considered, by extrapolation, to approximate a quantification of the severity of myocardial ischemia; 4) the validity of any prediction made on the basis of exercise performance similarly may be quantitatively assessed; 5) cost functions may be assigned, prejudging the direction in which one prefers to err, since the criterion of abnormality may be set at different levels.

**Sensitivity, Specificity and Correct Classification Rate**

Sensitivity, specificity and correct classification rate describe the diagnostic accuracy of the exercise test. The level of sensitivity and correct classification in the present study of 94% and 85%, respectively, are improvements over previously quoted figures. The inverse relationship between sensitivity and specificity is reflected in the 65% level of specificity, although this is not clinically important compared with high PA, TNR, and correct classification rate (87%, 80% and 85%, respectively). Sensitivity and specificity of an exercise test are determined by three principal factors:

1. The accuracy of the test: were a less informative test to be used, such as the submaximal Master Two Step, the sensitivity would fall to approximately 30%.

2. The severity of disease: mild coronary disease is undetected by stress tests in a large proportion of cases.

3. The criterion ("cut-point") to separate normal from abnormal test results: the reciprocal relationship between sensitivity and specificity when these are plotted against treadmill score exemplifies this concept. The more rigid the criterion for normality (the higher the score required to classify a patient as having disease), the more likely a proportion of patients with coronary disease will be erroneously classified as normal, the higher specificity being achieved at the expense of reduced sensitivity. Alternatively, the more liberal the criteria of abnormality, the lower the specificity and higher the sensitivity and TNR.

This inverse relationship between sensitivity and specificity is almost universal among test procedures. Therefore, in practice a compromise is made so that a criterion is chosen which sets sensitivity and specificity at levels that minimize false positive and false negative diagnoses. Were the decision made to decrease the probability of missing severe coronary disease, the criterion of abnormality might be liberalized (cost function), and the TNR would increase substantially so that there remained a lower probability that moderate-to-severe disease would go undetected (fig. 2). This results, however, in a concomitant decline in PA.

**Predictive Accuracy and True Negative Rate**

Although sensitivity and specificity allow quantification of overall accuracy of a procedure, they are not very useful in the interpretation of an individual test, as — having predetermined whether or not a patient has angiographically documented coronary disease — they define the probable treadmill test response. These parameters, therefore, view a test in an opposite manner to that in which it is used. To predict correctly disease based on a given exercise test result, one must use PA, TNR and PP, beginning with the interpretation of a treadmill test response and then extrapolating from it the probability of correctly recognizing presence and severity of coronary disease. The higher the PA or TNR, the more likely a positive or negative result is correct.

In this study, CAD was correctly recognized with a PA of 87% and TNR of 80%, but when predicting advanced degrees of coronary narrowing, sensitivity fell to 83%, PA to 73%, TNR to 83%, and correct classification rate to 78% (table 1). Thus, when the aim of a test is to identify advanced degrees of CAD compared with simple presence vs absence of CAD, the predictive strength of the test is lessened.

**Conversions of Pre-Test Probability to Post-Test Probability: Influence of Disease Prevalence**

PA, TNR and PP depend partially on the prevalence of disease in the subpopulation of individuals investigated. This means that any conclusion based on a test result — treadmill testing representing one example of this principle — must consider the clinical setting surrounding the patient — for example, typicality or atypicality of angina, age, asymptomatic state, etc. This setting may then be translated into a pre-test probability (prior probability) — the likelihood that the patient has disease before the test. Prior probability may thus be considered to be the prevalence of disease in the subpopulation being tested. The object of the test is to convert a pre-test probability into a revised post-test probability, the likelihood of disease being present once the test has been completed. The PP, PA and TNR may be considered different formats for representing this post-test probability (appendix C).

Using treadmill testing to screen asymptomatic, apparently normal individuals, the lower disease prevalence (assumed to lie between 10–20%) would result in a fall in PA (to 30–35% in this series); conversely, the TNR achieves a level of greater than 95%. Hence, a negative result suggests with a higher likelihood that CAD is absent.

**Potency of Each of the Independent Exercise Test Variables**

Configuration of depressed ST segments (whether they are flat, downsloping or non-ischemic), and the ST depth (at 80 msec) constituted those variables most influential in predicting CAD. Since the sign of the J and 80 msec ST differ in the dominant function, their algebraic difference provides a "hidden" index of slope incorporated into the treadmill score. Additional variables predictive of severity of coronary narrowing, though less important, included early onset of
ischemic ST-segment abnormalities and the more severe forms of exercise-induced ventricular ectopic activity, confirming prior observations. The sex of the subject being tested also achieved predictive significance, probably because coronary disease is much less prevalent among females.

The degree of blood pressure change with exercise, inappropriate increment in heart rate, and development of chest pain during the exercise test were not very important in contributing to the discriminant score. We did not observe exercise-related hypotension at moderate levels of exertion, occurring in association with anginal pain and/or ischemic ST-segment depression, which has been previously reported to reflect severe coronary disease, in enough patients to warrant firm conclusions, but our clinical experience agrees with the observation that hypotension during exercise generally reflects critical coronary disease. Inappropriate increment in heart rate ("chronotropic incompetence") also added little to the multivariate analysis, but the exclusion in this study of subjects not reaching 90% of maximal predicted heart rate prohibits further comment on the importance of this variable.

Finally, the duration of ischemia during the recovery period had a low, relatively insignificant standardized coefficient. Thus, although prolonged "duration of ischemia" does relate to severity of disease (table 2), the interdependence of this variable with configuration and depth of ST depression tends to minimize its potency in discrimination.

Conclusion

Exercise testing is emerging into an era of quantification. By incorporating several variables into the interpretation of the test, a treadmill score may be derived which relates severity of ischemia to the severity of CAD. The higher the treadmill score, the more likely the test truly reflects CAD, and the higher the probability that advanced degrees of CAD are present. Conversely, the lower the score, the more likely a patient is to be normal or have less severe degrees of coronary narrowing.

Appendix A — Determination of "Onset of Ischemia"

In cases with only non-ischemic responses during exercise testing, onset of ischemia was arbitrarily recorded as appearing at 12 minutes and at a heart rate of 200 beats/min. This adjustment was incorporated to eliminate variables of "infinite" size from the multivariate discriminant function analyses. When ischemic ST-segment changes were noted only in the recovery period (rather than during exercise), the time and heart rate at which ischemia was considered to have appeared was recorded as the maximal level achieved during the exercise. Corrections were not made for the resting heart rate.

Appendix B

Classification Parameters of Test Accuracy

TP and TN indicate true positive and negative responses; FP and FN indicate false positive and negative responses.

\[ Z = \frac{TP}{FP} \]

\[ Z = \frac{TN}{FN} \]

\[ Z = \frac{TP \times TN}{FP \times FN} = \frac{Sens \times Spec}{(1-Sens) \times (1-Spec)} \]

Classification Parameters Dependent on Disease Prevalence

Predictive accuracy = \[ \frac{TP}{TP + FP} \]

True negative rate = \[ \frac{TN}{TN + FN} \]

Correct classification rate = \[ \frac{TP \times TN}{TP \times TN + FP \times FN} \]

Risk ratio = \[ \frac{TP}{TP + FP} \]

\[ \frac{FN}{FN + TN} \]

\[ \frac{PA \times 1}{1-TNR} \]

Posterior probability — calculated based on the SPSS computer system.

Appendix C

Influence of Disease Prevalence on Predictors of Coronary Disease

Theoretically, sensitivity, specificity and odds ratio are un influenced by disease prevalence, as they are defined by a denominator that states that all patients either have disease or are free of it. Subgroups within

\[ *(PA \text{ and } TNR \text{ expressed as proportions rather than percentages.}) \]
a population may, however, have different quantitative and qualitative aspects of their disease (for example, a 50% coronary lesion in a symptomatic patient may differ anatomically and physiologically from what appears to be the same degree of narrowing in an asymptomatic subject). As our classification is based on a population having coronary disease, we feel that predictive statements using these concepts on a group of individuals with a different assumed disease prevalence are only rough estimates.

Appendix D
Calculations of Cost-Benefit Analysis

If the ratio of the costs for false negative and false positive misclassification are predetermined, a conceptual cost benefit disease prevalence (PREVCB) may be derived. The treadmill score at the 50% PP point of this new prevalence (PREVCB) thus defines that cutoff point which yields the specified ratio of costs.

\[
\text{PREV}_{1} = \frac{\text{PREV}_{1} + (C_2 \times \text{PREV}_{2})}{(C_1)}
\]

Where \( \text{PREV}_{1} \) = disease prevalence, \( \text{PREV}_{2} \) = prevalence of non-disease = 1 - \( \text{PREV}_{1} \)

\( C_1 \) = cost of FN diagnosis, and

\( C_2 \) = cost of FP diagnosis

References

Use of treadmill score to quantify ischemic response and predict extent of coronary disease.
K Cohn, B Kamm, N Feteih, R Brand and N Goldschlager