Development and validation of a simple exercise test score for use in women with symptoms of suspected coronary artery disease

Anthony P. Morise, MD,[a] Michael S. Lauer, MD,[b] and Victor F. Froelicher, MD[c] *Morgantown, WVa, Cleveland, Ohio, and Palo Alto, Calif*

Background Recently revised American College of Cardiology/American Heart Association guidelines for exercise electrocardiography (ExECG) have suggested that ExECG scores be used to assist in management decisions in patients with suspected coronary artery disease (CAD).

Methods We used 442 women who underwent both ExECG and coronary angiography (CAD ≥1 lesion with ≥50% stenosis; CAD prevalence was 32%) to derive an ExECG score including clinical and ExECG variables. By use of logistic regression analysis, variables were selected and relative weights were determined. Variable codes multiplied by respective weights were summed to produce a final ExECG score. The score was validated in separate populations concerning angiographic as well as prognostic end points.

Results Clinical variables selected and their weights included age (5), symptoms (2), diabetes (2), smoking (2), and estrogen status (1). ExECG variables selected and their weights included ST depression (2), exercise heart rate (4), and Duke Angina Index (3). For the validation group, score ranges are shown with the prevalence of CAD: <20 = 0/5 or 0%, 20-29 = 3/26 or 11%, 30-39 = 20/56 or 36%, 40-49 = 33/81 or 41%, 50-59 = 24/49 or 49%, 60-69 = 22/32 or 69%, and >70 = 7/7 or 100%. Frequency of death within 3 predetermined subgroups was as follows: low <40 = 3/1237 (0.2%), intermediate 40-60 = 9/383 (2.3%), high >60 = 4/54(7%); *P*< .0001.

Conclusion A simple ExECG score was developed for use specifically in women. When evaluated in separate cohorts, the score stratified women with suspected coronary disease into groups with a gradually increasing frequency of coronary disease and death. (Am Heart J 2002;144:818-25.)

Recent consensus guidelines for chronic stable angina recommend the treadmill exercise test for a variety of diagnostic and prognostic situations.[1] Earlier American College of Cardiology/American Heart Association guidelines for exercise testing suggested that exercise scores be used in the interpretation of exercise tests and in clinical decision-making.[2] Most scores in the literature are limited by their complexity and requirement for a calculator or computer. We have recently developed and validated a simple exercise test score to be used in men with suspected coronary disease.[3] The purpose of the present study was to develop and validate a similar score for use in women.

From the [a]Section of Cardiology, Department of Medicine, West Virginia University School of Medicine, Morgantown, WVa, [b]Department of Cardiology, Cleveland Clinic Foundation, Cleveland, Ohio, and the [c]VA Palo Alto Health Care System, Palo Alto, Calif.

Submitted April 27, 2001; accepted March 28, 2002.

Reprint requests: Anthony Morise, MD, Section of Cardiology, HSC South, WVU, Morgantown, WV 26506.
E-mail: amorise@pol.net

0002-8703/2002/$35.00 + 0 **4/1/125835**
doi:10.1067/mhj.2002.125835

Methods

Patient populations

We screened all women (inpatient and outpatient) aged ≥18 years referred by primary care physicians and cardiologists to the stress laboratory for their first exercise test. First exercise tests could include exercise electrocardiographic, nuclear, or echocardiographic studies. We included only symptomatic women referred with the purpose of evaluating the presence of coronary disease. We excluded asymptomatic women, those receiving digitalis preparations, those with a history of prior myocardial infarction or coronary angiography, and those with resting electrocardiograms that were considered uninterpretable (left ventricular hypertrophy, left bundle branch block, Wolff-Parkinson-White syndrome, or other significant downward displacement of the ST segment).

Patients who underwent coronary angiography within the 3 months after exercise testing at West Virginia University Hospital between 1981 and 1999 were used for the derivation group. A similar group who underwent angiography at the Cleveland Clinic between 1990 and 1998 served as the validation group. Another group seen at West Virginia University between 1995 and 2001 served as the prognostic validation group. These were consecutive women with symptoms of suspected coronary disease selected irrespective of whether they underwent angiography.

Baseline clinical information

We collected the following data from patients during a preexercise electrocardiogram interview: age, symptoms, medication usage at the time of the exercise electrocardiogram, and other coronary risk factors. Patients had height and weight recorded. We classified chest pain by use of the 3 categories of Diamond[4]: typical angina, atypical angina, nonanginal chest pain. Risk factors included the following: current or prior cigarette smoking, history of hypertension (on antihypertensive therapy), history of insulin or noninsulin-requiring diabetes, history of high cholesterol or receiving cholesterol-lowering therapy, a family history of premature (aged <60 years) coronary disease (infarction, coronary bypass or angioplasty, sudden death) in first degree relatives, obesity defined as a body mass index (kg/m^2) >27. We determined estrogen status by use of previously published criteria.[5,6] Women were estrogen-status negative if they were postmenopausal and not receiving estrogen replacement therapy. If they were premenopausal or receiving estrogen replacement therapy, they were considered as estrogen-status positive. Women who underwent hysterectomy without oopherectomy were considered estrogen-status positive if they were aged <50 years and without symptoms of estrogen deficiency. Otherwise, they were considered estrogen-status negative.

Exercise tests

All patients exercised using either Bruce or Cornell treadmill protocols. We did not use predetermined peak heart rates to determine when to stop exercise. We read all studies in a blinded fashion. By use of the 12 standard leads, we measured peak exercise or 3 minute recovery ST-segment changes (60 ms following the J point compared to the baseline between 2 PR segments). We qualitatively categorized peak exercise ST slope as upsloping, horizontal, or downsloping. Positive ST-segment criteria consisted of ≥1 mm horizontal/downsloping ST depression. We also recorded resting and peak exercise heart rate and blood pressure, exercise capacity estimated in metabolic equivalents from the final treadmill speed and grade, and exercise-induced angina. Angina during testing was classified according to the Duke Exercise Angina Index (2 if angina required stopping the test, 1 if angina occurred during or after treadmill test, and 0 for no angina).[7]

End point assessment

Coronary angiograms were performed for routine clinical indications and interpreted visually, blinded to the clinical and exercise electrocardiogram data. Coronary artery disease presence was defined as a ≥50% luminal diameter narrowing in at least 1 vessel. The 50% criterion was consistent with the cooperative trialists' findings.[8] Multivessel coronary artery disease was defined as ≥70% luminal diameter narrowing in at least 2 vessels.

Patients in the prognostic validation group had vital status and date of death determined by a search of the Social Security Death Index.

Statistical analysis

We used stepwise logistic regression analysis (JMP version 3.1) with the presence of coronary artery disease as the dependent variable. The general linear logistic regression model was as follows:

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

where a = intercept, b and c = coefficients, and x and y = variables.

Incremental value was defined as the additional information contributed by the exercise score over the information available from pretest clinical data. Incremental value was evaluated by use of receiver operating characteristic curve area analysis.[9]

The new simple exercise score was compared to a previously validated pretest score[10] and the Duke Treadmill Score.[7] Calibration was assessed qualitatively by use of graphs plotting prevalence as a function of exercise score and quantitatively by use of Hosmer-Lemeshow statistics. Comparison of frequencies was accomplished by use of χ^2 testing. Survival analysis was accomplished by use of Kaplan-Meier curve analysis. *P* values <.05 were considered significant.

Development of score

Previously, we developed and validated multivariable equations that can be used in much the same manner as the proposed exercise score.[11] These equations take the form of exponentials and require the use of a calculator to estimate coronary artery disease probability. The complexity of these equations has not facilitated their widespread application. To decrease this complexity, we sought to reduce the exponential equation to a simple linear score. To accomplish this, we first coded all variables with the same number of intervals such that coefficients derived from the regression analysis would be proportional. The interval with the largest value was associated with a higher probability of disease. For example, if intervals range from 0 to 5, then nominal variables such as diabetes would be coded as 0 if absent and 5 if present. On the other hand, continuous variables such as peak heart rate or age would be coded in several bins with higher heart rates and lower ages coded as 0 and lower heart rates and higher ages coded as 5. All codes would then be proportional to probability and the smallest coefficient would be associated with least important variable. The coefficient of this least important variable was divided into the coefficients of all other variables. The coefficient of the least variable would be reduced to 1 and all other variables to their proportional whole integer weights. This makes the relative importance of the selected variables very obvious. This results in a simple linear score in which each variable code is multiplied by its respective weight and summed to produce a composite score.

$$\text{Score} = ax + by + cz + \ldots$$

where a, b, c = weights and x, y, z = variable codes.

Table I. Clinical and exercise test characteristics

		Validation	
	Derivation	Angiographic	Prognostic
Number	442	256	1678
Age	54 ± 11	59 ± 10	47 ± 16
Symptoms (%)			
Typical	109 (25)	24 (9)	145 (9)
Atypical	186 (42)	42 (16)	696 (41)
Nonanginal	147 (33)	186 (75)	837 (50)
Diabetes (%)	88 (20)	47 (18)	190 (11)
Smoking (%)	155 (35)	46 (18)	665 (40)
Hyperlipidemia (%)	170 (39)	100 (39)	541 (32)
Hypertension (%)	219 (50)	128 (50)	603 (36)
Pretest score	8.4 ± 7.6	12.0 ± 3.5	8.1 ± 4.8
Pretest score group (%)			
Low	221 (50)	35 (14)	1044 (62)
Intermediate	140 (32)	177 (69)	451 (27)
High	81 (18)	44 (17)	183 (11)
Positive ST change (%)	105 (24)	96 (38)	230 (13)
Peak heart rate	146 ± 21	147 ± 20	153 ± 37
Peak systolic BP	162 ± 24	182 ± 30	162 ± 22
METs	5.8 ± 2.2	6.4 ± 1.9	6.9 ± 2.6
Duke angina (%)			
Limiting	63 (14)	8 (3)	135 (8)
Nonlimiting	106 (24)	109 (43)	141 (8)
Exercise score	40 ± 18	45 ± 13	28 ± 16
Exercise score group (%)			
Low	214 (48)	87 (34)	1246 (74)
Intermediate	165 (37)	132 (52)	384 (23)
High	63 (15)	37 (14)	48 (3)

BP, Blood pressure; *METs*, metabolic equivalent.

Results

Patient populations

Table I shows a summary of clinical and exercise test characteristics of the 3 populations. During the period of 1981 to 1999, 2606 symptomatic women with suspected coronary disease and interpretable electrocardiograms underwent exercise testing with or without simultaneous imaging. Of these women, 442 underwent coronary angiography. Coronary disease was found in 140/442, or 32% (52/442, or 12%, multivessel disease).

A similarly derived population of 256 women was obtained from the Cleveland Clinic to serve as the angiographic validation population. Coronary disease was found in 109/256, or 43%, (38/256, or 15%, multivessel disease). Compared with the derivation group, these women were older with differences in the frequency of symptom categories and smoking. Because data were insufficient, only 45, or 18%, of the women in the Cleveland group could be categorized concerning estrogen status (34/45, or 76%, were estrogen-status negative), compared with 193, or 44%, in the derivation group. (When estrogen status data are missing, both the pretest and exercise test scores are designed to have a 0 entered rather than adding or subtracting points.)

Another group of 1678 consecutive women with suspected coronary disease served as a prognostic validation population. All-cause mortality was 16/1678, or 0.95%. Compared with the derivation group, these women were younger, with differences in the frequency of symptom categories as well as lower frequencies of diabetes, hyperlipidemia, and hypertension. Only 458, or 27%, were estrogen-status negative.

Differences in exercise performance between the 3 groups are also noted. These performance differences, as well as the score results, indicate that the prognostic validation group was a lower risk group than either of the angiographic groups.

Score development

Independent variables considered included nominal (smoking, diabetes, hypertension, family history of premature coronary disease, hyperlipidemia, and obesity [body mass index >27]), ordinal (chest pain score, estrogen status, Duke Exercise Angina Index) and continuous (age, peak heart rate, peak systolic blood pressure, ST segment change, metabolic equivalents) variables. Nominal variables were coded as 0 or 5, as defined earlier. Ordinal and continuous variables were divided into 2 to 6 subgroups and coded as 0 to 5.

After logistic regression analysis, the variables in Table II were chosen. Clinical and exercise test variables are respectively listed according to the strength of the association with coronary disease presence. All codes ranged from 0 to 5, except for estrogen status, which ranged from −5 to 5 (may be coded as 0 for missing data). The coefficient for estrogen status, being the smallest, was divided into the other coefficients to derive the respective integer weights shown. Figure 1 displays the scoring method in an easy to follow tabular format similar to what was previously published for the men's exercise test score.[3]

Accuracy of simple score

The simple exercise score ranged from 5 to 91 points for the 442 women. Mean scores for the 3 groups are noted in Table I.

Discrimination. Table III displays and compares the receiver operating characteristic (ROC) curve areas for pretest and exercise test scores for both the derivation group and the angiographic validation group. In addition, ROC curve area comparisons are shown for each of the 3 pretest probability groups. Based on the derivation group, significant incremental value is clearly shown. However, whereas the pretest to exercise test comparisons in each of the validation sub-

Table II. Variables chosen for simple exercise score

Variable	P	Code range	Coefficient	Weight
Age (y)	.00001	0, 3, 5	0.38	5
Smoking	.003	0, 5	0.16	2
Diabetes	.004	0, 5	0.17	2
Estrogen status	.002	−5, 0, 5	0.08	1
Chest pain score	.01	1, 3, 5	0.20	2
Duke angina score	.0002	0, 3, 5	0.24	3
ST depression	.01	0, 3, 5	0.18	2
Peak heart rate	.05	0, 1, 2, 3, 4, 5	0.33	4

groups shows an increase in ROC curve area, in no instance did this reach statistical significance.

Calibration. Based on visual inspection of our derivation group data, we defined 3 risk groups: low (0-39 points), intermediate (40-60 points), and high (61-100 points). These are the same cut-points for the men's exercise score.[3] Table IV displays the prevalence of any coronary disease and multivessel disease in these 3 subgroups. The exercise test score stratification is highly significant ($P < .0001$). Figure 2 presents a visual display of calibration for both any coronary disease and multivessel disease (respective Hosmer-Lemeshow statistics, $\chi^2 = 4$, $P = .27$, and $\chi^2 = 2$, $P = .57$).

Table V breaks down the calibration from Table IV further by classifying the validation group data by both pretest and exercise test scores. It can be clearly seen that all 23 patients with coronary disease in the low exercise test group were originally from the intermediate pretest probability group. No women with a low pretest probability of disease were transformed to the high posttest probability group and only 1 woman with a high pretest probability was transformed to a low posttest probability. Within each of the 3 pretest probability subgroups there was stratification.

Prognosis. Our prognostic validation group was a very low risk group with only 16 deaths over 5 years of follow-up (mean 2.6 years). Stratification according to exercise test score group was as follows: low 3/1237, or 0.2%; intermediate 9/383, or 2.3%; high 4/54, or 7.4%. Despite the small number of deaths, 81% were captured in the intermediate- and high-risk groups. Kaplan-Meier analysis (Figure 3) also revealed a significant difference between these 3 groups ($P <$ .00001). The low- and intermediate-risk groups were different ($P = .0001$), showing separation after 3 years. The intermediate and high-risk groups were different ($P =$ .003), showing separation within the first year.

We also categorized the prognostic group by use of the Duke Treadmill score. Stratification according to Duke Treadmill score group was as follows: low 6/1034, or 0.6%; intermediate 9/620, or 1.5%; high 1/24, or 4.2%. The Duke Treadmill score did stratify these women, but the new score categorized a greater number into the low-risk group with more of the deaths categorized into the intermediate- and high-risk groups.

Figure 1

Variable	Choose response	Sum
Maximal Heart Rate (x4)	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 (x2)	1-2mm =6	
	> 2mm =10	
Age (x5)	>65 yrs =25	
	50 to 65 yrs = 15	
Angina History (x2)	Definite/Typical = 10	
	Probable/atypical =6	
	Non-cardiac pain =2	
Smoking? (x2)	Yes=10	
Diabetes? (x2)	Yes=10	
Exercise test induced Angina (x3)	Occurred =9	
	Reason for stopping =15	
Estrogen Status	Positive=-5, Negative=5	
	Total Score	

Women

Choose only one per group

<40=low probability

40-60= intermediate probability

>60=high probability

This table can be used as an aid at an exercise test workstation. The numbers in parentheses preceded by "x" represent the weights for each variable as noted on Table II. Each alternating shaded and white area represents a different variable within which a numeric choice should be selected for the Sum column. After a total of the sums for the 8 variables, a probability group can be assigned by use of the ranges on the right. See text for further comments.

Figure 4 displays exercise test score as a function of pretest score for the 1678 women in the prognostic validation group. Superimposed on the scatter plot are the borders defining the low-, intermediate-, and high-risk groups for both scores. Within each of the 9 sectors, the distribution of the women as well as those that died is noted.

Discussion

Exercise electrocardiography in women has a questionable reputation within both the cardiology and the

Table III. Incremental value of exercise score over pretest score by use of ROC curve areas

Pretest probability group	Derivation group			Validation group		
	Pretest	*P*	Exercise	Pretest	*P*	Exercise
All	74 ± 3	.0006	82 ± 2	72 ± 3	.25	70 ± 3
Low	48 ± 6	.0001	78 ± 5	65 ± 13	.38	70 ± 13
Intermediate	57 ± 5	.007	70 ± 5	58 ± 4	.32	61 ± 4
High	59 ± 7	.07	71 ± 7	68 ± 9	.07	78 ± 8

Table IV. Prevalence by simple exercise test score risk group

	Derivation			Validation		
	Low	Intermediate	High	Low	Intermediate	High
Number	191	162	89	87	132	37
% CAD	7	39	71	26	45	73
% MVD	1	12	34	6	17	30

CAD, Coronary artery disease; *MVD*, multivessel disease.

Table V. Prevalence of CAD and MVD by use of both pretest and exercise test scores

	Pretest score groups		
	Low	Intermediate	High
High (%)			
CAD	–	10 (56)	17 (89)
MVD	–	2 (11)	9 (47)
	n = 0	n = 18	n = 19
Intermediate (%)			
CAD	3 (17)	40 (44)	16 (67)
MVD	1 (5)	15 (17)	6 (25)
	n = 18	n = 90	n = 24
Low (%)			
CAD	0 (0)	23 (33)	0 (0)
MVD	0 (0)	5 (7)	0 (0)
	n = 17	n = 69	n = 1

Table VI. Exercise scores—comparative weights by sex

	Men	Women
Age	4	5
Symptoms	1	2
Diabetes	1	2
Hyperlipidemia	1	0
Smoking	0	2
Estrogen status	0	1
ST depression	5	2
Peak heart rate	6	4
Duke exercise angina score	1	3

primary care communities. This stems principally from concerns over false positives. Given the published data concerning the accuracy of exercise testing in women[12] and the lower accuracy compared with men,[13] this concern has merit. A specificity of 70% derived from a recent meta-analysis[12] would suggest that on average, 30% of positive exercise electrocardiograms in women are falsely positive. All the aforementioned studies have focused on ST-segment depression. When multivariable approaches have been used, the accuracy of testing in men and women was similar.[11]

Previously reported scores are accurate, but they are also complicated and difficult to use. The women's score proposed in this study is both clinically relevant and easy to apply in any exercise laboratory or clinic without need for a calculator or computer. The exercise test score would be difficult to memorize, but Figure 1 could be posted in an exercise laboratory to facilitate its use.

Given that the scores for men and women were derived by use of the same method, Table VI displays the weights for each variable for the 2 scores. In addition to estrogen status, which is unique to women, 2 other variables (hyperlipidemia and smoking) are not found in both scores. This inconsistency does not mean that hyperlipidemia in women and smoking in men are not important predictors, but these respective variables were not important predictors in these populations. Both of these variables appear in the pretest score,[10] so they can be factored into the diagnostic process of both men and women. Concerning the variables that appeared in both, diabetes and symptoms carried twice the weight in women compared with men. That symptoms were weighted more in women is somewhat surprising given the observations noted in the Coronary Artery Surgery Study.[14] In this analysis, the Diamond classification for chest pain was both less accurate and less precise in women compared with men. Not surprisingly, ST depression was weighted more heavily in men, but still was a predictor in women.

Figure 4 provides a framework for how pretest and exercise test scores might be used in clinical decision-making. Each of the 9 sectors on Figure 4 are numbered so as to identify each later in the discussion. Before selecting a test, the pretest score would discretely define the patient into 1 of 3 risk groups. According to Figure 4, women will be categorized by pretest probability as follows: low 60%, intermediate 25%, and high 15%.

Prior work from our laboratory indicates that as pretest probability increases, negative predictive value for exercise testing decreases.[15] In fact, those in the high pretest probability group have very low negative predictive value and little chance of being reduced to a low posttest probability group by a low exercise score (Figure 4, sector 7). These high pretest probability patients in sectors 7 to 9 might be better served by foregoing the stress test altogether and going directly to the catheterization laboratory.

Conversely, those patients with a pretest score of <4 might forego any stress testing because the likelihood of coronary disease and of being advanced to sectors 2 or 3 is virtually zero. For the remainder of those in the low pretest group with a pretest score >4, 95% will have a low exercise score. For those with exercise scores in sector 2, follow-up exercise imaging would be an appropriate next step.

Lastly, concerning the intermediate pretest group, the lower the exercise score, the better. Those in sectors 5 and 6 should undergo exercise imaging and angiography, respectively. Concerning those in sector 4, Table V suggests that despite being at lower risk than sectors 5 and 6, sector 4 patients may still have a considerable risk of coronary disease. The risk of multivessel disease and mortality is low, but perhaps a third of these women may have single vessel disease. In lieu of clinical trials giving clear guidance as to the most appropriate strategy, we suggest the following approach. Patients who fall into sector 4 should undergo exercise imaging unless their exercise test is free of any abnormal diagnostic or prognostic findings, that is, (1) <80% of age-predicted exercise capacity, (2) abnormal chronotropic response (without β-blockers present) or heart rate recovery, (3) horizontal or downsloping ST depression between 0 and 1 mm, (4) <20 mm Hg rise in systolic blood pressure, or (5) the presence of exercise-induced ventricular ectopy >10% of all beats.

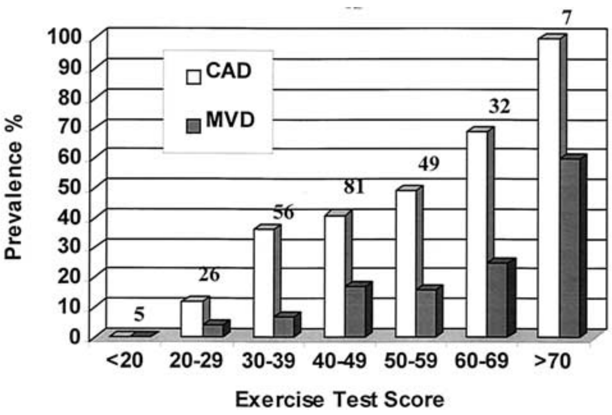
This study is limited by the use of angiographic populations referred to the catheterization laboratory after the exercise test. This referral bias tends to inflate sensitivity and deflate specificity. However, our prognostic validation group should not have been affected by posttest referral bias. We limited our exercise test-to-angiography time to 3 months. This would clearly limit the number for inclusion, but would also limit the error induced by changes in angiographic anatomy after the exercise test.

Some of our clinical variables, for instance, hypercholesterolemia, are less precisely defined than others. This softer definition allows easier data collection but makes for less precision when assessing the variable's true diagnostic worth. This may explain why hypercholesterolemia was not an independent predictor for our women, whereas it was for men. Nevertheless, considering the approach of Table V to combine pretest and exercise test scores, hypercholesterolemia is included in the pretest score for both men and women. Clinicians may choose to consider more detailed lipid profile data when deciding whether a significant lipid disorder exists.

The absence of statistically significant incremental value as reflected by the ROC curve analysis warrants further consideration. The qualitative increases in incremental value noted on the right side of Table III failed to achieve statistical significance. Is this the result of small sample size? Is this caused by the lack of adequate estrogen status data in >80% of the validation group? Perhaps, but the calibration analysis indicates that significant diagnostic and prognostic stratification was achieved, especially when both scores were used together. As previously stated by Diamond et al,[16] resolution (as reflected by the ROC curve results) is more relevant to epidemiologic (ie, group) decisions, whereas calibration (as reflected by Figure 2 and Table V) is more relevant to clinical (ie, individual) decisions. The exercise test score, while reasonably well calibrated concerning both diagnostic and prognostic endpoints, works best when combined with the pretest score. Therefore, the ROC curve area analysis does not tell the whole story when considering whether scores might be useful.

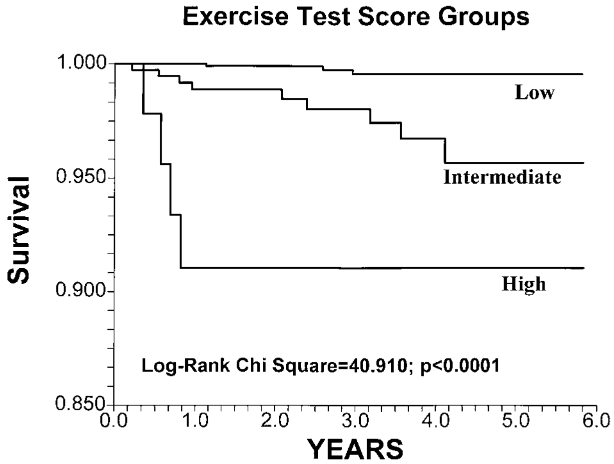
Finally, the simplicity of the model (or lack thereof) deserves discussion. Our intent was to develop a model that was simple enough to be usable and portable to other populations and complex enough to capture what is really occurring in a multifaceted biological process. Given the pervasiveness of computers, complex models are not difficult to apply in an exercise laboratory. Nevertheless, there seems to be a void between the appearance of complex models requiring a computer and their use in everyday practice. On the other hand, although complex models tend to do a better job fitting data, they are also more prone to be

Figure 2



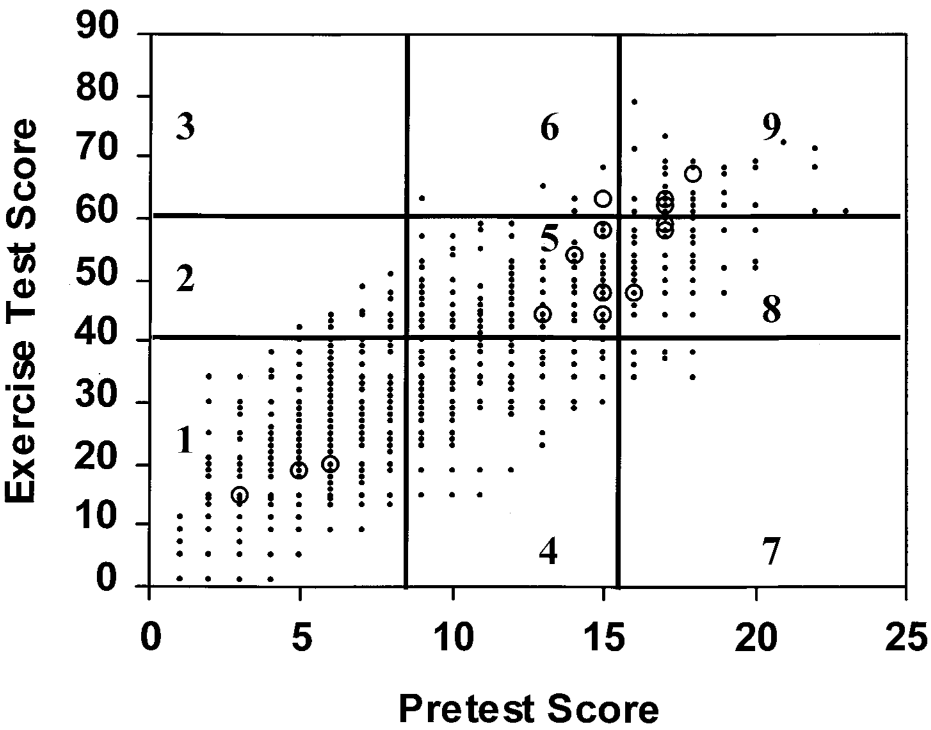
A vertical bar graph of increments of exercise score as a function of the prevalence of any coronary disease (CAD, *white*) and multivessel disease (MVD, *gray*). The calibrations for any disease and multivessel disease are reasonably linear. *Numbers above bars* represent the denominator of the frequency for each bar. Note that there is stratification within the low-risk exercise group (0-39 points).

Figure 3



Kaplan-Meier survival curves for 3 exercise test score probability groups. See text for discussion.

Figure 4



Scatter plot of exercise score (y-axis) as a function of pretest score (x-axis) for 1678 women. Points for those alive *(closed circles)* and dead *(open circles)* are displayed. Two deceased patients overlapped at the 13,44 point. Heavy gridlines represent cut-points for low-intermediate and intermediate-high risk groups for each score. The resulting 9 sectors are numbered from the left, *1-9*. See text for further comments.

unstable, leading to problems with external validation. Our exercise test score strikes a balance between simplicity, usability, and portability, and provides an immediately available tool without the need for a computer.

In conclusion, the proposed simplified exercise score for women has characteristics similar to a validated score for men. Confirmation of its calibration to predict angiographic prevalence and all-cause mortality in other populations of women suggest that its use in concert with the previously validated pretest score will improve decision-making for women with symptoms of suspected coronary disease. Further prospective observations in relevant populations, as well as clinical trials, should be pursued to determine whether cost effectiveness is enhanced.

References

1. Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients with Chronic Stable Angina). J Am Coll Cardiol 1999;33:2092-197.
2. Gibbons RJ, Balady GJ, Beasley JW, et al. ACC/AHA guidelines for exercise testing: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). J Am Coll Cardiol 1997;30:260-315.
3. Raxwal VK, Shetler K, Morise AP, et al. Simple validated treadmill score to diagnose coronary disease. Chest 2001;119:1933-40.
4. Diamond GA. A clinically relevant classification of chest discomfort. J Am Coll Cardiol 1983;1:574-5.
5. Morise AP, Dalal JN, Duval RD. Value of a simple measure of estrogen status for improving the diagnosis of coronary artery disease in women. Am J Med 1993;94:491-6.
6. Morise AP, Haddad WJ. Validation of estrogen status as an independent predictor of coronary artery disease presence and extent in women. J Cardiovasc Risk 1997;3:507-11.
7. Mark D, Hlatky M, Harrel F, et al. Exercise treadmill score for predicting prognosis in coronary artery disease. Ann Intern Med 1987;106:793-800.
8. Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomized trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet 1994;344:563-70.
9. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making 1984;4:137-50.
10. Morise AP, Haddad WJ, 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 1997;102:350-6.
11. Morise AP, Diamond GA, Detrano R, et al. Incremental value of exercise electrocardiography and thallium-201 testing in men and women for the presence and extent of coronary artery disease. Am Heart J 1995;130:267-76.
12. Kwok Y, Kim C, Grady D, et al. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999;83:660-6.
13. Morise AP, Diamond GA. Comparison of the sensitivity and specificity of exercise electrocardiography in biased and unbiased populations of men and women. Am Heart J 1995;130:741-7.
14. Diamond GA, Forrester JS. Probability of CAD. Circulation 1982;65:641-2.
15. Morise AP. Are the American College of Cardiology/American Heart Association guidelines for exercise testing for suspected coronary artery disease correct? Chest 2000;118:535-41.
16. Diamond GA, Denton TA. Risky business: prospective applicability of models. Circulation 1993;87:1054-5.