

Annals of Internal Medicine

The Electrocardiographic Exercise Test in a Population with Reduced Workup Bias: Diagnostic Performance, Computerized Interpretation, and Multivariable Prediction

Victor F. Froelicher, MD; Kenneth G. Lehmann, MD; Ronald Thomas, PhD; Steven Goldman, MD; Douglas Morrison, MD; Robert Edson, MS; Philip Lavori, PhD; Jonathan Myers, PhD; Charles Dennis, MD; Ralph Shabetai, MD; Dat Do, BA; Jeffrey Froning, MS; and the Veterans Affairs Cooperative Study in Health Services #016 (QUEXTA) Study Group*

Background: Empirical scores, computerized ST-segment measurements, and equations have been proposed as tools for improving the diagnostic performance of the exercise test.

Objective: To compare the diagnostic utility of these scores, measurements, and equations with that of visual ST-segment measurements in patients with reduced workup bias.

Design: Prospective analysis.

Setting: 12 university-affiliated Veterans Affairs Medical Centers.

Patients: 814 consecutive patients who presented with angina pectoris and agreed to undergo both exercise testing and coronary angiography.

Measurements: Digital electrocardiographic recorders and angiographic calipers were used for testing at each site, and test results were sent to core laboratories.

Results: Although 25% of patients had previously had testing, workup bias was reduced, as shown by comparison with a pilot study group. This reduction resulted in a sensitivity of 45% and a specificity of 85% for visual analysis. Computerized measurements and visual analysis had similar diagnostic power. Equations incorporating nonelectrocardiographic variables and either visual or computerized ST-segment measurement had similar discrimination and were superior to single ST-segment measurements. These equations correctly classified 5 more patients of every 100 tested (areas under the receiver-operating characteristic curve, 0.80 for equations and 0.68 for visual analysis; $P < 0.001$) in this population with a 50% prevalence of disease.

Conclusions: Standard exercise tests had lower sensitivity but higher specificity in this population with reduced work-up bias than in previous studies. Computerized ST-segment measurements were similar to visual ST-segment measurements made by cardiologists. Considering more than ST-segment measurements can enhance the diagnostic power of the exercise test.

This paper is also available at <http://www.acponline.org>.

Ann Intern Med. 1998;128:965-974.

For author affiliations and current author addresses, see end of text.

* For members of the Veterans Affairs Cooperative Study in Health Services #016 (QUEXTA) Study Group, see Appendix 2.

The standard exercise test is still the first step in the evaluation of the stable patient with chest pain that may be due to coronary artery disease. This is because simple ST-segment measurements are as diagnostic as other tests that can be performed by the clinician (1, 2). Although studies suggest that the discrimination of multivariable equations (3), heart rate adjustment (4), and scores (5) is superior to that of ST-segment measurements, failure to validate this superiority has impeded acceptance of these tools. Even in correlation studies that have appropriately enrolled consecutive patients who have had both exercise testing and coronary angiography, workup bias has been a limitation. Patients in these studies were selected for angiography if a physician judged that the likelihood of coronary disease was high enough to warrant this invasive procedure. This selection process makes patients with abnormal exercise test results more likely to be chosen and excludes patients with normal test results and high exercise capacity; this results in a higher prevalence of disease than would be seen in a clinical population.

Prediction equations, scores, and heart rate adjustment algorithms have been derived from popu-

See editorial comment on pp 1035-1037.

Table 1. Clinical Characteristics of Patients in the QUEXTA Study (n = 814)*

Variable	Mean Value \pm SD	Frequency, n (%)	Difference between Value for Patients with CAD and Value for Patients without CAD†		P Value
			Mean	Percentage	
Age, y‡	58.4 \pm 10.8		-6.6		<0.001
Body mass index, kg/m ²	28.0 \pm 4.8		-0.1		>0.2
Pack-years of smoking‡	32.9 \pm 34.3		-3.1		0.04
Diastolic BP before exercise, mm Hg‡	79.0 \pm 10.7		2.4		0.2
Systolic BP before exercise, mm Hg‡	125.5 \pm 19.7		-2.1		0.2
Heart rate before exercise, beats/min‡	78.7 \pm 15.3		1.7		>0.2
Definite angina‡		254 (31.2)		-20.4	<0.001
Probable angina		454 (55.8)		19.8	<0.001
Nonspecific chest pain		84 (10.3)		1.7	0.043
Chest pain at rest‡		545 (67.0)		16.8	<0.001
Chest pain only during exercise		247 (30.3)		-15.7	<0.001
Diabetes‡		135 (16.6)		-10.5	<0.001
Hypercholesterolemia		305 (37.5)		-11.0	<0.001
Current cigarette use		287 (35.3)		5.3	0.20
Obesity		470 (57.7)		-0.4	>0.2
Peripheral vascular disease		76 (9.3)		0.3	>0.2
Congestive heart failure		19 (2.3)		0.2	>0.2
Chronic obstructive pulmonary disease		60 (7.4)		-0.7	>0.2
Family history of CAD‡		440 (54.1)		1.1	>0.2
Hypertension		439 (53.9)		-6.5	0.032
Low high-density lipoprotein cholesterol level		122 (15.0)		-5.8	0.0037
Stroke		31 (3.8)		-3.3	0.014
Receiving digoxin		21 (2.6)		-0.8	>0.2
Receiving β -blocker		216 (26.5)		-10.0	<0.001

* BP = blood pressure; CAD = coronary artery disease; QUEXTA = Quantitative Exercise Testing and Angiography.

† 402 patients (49.4%) had CAD.

‡ Candidate for multivariate modeling.

lations with extensive workup bias and are unlikely to be applicable to patients who present with chest pain (6). Our study reduced workup bias prospectively by following a protocol that required patients to agree to undergo both exercise testing and coronary angiography. A pilot study that did not avoid workup bias was done in 687 patients at two sites from October 1990 to August 1994. The main study, Quantitative Exercise Testing and Angiography (QUEXTA), enrolled 1274 patients at 12 sites from August 1994 to September 1995.

Methods

Patients

To be included in QUEXTA, patients had to be men 18 years of age or older with probable or definite stable angina. Standard exclusion criteria were used, and patients with previous myocardial infarction or previous abnormal angiograms were excluded. To further minimize workup bias, the study allowed no more than 25% of the patients at any one site to have had a recent treadmill test. The preferred entry point was the clinic, but less than 25% of patients could come from either the exercise or angiography laboratories. Of 1274 consecutive male patients who were enrolled at 12 Veterans Affairs Medical Centers between 22 August 1994 and 15 September 1995, 814 had no myocardial

infarction on electrocardiography or history, underwent both coronary angiography and treadmill testing, and had complete data. Institutional review was done centrally and at each study site, and all patients signed a consent form approved for this study. Coronary angiography and treadmill testing had to be done within 30 days of each other. For validation purposes, the 814 patients were divided into a training set of 543 patients (two thirds of the total sample) and a test set of 271 patients (one third of the total sample). Approximately 7000 patients had exercise testing, and 1328 patients were enrolled during the recruitment period.

Clinical variables obtained at the initial evaluation were recorded on a standard form. Chest pain was coded as 1 for definite angina, 2 for probable angina, 3 for nonanginal pain, and 4 for no pain. All other clinical variables, except age, body mass index, resting ST-segment depression, hemodynamic variables, and pack-years of cigarette smoking, were coded as present or absent.

Exercise Testing

All patients had exercise testing done with a ramp treadmill protocol (7). ST-segment depression was measured at the J junction to the nearest quarter millimeter, and ST slope, measured over the following 60 milliseconds of the ST segment, was classified as upsloping, horizontal, or downsloping. ST slope was coded as 1 for abnormal (horizontal

or downsloping and at least 1 mm of depression) or 0 for normal slope. The 12-lead electrocardiograms were read by two cardiologists at each site on separate days, once by using raw signals and once by using the device averages. The cardiologists were blinded to patient identity and test results. Maximal and delta values for hemodynamic variables, along with exercise-induced hypotension, exercise-induced angina, and exercise capacity estimated in metabolic equivalents (METs) from the final treadmill speed and grade, were recorded. Angina during testing was classified according to the Duke Angina Index (2 if angina required that the test be stopped, 1 if angina occurred during or after the test, and 0 if no angina occurred) (8). No test result was classified as indeterminate (9). Medications were withheld only on the day of testing, and no maximal heart rate targets were applied (10).

Computer Analysis

Electrocardiographic devices were used at all sites to simultaneously record in digital format all 12 electrocardiographic leads through exercise and recovery at 500 samples per second (Mortara Instrument, Milwaukee, Wisconsin) on optical disks (11). Optical disk recordings were processed off-line by using a microcomputer at the exercise electrocardiography core laboratory. After the raw data were averaged, QRS measurement landmarks were determined by using software developed by Sunny-side Biomedical (Vista, California) (12).

Coronary Angiography

Coronary angiography was done with standard techniques after administration of nitroglycerin. Trained observers at each site made blinded quan-

titative measurements. All stenoses with visual percentage narrowing greater than 30% were measured. Raw measurements were sent to the core angiography laboratory in Seattle, where they were converted to true diameters after correction for distortion. In a randomized selection, the mean difference per stenosis between the measurement in the core laboratory and measurements at participating sites was 0.9%, with a mean absolute value of 11.4%. Patients were categorized as having significant coronary artery disease if at least one stenosis with narrowing of 50% or more by quantitative measurement was present in any artery or branch with a reference diameter of at least 1 mm.

Statistical Analysis

The summary statistics were examined, and several variables were eliminated from the model building because of their low prevalence or low variance. Examination of the distribution of the visual ST-segment measurements and their relation to the angiographic results led to the choice to use raw visual interpretation of the maximal abnormal ST-segment depression in either exercise or recovery. On the basis of these results, 20 variables were chosen for multivariate analyses (Tables 1 and 2). The training set for diagnosis of any coronary artery disease was divided into two groups, one with and one without significant angiographic coronary artery disease. After a logistic regression equation was developed for predicting pre-exercise test probability for coronary artery disease, the exercise test hemodynamic and nonelectrocardiographic variables were added to the pre-exercise test variables as candidates. This allowed variable selection for three additional models to predict post-exercise test proba-

Table 2. Hemodynamic and Visual Electrocardiographic Characteristics of Patients in the QUEXTA Study*

Variable	Mean Value \pm SD	Frequency, n (%)	Difference between Value for Patients with CAD and Value for Patients without CAD		P Value
			Mean	Percentage	
Heart rate at maximal exercise, <i>beats/min</i> [†]	128.4 \pm 24.0		12.9		<0.001
Change in heart rate, <i>beats/min</i> [†]	49.6 \pm 22.8		11.2		<0.001
Systolic BP at maximal exercise, <i>mm Hg</i> [†]	169.6 \pm 30.4		6.6		0.007
Change in systolic BP, <i>mm Hg</i> [†]	44.1 \pm 25.6		8.7		<0.001
Double product (heart rate \times systolic BP/1000) [†]	22.1 \pm 6.6		3.0		<0.001
Change in double product [†]	12.2 \pm 6.0		3.0		<0.001
Metabolic equivalents [†]	7.4 \pm 3.2		1.9		<0.001
Angina pectoris score [†]	0.53 \pm 0.71		-0.23		<0.001
Maximal ST-segment depression during exercise or recovery, <i>mm</i> [†]	0.51 \pm 0.83		-0.55		<0.001
Maximal ST-segment depression during exercise [†]	0.43 \pm 0.78		-0.46		<0.001
Maximal ST-segment depression during recovery [†]	0.35 \pm 0.66		-0.44		<0.001
ST-segment depression at rest	0.06 \pm 0.21		-0.04		0.030
Abnormal resting electrocardiogram		300 (36.9)		-14.2	0.002
Hypotension		22 (2.7)		-2.0	0.07
Abnormal visual ST-segment depression		236 (29.0)		-28.2	<0.001
Abnormal visual ST-segment slope		279 (34.3)		-28.1	<0.001

* BP = blood pressure; CAD = coronary artery disease.
[†] Candidate for multivariate modeling.

Table 3. Results from the QUEXTA Test Set Obtained by Using Unsimplified Multivariable Equations*

Variable	Visual ST-Segment Analysis	Equation Using Clinical Variables (Step 1)	Equation Using Clinical and Hemodynamic Variables (Step 2)	Equation Using Clinical, Hemodynamic, and Visual ST-Segment Analysis (Step 3a)	Equation Using Clinical, Hemodynamic, and Computerized ST-Segment Variables (Step 3b)
ROC curve	0.68	0.74	0.74	0.80	0.80
Sensitivity (95% CI), %	45 (32–58)	50 (32–62)	38 (23–50)	54 (45–69)	57 (48–67)
Difference in ROC curve†	Reference	0.14	0.21	0.0001	0.0003
Positive likelihood ratio	3	3.3	2.5	3.6	3.8
Negative likelihood ratio	1/1.5	1/1.7	1/1.4	1/1.8	1/2.0

* Test set with specificity set at value obtained by using visual analysis cut-off criterion of 1 mm of horizontal or downsloping ST-segment depression at any time during exercise or recovery (85%); sum of the depression at ST60 in leads II, V5, and V2; the most ST60 depression in the three leads; the time area in recovery of the slope and ST60 for V5; heart rate index (ST60 V5); and the simplest measurement of ST60 V5 at 3.5 minutes of recovery. The statistics are for the complex rather than the simplified models. HR = heart rate; ROC = receiver-operating characteristic; ST60 = ST amplitude 60 milliseconds after QRS end; QUEXTA = Quantitative Exercise Testing and Angiography.

† Statistical difference of the ROC between visual ST-segment analysis and the various equations and ST-segment measurements.

bility for coronary artery disease and to compare the discriminating power of computerized and visual measurements.

On the basis of the protocol and previous publications, the computer variables considered for the equations were 1) ST/HR (heart rate) index calculated at ST0 and ST60, 2) the Hollenberg score, 3) depression at ST60 (ST amplitude 60 milliseconds after J junction) in V5 at a heart rate of 100 beats/min, 4) ST integral in V5 at 3.5 minutes of recovery, 5) ST slope in V5 at maximal exercise, 6) ST slope in V5 at 3.5 minutes of recovery, 7) ST amplitude at J junction with a horizontal ST slope in V5 at maximal exercise, 8) ST amplitude at J junction with a horizontal ST slope in V5 at 3.5 minutes of recovery, 9) ST60 in V5 at 3.5 minutes of recovery, and 10) ST60 in II at 3.5 minutes of recovery. The most ST60 depression and the sum of ST60 depression in the three major perpendicular leads (II, V2, and V5) at maximal exercise and 3.5 minutes of recovery, as well as ST60, ST0, and ST integral in V2 and II, were considered individually. Comparisons were based on the area under the receiver-operating characteristic (ROC) curve and on sensitivity at the fixed specificity for visual ST-segment depression. **Table 3** shows the results obtained by comparing visual ST-segment depression separately with every other model.

Comparison with the pilot population (**Appendix Table 2**), which had a prevalence of disease similar to that of the study population and was tested by using the same methods, showed how effective the protocol was in reducing workup bias.

Results

Clinical and Resting Electrocardiographic Variables

Table 1 shows summary statistics for clinical variables in the full diagnostic group. According to the angiographic criteria, 276 patients in the training set

and 135 patients in the test set had coronary disease. We noted that in our patients, all of whom had stable chest pain, the probability of coronary artery disease was almost halved if the pain ever occurred at rest. Thus, pain at rest was included as a candidate variable. The pre-exercise test variables chosen by the logistic model for the pre-exercise test equation included age (explaining 60% of total variance), chest pain type (explaining 30% of total variance), diabetes, and pack-years of smoking.

Exercise Test Variables

Table 2 shows summary statistics for hemodynamic and visual electrocardiographic variables measured after the exercise test. The post-exercise test variables chosen for the equation by the logistic model included the Duke Angina Score (explaining 10% of total variance), METs, and maximal heart rate; age explained 20% of the variance. Adding the visual or computerized electrocardiographic variables for consideration in the model produced equations in which ST-segment variables explained 30% of the total variance, age explained 10%, METs explained 10%, and maximal heart rate explained 10%. Of the computerized measurements considered, the logistic model chose five: two empirical adjustments to simulate visual analysis, the ST/HR index, and slope and ST60 in V5 at 3.5 minutes of recovery. However, for the equations that considered only computerized electrocardiographic measurements, ST60 in V5 at 3.5 minutes of recovery was used because it discriminated as well as the equations with five variables (**Appendix Table 1**).

Area under the Receiver-Operating Characteristic Curve Analysis

Receiver-operating characteristic curves were generated from probability scores derived from the unsimplified pre-exercise test and post-exercise test equations as well as from the important ST-segment measurements, including the Hollenberg score, the ST/HR index, and visual analysis. The ROC plots

Table 3—Continued

Equation Using Clinical, Hemodynamic, Visual, and Computerized ST-Segment Variables (Step 4)	Hollenberg Score	ST/HR Index	Sum of ST60 Depression in Leads V5, V2, and II	ST-Segment Integral	Most ST60 Depression in Leads V5, V2, or II	ST60 in V5 at 3 Minutes of Recovery
0.80	0.68	0.70	0.70	0.69	0.70	0.72
56 (42–63)	35 (24–47)	49 (40–59)	37 (24–48)	37 (24–49)	34 (20–46)	47 (31–60)
0.0001	0.87	0.61	0.53	0.78	0.51	0.19
3.7	2.3	3.3	2.5	2.5	2.3	3.1
1/1.9	1/1.3	1/1.7	1/1.3	1/1.3	1/1.3	1/1.6

for the models and the univariate ST-segment measurements for the test set are shown in **Figures 1 and 2**. These curves were also used to determine the sensitivity of the various approaches at a specificity matching that of the standard 1-mm criterion. For the standard ST-segment criteria, the sensitivity was 45% and the specificity was 85% in both the training and test sets; to facilitate comparison with this standard of interpretation, the sensitivities for the other methods were compared at a matched specificity of 85%. The multivariable analysis method significantly improved prediction over that seen with the visual ST-segment criteria, increasing predictive accuracy from 65% to 70% with visual analysis and to 71% with the addition of computer analysis. Therefore, in this population with a 50% prevalence of coronary artery disease, 5 or 6 (95% CI, 3.3 to 7.3) more patients of every 100 tested would be correctly classified with the use of multivariable techniques. The two most advocated approaches (the ST/HR index and the Hollenberg score), in addition to the “best” single ST-segment measurement from previous studies (ST60 in V5 at 3.5 minutes of recovery) (13, 14), are shown in **Tables 3 and 4**.

For the test set data, the ROC area increased for the post-exercise test equation that included the exercise test variables and computerized ST-segment measurements. The ROC areas for the post-exercise test equations, including visual or computerized ST-segment measurements, were significantly greater than those for the standard electrocardiographic criteria ($P < 0.001$). The Hollenberg score, the ST/HR index, and the computerized measurements were not superior to standard visual criteria, but they were similar and served as reasonable substitutes for visual analysis.

Tables 3 and 4 list our major findings. Comparison is always in reference to the standard visual criteria for abnormal results: 1 mm of horizontal or downsloping ST-segment depression occurring at any time in any of the 12 leads during exercise or recovery. The ROC results using the 0 to 1 scores

(calculated probabilities) of the unsimplified multivariable equations are provided, as are the likelihood ratios. These ratios tell the clinician how much the odds of disease are increased by a positive test result or decreased by a negative test result.

Appendix Table 2 provides similar analyses from the pilot study to show the effects of selection (workup bias) in a different population in which the same technical methods were used.

Other Leads

Review of the 12-lead visual electrocardiographic interpretations confirmed that changes isolated to the inferior leads were rare in our patients, who had no diagnostic Q waves. Considering the sum of ST-segment depression or the most depression in the three leads representing the three main areas of the myocardium (II, V2, and V5) did not improve the diagnostic capacity of the test, as shown by the ROC values (0.67 compared with 0.69; $P > 0.2$).

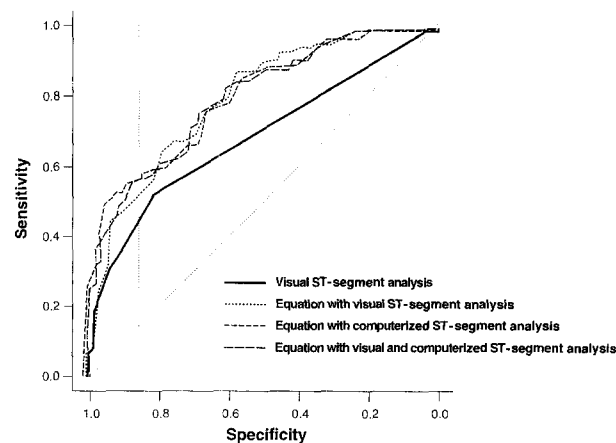


Figure 1. Receiver-operating characteristic curves comparing the diagnostic capacity of standard visual ST-segment analysis with that of the major logistic regression equations. The vertical line is at the specificity obtained with 1 mm of visual ST-segment depression (85%). The equation from step 4 (**Tables 3 and 4**) that included clinical, hemodynamic, visual ST-segment, and computerized measurements is similar to the equation from step 3b, and both equations have better discrimination than single electrocardiographic measurements do.

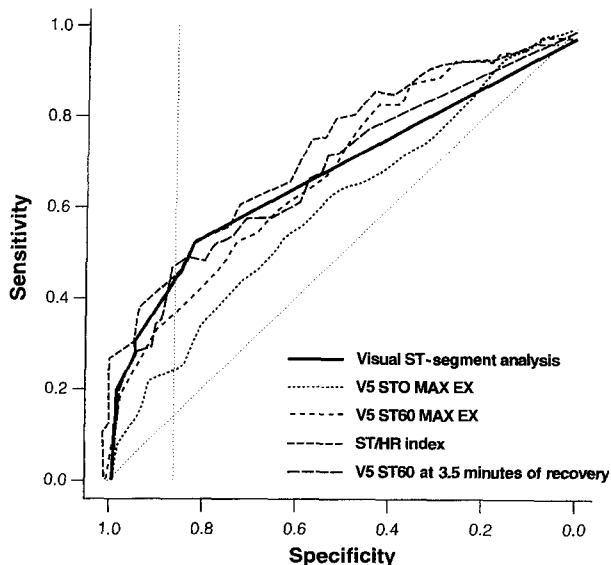


Figure 2. Receiver-operating characteristic curves comparing the diagnostic capacity of standard visual ST-segment analysis with that of several computerized measurements. The computer measurements are similar but not superior to visual analysis. MAX EX = maximal exercise; STO = beginning of ST segment; ST60 = ST amplitude 60 milliseconds after QRS end.

R-Wave Adjustment

Adjusting the computerized measurements by using the R-wave amplitude did not significantly improve the ROC areas (0.69 compared with 0.69, $P > 0.2$).

Recovery Measurements

Although the visual analysis considered abnormal ST-segment depression in exercise or recovery, a separate analysis of the training set showed that 43 of the 156 patients with abnormal ST-segment responses achieved the 1-mm ST-segment criteria only during exercise and that 21 were abnormal only during recovery. The ROC values for measurements during recovery always tended to be greater than comparable measurements during maximal exercise.

Discussion

In 1989, Philbrick and colleagues (6) did a methodologic review of 58 studies that included 7501

patients who had undergone both exercise testing and coronary angiography. They identified seven methodologic standards necessary to eliminate any bias or condition that would lead to an inaccurate assessment of exercise electrocardiographic criteria. In QUEXTA, all of these criteria were fulfilled and workup bias was reduced through a protocol that required patients to agree to both coronary angiography and treadmill testing before study enrollment.

The results of the 58 studies examined in a meta-analysis (15) that appropriately removed patients who had previously had myocardial infarction provide the best estimate of the performance of the exercise tests with workup bias. These studies showed exercise testing to have a mean sensitivity of 67% and a mean specificity of 72%. In one of the few studies in which workup bias was lessened, the sensitivity was 40% and the specificity was 96% (16); these results are similar to ours. Although a lower prevalence and severity of disease could account for the decreased sensitivity, comparison with the studies in the meta-analysis and the pilot data (**Appendix 1** and **Appendix Table 2**) did not support this explanation. Thus, we showed the sensitivity and specificity of the exercise test as it functions using the 1-mm criterion in a typical office population.

Six studies (17–24) have compared multiple computerized exercise electrocardiographic criteria much like QUEXTA did. Two of the six found computerized measurements to be superior to visual measurements (18, 19), and one found the two types of measurement to be similar (21). One found that multivariable techniques were superior to visual measurements but that the computerized ST-segment measurements were not necessary in the prediction equation (25). The other two studies did not consider visual measurements and found a multivariable model to be superior to any single ST-segment measurement (22, 23). Ours was the only study to compare all major computerized and visual ST-segment measurements and to apply multivariable techniques and confirm the consensus of previous studies.

Controversy surrounds the role of heart rate ad-

Table 4. Results from the QUEXTA Training Set Obtained by Using Unsimplified Multivariable Equations*

Variable	Visual ST-Segment Analysis	Equation Using Clinical Variables (Step 1)	Equation Using Clinical and Hemodynamic Variables (Step 2)	Equation Using Clinical and Hemodynamic Variables and Visual ST-Segment Analysis (Step 3a)	Equation Using Clinical, Hemodynamic, and Computerized ST-Segment Variables (Step 3b)
ROC curve	0.65	0.73	0.76	0.79	0.82
Sensitivity, %	44	35	43	53	61
Cut-off	-0.1 mv	0.65	0.66	0.63	0.61

* Training set with specificity set at value obtained by using visual analysis cut-off criterion of 1 mm of horizontal or downsloping ST-segment depression at any time during exercise or recovery (85%); sum of the depression at ST60 in leads II, V5, and V2; the most ST60 depression in the three leads; the time area in recovery of the slope and ST60 for V5; heart rate index (ST60 V5); and the simplest measurement of ST60 V5 at 3.5 minutes of recovery. The statistics are for the complex rather than the simplified models. HR = heart rate; ROC = receiver-operating characteristic; QUEXTA = Quantitative Exercise Testing and Angiography.

justment in the interpretation of exercise test results (25–28). The studies with positive results seem to have had limited challenge, whereas such studies as QUEXTA, which include all consecutive patients presenting with chest pain, have not found an improvement in test characteristics with the ST/HR index. The inclusion of normal persons exaggerates the performance of heart rate correction algorithms because of the differences in maximum heart rate between normal persons and patients with chest pain (29, 30). In QUEXTA, the discriminating power of the ST/HR index was similar to that of computerized variables and visual analysis.

Hollenberg and coworkers (31) proposed a treadmill exercise score that considers all ST-segment measurements made during exercise and recovery and combines them with heart rate and METs. They reported that the treadmill exercise score gave the exercise test a sensitivity of 85% and a specificity of 95% and was superior to visual measurement (31). The results of QUEXTA, Detrano and associates (22), and Deckers and colleagues (21) did not validate the findings of Hollenberg and coworkers.

After it was introduced by Sheffield and coworkers (32), ST integral was studied by both Ascoop and colleagues (17) and Simoons (18), who concluded that it had a lower diagnostic value than other criteria. The appeal of ST integral is that it combines ST-segment slope and depression as one variable. Our results indicate that ST integral can be diagnostic, especially when measured during recovery, but the observed results were similar to those of standard visual analysis.

Review of the 12-lead visual electrocardiographic interpretations confirmed that changes isolated to the inferior leads were rare and confirmed previous study results showing that they added nothing to the diagnostic characteristics of the test when the resting electrocardiogram was normal (33). In our study, computerized measurements considering either the sum of ST-segment depression or the most depression in the three leads representing the three main areas of the myocardium failed to improve the diagnostic accuracy of the test.

The importance of recovery measurements was consistent with previous experience from visual analysis (34). For this measurement to function as it did in our study, the patient must lie down immediately after exercise and not perform a cool-down walk. Because ST60 in V5 at 3.5 minutes of recovery is a simple measurement that is less contaminated by noise, it has much to recommend it.

Previous studies have suggested that adjusting ST-segment depression measurements by using R-wave amplitudes yields greater diagnostic results than using ST-segment depression measurements alone (35). This is because patients with small R-wave amplitudes do not manifest as much ST-segment depression with exercise despite the presence of coronary artery disease, whereas patients with large R-wave amplitudes tend to have exaggerated ST-segment changes. We found no differences in ROC areas when we used the computerized measurements in V5 at maximal exercise or during recovery by adjusting for R-wave amplitude (ROC, 0.69 compared with 0.69).

The application of multivariate analysis (using multivariable techniques) to clinical and exercise test variables has been shown to improve on the standard application of exercise electrocardiography for diagnosing coronary artery disease. A recent meta-analysis (3) of 24 studies that considered exercise test and clinical variables to predict the presence of any angiographic disease found the following variables to be significant predictors in more than half of the studies: sex, chest pain, age, cholesterol levels, ST slope and depression, and maximal heart rate.

This meta-analysis (3) showed that two of the clinical variables used in QUEXTA (age and chest pain) are important clinical predictors. However, we found that diabetes and pack-years of smoking were significant predictors, whereas history of hypercholesterolemia was not. Of the exercise hemodynamic and electrocardiographic variables, ST-segment depression, slope, and maximal heart rate were significant predictors. The same variables were noted to be significant by meta-analysis in more than half of

Table 4—Continued

Equation Using Clinical, Hemodynamic, Visual, and Computerized ST-Segment Variables (Step 4)	Hollenberg Score	ST/HR Index	Sum of ST60 Depression in Leads V5, V2, and II	ST-Segment Integral	Most ST60 Depression in Leads V5, V2, or II	ST60 in V5 at 3 Minutes of Recovery
0.82	0.65	0.68	0.67	0.67	0.67	0.69
61	42	44	38	42	40	43
0.60	−3.65	−0.002	−0.083 mv	−0.063	−0.051 mv	−0.051 mv

the studies, whereas METs and exercise-induced angina were significant predictors in our study but not in most other studies. This discrepancy is probably due to reduced workup bias. The equation based on clinical variables outperformed standard ST-segment criteria alone, and the addition of the treadmill responses, including ST-segment responses, further improved discrimination. Although more complicated computerized measurements had diagnostic power, the measurement of ST60 at 3.5 minutes of recovery served equally well. In addition to being simple to perform, this ST-segment measurement avoids the noise inherent in exercise.

In QUEXTA, investigators followed a strict protocol in accordance with published guidelines; similar care must be taken to obtain similar results in practice. A limitation of our study is the lack of women; no women were included because less than 2% of patients seen in the Veterans Affairs Medical Centers are female. However, previous studies have shown that sex-specific logistic regression equations can predict coronary artery disease in women (16). We also failed to remove all workup bias, and selection was still a problem because only 814 of all patients tested during the time period consented to participate. Finally, our study does not give the physician a testing strategy that will improve the poor sensitivity of the standard exercise test. When appropriate, stress tests with imaging should be considered, as should strategies for the standard exercise test that use multivariable techniques and probability thresholds (36). This latter approach uses multivariable equations listed in the American Heart Association/American College of Cardiology guidelines (2), demonstrating the portability of these equations to other populations, including women.

This study of a population with reduced workup bias shows that the real clinical strength of standard exercise electrocardiography lies in its high specific-

ity. For the diagnosis of angiographic coronary artery disease, computerized exercise ST-segment measurements were similar to visual ST-segment measurements and could be used to supplement a physician's interpretation, much like the commonly used resting electrocardiographic analysis programs (37). Although neither heart rate adjustment nor the Hollenberg score improved diagnostic classification compared with simpler measurements, equations including clinical variables and exercise test results showed the greatest discrimination and significantly improved the diagnostic power of the standard exercise test. These equations permitted the correct classification of 5 more patients of every 100 that received exercise testing in this population with a 50% prevalence of coronary artery disease. The number of additional patients correctly classified will increase and decrease with prevalence.

Appendix 1

To further validate our results, we tested the performance of the variables and the models in the diagnostic subset of the patients in the pilot study who had computer data. The pilot study was done to help test the technology required and to assess the recruitment rate at two study centers. The prevalence of significant angiographic coronary artery disease was similar to that in the data set. Similar results were obtained for discrimination (reflected by the ROC area and the sensitivities at matched specificity), but the calibration was different. The cut-off criterion for probability of coronary disease, using the prediction equations that contained ST-segment variables set to achieve 85% specificity, was roughly 60% in QUEXTA and 70% in the pilot study. The sensitivity (65%) and specificity (69%) for 1 mm of ST-segment depression in the pilot study are more like the results of a meta-analysis (mean sensitivity, 67%; mean specificity, 72%) based on studies with workup bias (15). The same

Appendix Table 1. Odds Ratios for the Variables Chosen in the Models*

Variable	Odds Ratio (95% CI)†				
	Model 1	Model 2	Model 3a	Model 3b‡	Model 4‡
Baseline odds	0.03	0.26	1.19	0.63	1.10
Age	1.06 (1.02–1.11)	1.05 (0.99–1.12)	1.03 (0.97–1.10)	1.04 (0.97–1.11)	1.03 (0.97–1.10)
Diabetes	1.73 (1.01–2.98)	1.76 (0.99–3.14)	1.78 (0.98–3.25)	1.77 (0.97–3.26)	1.81 (1.0–3.29)
Chest pain type, other	0.45 (0.29–0.68)	0.60 (0.38–0.95)	0.56 (0.35–0.92)	0.64 (0.39–1.04)	0.59 (0.36–0.97)
Chest pain, not at rest	1.78 (1.18–2.69)	1.58 (1.27–1.97)	1.33 (1.06–1.67)	1.14 (1.13–1.75)	1.33 (1.06–1.68)
Pack-years	1.01 (1.00–1.11)	1.00 (1.00–1.11)			
Duke Angina Index score 1		1.62 (1.10–2.57)	1.31 (0.80–2.13)	1.31 (0.81–2.14)	1.27 (0.78–2.06)
Duke Angina Index score 2		2.17 (1.10–4.27)	1.50 (0.72–3.11)	1.83 (0.89–3.75)	1.54 (0.74–3.20)
METs (per unit MET)		0.91 (0.69–1.20)	0.91 (0.73–1.13)	0.90 (0.73–1.12)	0.91 (0.83–0.99)
Maximal heart rate (per 10 beats/min)		0.92 (0.98–1.0)	0.88 (0.98–1.0)	0.92 (0.98–1.0)	0.89 (0.17–18.2)
Visual abnormal ST-segment depression (per mm of ST-segment depression)			2.24 (0.31–16.4)		1.78 (0.17–18.2)
ST60 in V5 at 3.5 minutes of recovery (per –10 mv)				2.11 (0.52–8.56)	1.46 (0.03–6.2)

* MET = metabolic equivalent.

† Odds of having coronary disease with the variable abnormal versus normal.

‡ ST60 in V5 at 3.5 minutes of recovery was the only computerized variable considered.

Appendix Table 2. Values in Pilot Study Group Derived by Using Unsimplified Equations at the Specificity Matching Visual Analysis (69%) in the Pilot Group and the Specificity Matching Visual Analysis in the QUEXTA Test Set (85%)*

Variable	Visual Analysis ST-Segment	Clinical Variables	Clinical and Hemo-dynamic Variables	Clinical and Hemo-dynamic Variables and Visual ST-Segment Analysis	Clinical, Hemo-dynamic, and Computerized ST-Segment Variables	Clinical, Hemo-dynamic, Visual, and Computerized ST-Segment Variables	Hollenberg Score	ST/HR Index	Sum of ST60 Depression in Leads V5, V2, and II	ST-Segment Integral	Most ST60 Depression	ST60 in V5 at 3 Minutes of Recovery
ROC curve	0.70	0.73	0.77	0.80	0.80	0.80	0.72	0.72	0.77	0.73	0.77	0.75
Sensitivity, %	65	66	71	72	75	77	62	69	70	65	67	65
Cut-off for 69% specificity	-0.1 mv	0.54	0.51	0.55	0.51	0.52	-2.92	-0.002	-0.56 mv	-0.054	-0.041 mv	-0.043 mv
Cut-off for 85% specificity	-0.2 mv	0.67	0.67	0.69	0.7	0.7	-5.69	-0.0036	-0.118 mv	-0.087	-0.071 mv	-0.082 mv

* The pilot study group comprised 321 patients, 158 of whom had coronary artery disease. The pilot group had specificity set at 1) value obtained by using visual analysis criterion of 1 mm of horizontal or downsloping ST-segment depression at any time during exercise or recovery (69%) and 2) specificity of 85% to demonstrate the difference in calibration between the QUEXTA group and the pilot group; sum of the depression at ST60 in II, V5, and V2; the most ST60 depression in the three leads; the time area in recovery of the slope and ST60 for V5; heart rate index (ST60 V5); and the measurement of ST60 V5 at 3.5 minutes of recovery. HR = heart rate; QUEXTA = Quantitative Exercise Testing and Angiography; ROC = receiver-operating characteristic.

equations were used in the pilot study and the QUEXTA training set. Different variables would have been chosen if we had developed new models in the pilot group. The differing calibration seems to be due to a higher prevalence of abnormal resting electrocardiograms and abnormal exercise test results in the pilot group compared with the QUEXTA group. In the QUEXTA group, 29% of patients had abnormal ST-segment depression (≥ 1 mm in exercise or recovery with abnormal slope, horizontal or downsloping) compared with 47% of patients in the pilot group (152 of 321).

Appendix 2

The following Veterans Affairs Medical Centers and persons participated in this study or the pilot study.

Chairman's Office, Veterans Affairs Medical Center, Palo Alto, California: Victor F. Froelicher, MD (*Co-Chair*), Tianna Umann (*Electrocardiographic Data Assistant*). Veterans Affairs Medical Center, Seattle, Washington: Kenneth G. Lehmann, MD (*Co-Chair*), Mimi Platt (*Administrative Assistant*).

Augusta, Georgia: Marandapalli Sridharan, MD, Christopher Pallas, MD, Horace Killam, MD (past), Anita Wylds, and Pat Orander. Birmingham, Alabama: Gilbert Perry, MD, Sriram Iyer, MD, Barbara Sanders, and Scottie Wilkins. Brockton, Massachusetts: Thomas P. Rocco, MD, Daniel Pietro, MD, Diane Lapsley, and Terry Fortin. Buffalo, New York: Eli Farhi, MD, Avery Ellis, MD, PhD, Linda Sherer, and Rose Marie Liedke. Denver, Colorado: Doug Morrison, MD, Stephen Crowley, MD; Jennifer Mignoli, Donna Lemaster (past), Victoria Hall (past), and Karen Babcock. Durham, North Carolina: Kenneth Morris, MD, Mitchell Krucoff, MD, Susan Daughtery, Suzanne Crater (past). Houston, Texas: Alfredo Montero, MD, Nadir Ali, MD, Mohamed Jeroudi, MD, Tracy Ferrando, and Jacklon Hicks. Milwaukee, Wisconsin: Virinderjit Bamrah, MD, Felix Tristani, MD, Gloria Luckett, and Christian Hanson. Oklahoma City, Oklahoma: Udho Thadani, MD, Elliott Schechter,

MD, John Turner, and Janet Davis. Portland, Oregon: Greg Larsen, MD, George Giraud, MD, Kathy Avalos, and Lisa Aizawa. San Diego, California: Ralph Shabetai, MD, William Penny, MD, Catherine Nielsen, RN, and Stacie Reynolds. Tucson, Arizona: Stephen Goldman, MD, Thomas Raya, MD, Susan Bigda, and E. Ray Holcombe.

Executive Committee: Victor F. Froelicher, MD (*Co-Chair*); Kenneth Lehmann, MD (*Co-Chair*); Charles Dennis, MD, Browns Mills, New Jersey; Robert Edson, MA, Palo Alto, California; Stephen Goldman, MD; Doug Morrison, MD; Ralph Shabetai, MD; and Ronald Thomas, PhD, La Jolla, California.

Data Monitoring Board: Paul Ribisl, PhD (*Chair*), Winston-Salem, North Carolina; Bernard Chaitman, MD, St. Louis, Missouri; Katherine Detre, MD, DrPH, Pittsburgh, Pennsylvania; Lars Ekelund, MD, PhD, Chapel Hill, North Carolina; Samuel Fox, MD, Mt. Desert, Maine; and William French, MD, Torrance, California.

Center for Cooperative Studies in Health Services: Philip Lavori, PhD (*Co-Director*); Rudolf Moos, PhD (*Co-Director*); Kathy Small (*Administrative Officer*, past); Carol Kerner; Ronald Thomas, PhD (*Biostatistician*, past); Robert Edson, MA; Bruce Chow (*Programmer*, past); Neil Shatz; Lenore Sheridan (*Statistical Assistant*); Winnie Koo (past); Jeannine Batmale (past); Johanna Macol (*Research Assistant*); and Felicitas Sibley (past).

Core Exercise Electrocardiography Laboratory: Jon Myers (*Coordinator*), Jeff Froning (*Programmer*).

Core Angiography Laboratory: Brad Bisson (*Coordinator*), Samantha Heath-Lange (*Programmer*).

Cooperative Studies in Health Services Central Administration: Daniel Deykin, MD (*Director*, past); Shirley Meehan, PhD, MBA (*Acting Director*); Janet Gold (*Administrative Officer*); J. Joseph Gough, MA (*Center for Cooperative Studies in Health Services Projects Officer*); and Carolyn Smith (*Staff Assistant*).

Requests for Reprints: Victor F. Froelicher, MD, Cardiology Division (111C), Palo Alto Veterans Affairs Health Care Systems, 3801 Miranda Avenue, Palo Alto, CA 94304.

Current Author Addresses: Drs. Froelicher, Lavori, and Myers and Mr. Edson, Mr. Do, and Mr. Froning: Palo Alto Veterans Affairs Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304.

Dr. Lehmann: Veterans Affairs Medical Center (111C), 1660 South Columbia Way, Seattle, WA 98108.

Dr. Thomas: University of California, San Diego, 9500 Gilman Drive, MC 0949, La Jolla, CA 92093-0949.

Dr. Goldman: Tucson Veterans Affairs Medical Center, Cardiology Division (111C), 3601 South 6th Avenue, Tucson, AZ 85723.

Dr. Morrison: Denver Veterans Affairs Medical Center, Cardiology Division, 1055 Clermont Street, Denver, CO 80220.

Dr. Dennis: Cardiology, Deborah Heart Center, 200 Trenton Road, Browns Mills, NJ 08015.

Dr. Shabetai: Veterans Affairs Medical Center, Cardiology Section (111A), 3350 La Jolla Village Drive, San Diego, CA 92161.

References

1. Fletcher GF, Froelicher VF, Hartley LH, Haskell WL, Pollock ML. Exercise standards. A statement for health professionals from the American Heart Association. *Circulation*. 1990;82:2286-322.
2. Gibbons RJ, Balady GJ, Beasley JW, Bricker JT, Duvernoy WF, Froelicher VF, 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;30:260-311.
3. Yamada H, Do D, Morise A, Atwood JE, Froelicher V. Review of studies using multivariable analysis of clinical and exercise test data to predict angiographic coronary artery disease. *Prog Cardiovasc Dis*. 1997;39:457-81.
4. Okin PM, Kligfield P. Heart rate adjustment of ST segment depression and performance of the exercise electrocardiogram: a critical evaluation. *J Am Coll Cardiol*. 1995;25:1726-35.
5. DelCampo J, Do D, Umann T, McGowan V, Froning J, Froelicher V. Comparison of computerized and standard visual criteria of exercise ECG for diagnosis of coronary artery disease. *Annals of Non-Invasive Electrocardiography*. 1996;1:430-42.
6. Philbrick JT, Horowitz SW, Feinstein AR. Methodological problems of exercise testing for coronary artery disease: groups, analysis and bias. *Am J Cardiol*. 1989;64:1117-22.
7. Myers J, Do D, Herbert W, Ribisi P, Froelicher VF. A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data. *Am J Cardiol*. 1994;73:591-6.
8. Mark DB, Hlatky MA, Harrell FE Jr, Lee KL, Califf RM, Pryor DB, et al. Exercise treadmill score for predicting prognosis in coronary artery disease. *Ann Intern Med*. 1987;106:793-800.
9. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA*. 1995;274:645-51.
10. Fletcher GF, Balady G, Froelicher VF, Hartley LH, Haskell WL, Pollock ML. Exercise standards. A statement for healthcare professionals from the American Heart Association. Writing Group. *Circulation*. 1995;91:580-615.
11. Froning JN, Froelicher VF, Olson MD. A real-time data-logger system using an optical disk WORM for archiving continuous 12-lead ECG data during exercise testing. *J Electrocardiol*. 1988;21:S141-8.
12. Froelicher VF, Myers J, Follansbee WP, Labovitz AJ. Exercise and the Heart. St. Louis: Mosby; 1993:48-69.
13. Herbert WG, Dubach P, Lehmann KG, Froelicher VF. Effect of beta-blockade on the interpretation of the exercise ECG: ST level versus delta ST/HR index. *Am Heart J*. 1991;122(4 pt 1):993-1000.
14. Ribisi PM, Liu J, Mousa I, Herbert WG, Miranda CP, Froning JN, Froelicher VF. Comparison of computer ST criteria for diagnosis of severe coronary artery disease. *Am J Cardiol*. 1993;71:546-51.
15. Gianrossi R, Detrano R, Mulvihill D, Lehmann K, Dubach P, Colombo A, et al. Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. *Circulation*. 1989;80:87-98.
16. 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.
17. Ascoop CA, Distelbrink CA, De Lang PA. Clinical value of quantitative analysis of ST slope during exercise. *Br Heart J*. 1977;39:212-7.
18. Simoons M. Optimal measurements for detection of coronary artery disease by exercise electrocardiography. *Comput Biomed Res*. 1977;10:483-99.
19. Simoons ML, Hugenholtz PG. Estimation of the probability of exercise-induced ischemia by quantitative ECG analysis. *Circulation*. 1977;56:552-9.
20. Detry JM, Robert A, Luwaert RJ, Rousseau MF, Brasseur LA, Melin JA, et al. Diagnostic value of computerized exercise testing in men without previous myocardial infarction. A multivariate, compartmental and probabilistic approach. *Eur Heart J*. 1985;6:227-38.
21. Deckers JW, Rensing BJ, Tijssen JG, Vinke RV, Azar AJ, Simoons ML. A comparison of methods of analyzing exercise tests for diagnosis of coronary artery disease. *Br Heart J*. 1989;62:438-44.
22. Detrano R, Salcedo E, Leatherman J, Day K. Computer-assisted versus unassisted analysis of the exercise electrocardiogram in patients without myocardial infarction. *J Am Coll Cardiol*. 1987;10:794-9.
23. Detrano R, Salcedo E, Passalacqua M, Friis R. Exercise electrocardiographic variables: a critical appraisal. *J Am Coll Cardiol*. 1986;8:836-47.
24. Pruvost P, Lablanche JM, Beuscart R, Fourrier JL, Traisnel G, Lombart F, et al. Enhanced efficacy of computerized exercise test by multivariate analysis for the diagnosis of coronary artery disease. A study of 558 men without previous myocardial infarction. *Eur Heart J*. 1987;8:1287-94.
25. Okin PM, Kligfield P. Heart rate adjustment of ST segment depression and performance of the exercise electrocardiogram: a critical evaluation. *J Am Coll Cardiol*. 1995;25:1726-35.
26. Lachterman B, Lehmann KG, Detrano R, Neutel J, Froelicher VF. Comparison of ST segment/heart rate index to standard ST criteria for analysis of exercise electrocardiogram. *Circulation*. 1990;82:44-50.
27. Morise AP, Duval RD. Accuracy of ST/heart rate index in the diagnosis of coronary artery disease. *Am J Cardiol*. 1992;69:603-6.
28. Bobbio M, Detrano R. A lesson from the controversy about heart rate adjustment of ST segment depression. *Circulation*. 1991;84:1410-3.
29. Rodriguez M, Froning J, Froelicher V. ST0 or ST60 [Editorial]. *Am Heart J*. 1993;126(3 Pt 1):752-4.
30. Morris CK, Myers J, Froelicher VF, Kawaguchi T, Ueshima K, Hideg A, et al. Nomogram based on metabolic equivalents and age for assessing aerobic exercise capacity in men. *J Am Coll Cardiol*. 1993;22:175-82.
31. Hollenberg M, Budge WR, Wisneski JA, Gertz EW. Treadmill score quantifies electrocardiographic response to exercise and improves test accuracy and reproducibility. *Circulation*. 1980;61:276-85.
32. Sheffield LT, Holt JH, Lester FM, Conroy DV, Reeves TJ. On-line analysis of the exercise electrocardiogram. *Circulation*. 1969;40:935-44.
33. Miranda CP, Liu J, Kadar A, Janosi A, Froning J, Lehmann KG, Froelicher VF. Usefulness of exercise-induced ST-segment depression in the inferior leads during exercise testing as a marker for coronary artery disease. *Am J Cardiol*. 1992;69:303-8.
34. Lachterman B, Lehmann KG, Abrahamson D, Froelicher VF. "Recovery only" ST-segment depression and the predictive accuracy of the exercise test. *Ann Intern Med*. 1990;112:11-6.
35. Berman JA, Wynne J, Mallis G, Cohn PF. Improving diagnostic accuracy of the exercise test by combining R-wave changes with duration of ST segment depression in a simplified index. *Am Heart J*. 1983;105:60-6.
36. Do D, West JA, Morise A, Atwood E, Froelicher V. A consensus approach to diagnosing coronary artery disease based on clinical and exercise test data. *Chest*. 1997;111:1742-9.
37. Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*. 1991;325:1767-73.