JACC Vol. 20, No. 5
November 1, 1992:1187-96

Development and Validation of a Logistic Regression–Derived Algorithm for Estimating the Incremental Probability of Coronary Artery Disease Before and After Exercise Testing

ANTHONY P. MORISE, MD, FACC, ROBERT DETRANO, MD, PHD,* MARCO BOBBIO, MD,† GEORGE A. DIAMOND, MD, FACC‡

Morgantown, West Virginia; Torrance and Los Angeles, California; Turin, Italy

Objectives. Our goals were to develop and validate a multivariate algorithm for estimating the incremental probability of the presence of coronary artery disease.

Background. Multivariate methods, including logistic regression analysis, have been extensively applied to diagnostic exercise testing. However, few previous studies have included both an incremental design and external validation.

Methods. A retrospective collection of clinical, exercise test and catheterization data was performed involving four U.S. referral medical centers. All patients had no prior history of coronary disease and had undergone coronary angiography ≤3 months after exercise stress testing. An algorithm was developed in one center (590 patients with a 41% prevalence of coronary artery disease) with the use of logistic regression analysis and was validated in the other three centers (1,234 patients, 70% prevalence). The algorithm incorporated pretest variables (age, gender, symptoms, diabetes, cholesterol), exercise electrocardiographic (ECG) variables (mm of ST segment depression, ST slope, peak heart rate, metabolic equivalents [METs], exercise angina) and one thallium variable. Discrimination was measured with receiver operating characteristic curve analysis. Calibration (that is, reliability) was assessed from a comparison of probability estimates and the actual prevalence of disease.

Results. The overall incremental receiver operating characteristic curve areas for the validation group were pretest, −0.738 ± 0.016; postexercise ECG, 0.78 (SE 0.017); and postthallium, 0.82 (SE 0.016); $p < 0.01$ for both increments. Within the three validation institutions, the institution with a disease prevalence closest to that of the derivation institution had the best incremental receiver operating characteristic curve areas. There was a stepwise incremental improvement in calibration especially from exercise ECG to thallium testing.

Conclusions. An incremental multivariate algorithm derived in one center reliably estimated disease probability in patients from three other centers. The incremental value of testing was best demonstrated when the derivation and validation groups had a similar disease prevalence. This algorithm may be useful in decision making that relates to the diagnosis of coronary disease.

(J Am Coll Cardiol 1992;20:1187–96)

A clinical diagnosis is usually accomplished by the progressive accumulation of data that either support or refute the diagnosis. Clinicians use the incremental weighing of facts as one means of making clinical diagnoses. When facts are concordant, clinical diagnoses are frequently not difficult to make. When they are discordant, clinical diagnoses tend to be uncertain, forcing a decision maker to use higher level testing. When coronary artery disease is suspected, this process usually begins with an appraisal of the clinical history and physical examination and often progresses to rest and exercise electrocardiography or other noninvasive imaging methods. This incremental process frequently culminates in coronary angiography.

Because many factors relate to the diagnosis of coronary disease, multivariate methods have been extensively applied to the diagnostic evaluation process (1–20). A minority of the reported studies validated their results in separate populations (3,5,7,8,10,14,17,19,20) and fewer still in populations outside the original institution (5,10,17). Only one of these studies (5) used logistic regression analysis and an incremental approach. The purpose of this study was twofold: 1) to develop an incremental multivariate analytic approach to the clinical diagnosis of coronary artery disease that allows for estimation of the probability of coronary disease both before and after exercise testing, and 2) to validate the method in a

From the Section of Cardiology, Department of Medicine, West Virginia University School of Medicine, Morgantown, West Virginia; *Division of Cardiology, Department of Medicine, St. Johns Cardiac Research Center, Torrance, California; †Division of Cardiology, University of Turin, Turin, Italy, and the ‡Division of Cardiology and Department of Medicine, Cedars-Sinai Medical Center, and the School of Medicine, University of California, Los Angeles, California. This study was supported in part by an Individual Investigator (RO1) grant from the Agency for Health Care Policy Research (HS-06065), Rockville, Maryland, by a Specialized Center of Research (SCOR) grant from the National Institutes of Health (HL-17651), Bethesda, Maryland, and by a Grant-in-Aid from the American Heart Association, West Virginia Affiliate.

Manuscript received February 19, 1992; revised manuscript received April 13, 1992, accepted April 22, 1992.

Address for correspondence: Anthony P. Morise, MD, Section of Cardiology, HSC-South, West Virginia University, Morgantown, West Virginia 26506.

0735-1097/92/$5.00

population drawn from several centers outside the derivation institution. The important aspects of the extent and prognostic impact of coronary artery disease were not evaluated in this study.

Methods

Derivation group. All patients referred to the stress laboratory at West Virginia University Medical Center between June 1981 and December 1989 were screened. Only those referred specifically for evaluation of the presence of coronary disease and who did not have a history of prior myocardial infarction or coronary arteriography were considered. Patients who met these criteria and who underwent coronary angiography within 3 months after stress testing constitute our derivation group. Because this study involved only a review of data previously acquired for clinical indications, review by the Human Subjects Committee was not required. However, all data were handled to ensure patient confidentiality.

Baseline clinical information. Data were collected from patients in the derivation group during a preexercise test interview. The variables tabulated were patient age, gender, symptoms and history of current cigarette smoking and diabetes mellitus. Chest pain was classified according to the four categories of Diamond (21). Rest heart rate, systolic blood pressure, height and weight were also measured. If the patient consented, blood specimen was collected to measure total serum cholesterol. Body mass index was calculated by Weight(kg)/(Height[m])^2.

All rest electrocardiograms (ECGs) were categorized as completely normal, equivocal or abnormal. Equivocal ECGs had ST-T changes with no significant downward displacement of the ST segment (0.08 s from the J point when compared with the baseline between two PR segments for at least three cycles) that would render them difficult to interpret. Abnormal ECGs had ST-T changes that would make an exercise ST response uninterpretable. These included ECG patterns of left ventricular hypertrophy, left bundle branch block, digitalis effect, Wolff-Parkinson-White syndrome or other significant downward displacement of the ST segment. Patients with abnormal Q waves were excluded.

Exercise tests. Most patients exercised with the use of the Bruce treadmill protocol (10% used the Naughton treadmill or arm ergometer protocols). The following data were collected: peak exercise heart rate, systolic blood pressure and metabolic equivalents (METs) achieved, the presence or absence of exercise-induced angina, ST segment depression in mm, and ST segment slope were collected. The rate-pressure product was calculated by the formula: (Peak heart rate × Peak systolic blood pressure)/1,000. Exercise-related ST segment changes were measured 0.08 s after the J point and were compared with the baseline between two PR segments. Peak exercise ST segment slope was qualitatively categorized as upsloping, horizontal or downsloping. All studies were read by one of the authors (A.P.M.) in blinded manner. Positive ST segment criteria consisted of ≥1-mm horizontal or downsloping ST depression 0.08 s after the J point for three consecutive cycles. If the ST segments were upsloping, ≥1.5 mm was necessary to define positivity. Any episodes of three or more consecutive ventricular premature beats were recorded as positive for exercise-induced ventricular arrhythmias. A decrease in systolic blood pressure ≥10 mm Hg below the maximal level was recorded as exertional hypotension.

Data were also compiled from those patients who underwent concurrent thallium scintigraphy. All thallium scintigraphic studies were read by one of us (A.P.M.) in blinded manner. Thallium studies were categorized as showing reversible, fixed or no defects. Other thallium variables, such as lung/heart activity ratio and number of thallium defects, were not considered because comparable data were not available for the validation group.

Coronary angiography. Angiograms were read by two cardiologists in blinded fashion and differences were resolved by consensus. Coronary artery disease was defined as the presence of at least one vessel with ≥50% lumen diameter narrowing.

Validation group. This group consisted of 1,234 patients with suspected coronary disease who underwent exercise electrocardiography and coronary angiography (778 underwent thallium scintigraphy). These patients were assembled for the purpose of validating other testing modalities and were therefore suitable for validating our algorithm. The group included patients from the Cleveland Clinic, the Veterans Administration Medical Center in Long Beach, California and the Cedars-Sinai Medical Center in Los Angeles, California.

Development of multivariate algorithm. We used logistic regression analysis and NCSS software (Number Cruncher Statistical System, Version 5.3, 1988) to develop the multivariate algorithm. We chose variables with good multivariate correlations ($p < 0.01$) and included additional variables because of their use in current clinical practice. We sought to include no more than one variable for every 10 outcomes (22) and to allow for incorporation of non-ECG exercise variables for patients without interpretable rest ST segments. We allowed for the separate determination of preexercise (pretest), postexercise ECG and postthallium probabilities to reflect the actual flow of clinical decision making.

Methods of statistical comparison and analysis. *Discrimination (or resolution).* Discrimination defines the ability of a method to separate (resolve) a group of patients into those with and without disease. The area under a receiver operating characteristic curve is an index of discrimination that ranges from 0 to 1. This single index indicates the probability that a method will perfectly resolve a group into diseased and nondiseased subgroups. With the probability data, the methods of Hanley and McNeil were used both to generate the areas under the curve (23) and to compare them (24). The receiver operating characteristic curve areas are expressed as area ± SE.

Calibration (or reliability). For probability estimates, the results should mean what they say. In other words, a

Table 1. Clinical Population Data

	Referral Group	Derivation Group	Male Subjects	p Value (male vs. female)	Female Subjects	AllvVal	Validation Group
Subjects	1,888	590	326		264		1,234
Age (yr)	50 ± 13	54 ± 12	54 ± 12	*	56 ± 11	*	59 ± 10
Men	1,076 (57)	326 (55)	100		—	†	1,015 (82)
Symptoms (1 = no pain to 4 = angina)	2.5 ± 0.9	2.8 ± 0.9	2.8 ± 0.9	NS	2.8 ± 0.9	*	3.0 ± 1
Cholesterol (mg/dl)	239 ± 56 (n = 852)	244 ± 57 (n = 435)	233 ± 48 (n = 236)	†	256 ± 65 (n = 199)	NS	236 ± 54 (n = 732)
Rest ECG							
Normal	1,512 (80)	409 (69)	232 (71)	NS	177 (67)	†	601 (49)
Uninterpretable	156 (8)	68 (12)	36 (11)	NS	32 (12)	†	412 (33)
Diabetes	227 (12)	130 (22)	68 (21)	NS	62 (23)	NS	198 (16)
Positive ST changes	384 (22) (n = 1,732)	203 (39) (n = 522)	114 (39) (n = 290)	NS	89 (38) (n = 232)	NS	381 (54) (n = 705)
Peak HR (beats/min)	151 ± 25	141 ± 23	140 ± 24	NS	142 ± 23	NS	140 ± 25 (n = 585)
Peak METs	7.6 ± 3	6.4 ± 3	7.2 ± 3	†	5.5 ± 2	†	7.9 ± 3 (n = 1,032)
Exercise angina	428 (23)	243 (41)	124 (38)	NS	119 (45)	†	526 (51) (n = 1,027)
Abnormal thallium test	127 (22) (n = 583)	92 (45) (n = 206)	52 (51) (n = 101)	NS	40 (38) (n = 105)	†	555 (71) (n = 778)
CAD (%)	—	243 (41)	150 (46)	†	93 (35)	†	863 (70)

Data are presented as mean value ± SD or number (percent) of subjects. *$p < 0.05$. †$p < 0.01$. AllvVal = derivation versus validation p value; CAD = coronary artery disease; ECG = electrocardiogram; HR = heart rate; METs = metabolic equivalents; Val = validation group.

probability of 40% should mean that the disease prevalence is 40% when all patients with a probability of 40% are considered as a group. Visual comparisons of calibration were made by plotting deciles of pretest or posttest probability (x axis) versus actual disease prevalence (y axis). Statistical comparisons were made by generating chi-square goodness-of-fit statistics. Chi-square ratios (F statistic) were used to compare one incremental calibration with another.

Mean values (± 1 SD). These values were compared by *t* test. Comparison of proportions was made by a nonparametric two-sample test of proportions. p values > 0.05 were not considered significant.

Results

Derivation group. A total of 1,888 patients at West Virginia University Medical Center were considered for this study. The 590 patients who subsequently underwent angiography constituted the derivation group. Table 1 lists the clinical characteristics and differences of the referral, derivation and validation groups.

Bivariate and multivariate analysis. We evaluated 10 pretest, 9 exercise ECG and 3 thallium variables within the derivation group. The 10 continuous variables included age, cholesterol, body mass index, rest and peak heart rate and blood pressure, peak rate-pressure product and METs and mm ST depression. The 12 discrete variables were further classified into 5 categoric and 7 dichotomous (yes or no) variables. Categoric variables included gender, symptoms, rest ECG, ST slope and qualitative thallium defect. Dichotomous variables included smoking, diabetes, angina, decrease in blood pressure, arrhythmias, any thallium defect and any reversible thallium defect. Table 2 lists the results of bivariate analysis. All but four variables were good predictors ($p < 0.05$) of the presence of coronary artery disease.

Table 3 shows the results of incremental multivariate analysis of 10 pretest, 7 exercise ECG and 1 thallium test variable. We eliminated the variables of peak rate-pressure product and blood pressure because peak systolic blood pressure was a poor bivariate predictor of disease and peak heart rate was as good as rate-pressure product. Similarly, we eliminated any thallium and any reversible thallium defect because qualitative thallium defect was as good a predictor and it considered three categories. The variables pretest and posttest are probabilities (0 to 1). Pretest is a cluster variable generated by applying the logistic regression formula to the pretest variables noted in Table 3. Posttest is another cluster variable generated by applying the logistic regression formula to the exercise ECG variables (in addition to pretest) noted in Table 3. Pretest analysis indicated that age, gender, symptoms, diabetes and cholesterol were independent predictors ($p < 0.01$) of the presence of disease. Postexercise ECG analysis indicated that ST slope, peak heart rate, exercise angina as well as pretest were independent predictors. Postthallium test analysis indicated that qualitative thallium defect as well as posttest were both independent predictors.

Because mm ST depression was not an independent

Table 2. Bivariate Analysis

	Pretest (n = 435)			Exercise ECG (n = 522)			Thallium (n = 206)	
	Chi-Square	p Value		Chi-Square	p Value		Chi-Square	p Value
Age (yr)	70.7	0.00001	mm ST	37.5	0.00001	Qual def	8.0	0.005
Gender	6.9	0.008	ST slope	55.4	0.00001	Any def	8.0	0.005
Symptoms	11.6	0.0007	Peak HR	41.3	0.00001	Rev def	5.2	0.03
Smoking	0.6	0.46	Peak SBP	0.4	0.52			
Diabetes	7.3	0.007	Peak RPP	23.3	0.00001			
Cholesterol	12.5	0.0004	METs	10.4	0.001			
BMI	11.7	0.0006	Angina	23.8	0.00001			
Rest HR	10.4	0.001	SBP decrease	1.4	0.23			
Rest SBP	3.9	0.046	Arrhythmia	2.9	0.08			
Rest ECG	4.9	0.03						

BMI = body mass index; def = thallium defect; ECG = electrocardiogram; HR = heart rate; METs = metabolic equivalents; mm ST = mm of any ST depression; Qual = qualitative; Rev = any reversible; RPP = rate-pressure product; SBP = systolic blood pressure.

predictor of disease (Table 3) and the accuracy of exercise-induced ST changes is lower in women than in men, we examined whether the analysis would differ between men and women. The middle columns of Table 4 indicate that mm ST depression was, in fact, an independent predictor in men but not in women.

Derivation of algorithm. Table 4 has seven data columns, one for each separate regression analysis. The first two columns consider the pretest variables with and without cholesterol; the middle four consider the exercise ECG results separately for men and women and whether the rest ST segments were interpretable, and the last column considers the results of the thallium test. Table 4 also shows the beta estimate and intercept information representing the coefficients for the logistic regression model noted in the Appendix.

Validation of algorithm. The equations derived from the results in Table 4 were applied to our validation group, and Table 5 and Figures 1 and 2 show some of these results. The receiver operating characteristic curve areas for the three incremental probability estimates for all patients with available data were as follows: pretest (n = 1,234) -0.738 ± 0.016, postexercise ECG (n = 1,023), -0.78 ± 0.016, and postthallium test (n = 778), -0.82 ± 0.016. Each posttest receiver operating characteristic curve area was significantly greater than the preceding area ($p < 0.01$). Because the number of patients at each incremental evaluation was unequal, we performed a second comparison in 778 patients who had data to satisfy all three incremental levels. With this additional comparison (Fig. 1), there continued to be a steady significant ($p < 0.01$) incremental increase in the receiver operating characteristic curve area.

Figure 2 shows the calibration results for these 778 patients. The reliability of each decile of estimated probability was compared with the expected ranges (stairsteps) and with the other incremental levels (pretest and two posttest). Chi-square goodness-of-fit testing indicated that the postthallium test results were different from the postexercise ECG results. Visual inspection of Figure 2 indicates that this difference reflects an improvement in the calibration of the postthallium test results. All three incremental levels underestimated disease prevalence in the deciles representing the 20% to 70% range (bar above expected range). However, at decile ranges <20, the postthallium test results were better calibrated qualitatively and quantitatively (that is, the data were within the expected range) than were the postexercise ECG results.

Table 3. Incremental Multivariate Analysis

	Pretest (n = 435)			Postexercise ECG (n = 522)			Thallium Test (n = 206)	
	Chi-Square	p Value		Chi-Square	p Value		Chi-Square	p Value
Age (yr)	32.5	0.00001	Pretest	65.1	0.00001	Posttest	37.0	0.00001
Gender	14.1	0.0002	mm ST	2.1	0.15	Defect	10.2	0.001
Symptoms	11.2	0.0008	ST slope	13.3	0.0003			
Smoking	0	0.97	Peak HR	4.9	0.03			
Diabetes	9.3	0.002	METs	0.1	0.94			
Cholesterol	13.5	0.0002	Angina	4.5	0.03			
BMI	4.0	0.046	SBP decrease	0.8	0.36			
Rest HR	0.2	0.67	Arrhythmia	2.8	0.10			
Rest SBP	1.1	0.31						
Rest ECG	0.8	0.36						

Pretest = pretest probability; Posttest = posttest probability; other abbreviations as in Table 2.

Table 4. Final Multivariate Algorithm: Chi-Square Results

	Pretest			Postexercise ECG				Postthallium Test	
				Men		Women			
	Chol +	Chol −		RECG +	RECG −	RECG +	RECG −		
Age (yr)	47.2†	78.8†	Pretest probability	27.6†	34.8†	29.7†	35.8†	Posttest probability	37.0†
	(0.0757)	(0.0857)		(4.325)	(3.926)	(5.169)	(5.119)		(4.329)
Gender	17.1†	17.3†	mm ST	8.6†		1.2		Defect quality	8.7†
	(−0.971)	(−0.8111)		(0.6842)		(−0.3057)			(0.5857)
Symptoms	9.9†	14.3†	ST slope	5.7†		7.7†			
	(0.3962)	(0.4056)		(0.6619)		(0.9022)			
Diabetes	10.1†	10.8†	Peak HR	6.5*	4.5*	0.1	1.6		
	(0.7987)	(0.7357)		(−0.020)	(−0.0138)	(−0.0176)	(−0.0096)		
Cholesterol	14.1†		METs	0.1	0.1	0.4	0.1		
	(0.0077)			(−0.015)	(−0.0139)	(−0.0614)	(−0.0059)		
			Angina	1.7	6.6*	4.1*	3.8*		
				(0.3979)	(0.6695)	(0.7025)	(0.5935)		
Intercept	(−7.285)	(−6.041)		(−1.073)	(−0.2229)	(−3.556)	(−1.485)		(−2.969)

*$p < 0.05$. †$p < 0.01$. Numbers in parentheses are the beta estimates. CHOL + = cholesterol measured; CHOL − = cholesterol unmeasured; METs = peak metabolic equivalents achieved; RECG + = interpretable rest electrocardiogram; RECG − = rest electrocardiogram that could not be interpreted; other abbreviations as in Table 2.

Table 5 compares the results from the three contributing validation institutions. The Cleveland group most closely approximated the derivation group described in Table 1 with respect to age, percent men, percent abnormal exercise ST and thallium responses and, most important, the prevalence of coronary disease. The other two groups were significantly

Table 5. Comparison of Validation Institutions

	Cleveland Clinic	Long Beach VA MC	Cedars-Sinai MC
n	303	503	428
Age (yr)	54 ± 9	60 ± 9	60 ± 10
% men (mean ± SD)	68	98	74
% positive ST	40	74	59
	(n = 299)	(n = 137)	(n = 269)
≥1 mm ST depression			
Sensitivity/specificity	63/79	80/56	63/62
Predictive value (+/−)	71/72	90/39	90/24
% abnormal thallium	41	76	89
	(n = 301)	(n = 58)	(n = 419)
% coronary disease	46	75	81
Probability			
Pretest	45	60	52
After exercise ECG	46	72	57
After thallium	51	74	67
ROC—all			
Pretest	0.78 (76 − 79)	0.71 (69 − 72)	0.71 (69 − 72)
After exercise ECG	0.82 (81 − 83)	0.70 (68 − 72)	0.76 (74 − 78)
	n = 303	n = 292	n = 428
After thallium	0.86 (85 − 87)	0.68 (64 − 74)	0.77 (75 − 78)
	n = 301	n = 58	n = 419
ROC—all with thallium			
Pretest	0.78 (76 − 79)	0.67 (64 − 73)	0.71 (69 − 72)
After exercise ECG	0.82 (81 − 83)	0.71 (68 − 74)	0.76 (75 − 77)
After thallium	0.86 (85 − 87)	0.68 (64 − 74)	0.77 (75 − 78)

MC = Medical Center; ROC = receiver operating characteristic curve area; VA = Veterans Administration; other abbreviation as in Table 1.

JACC Vol. 20, No. 5
November 1, 1992:1187–96

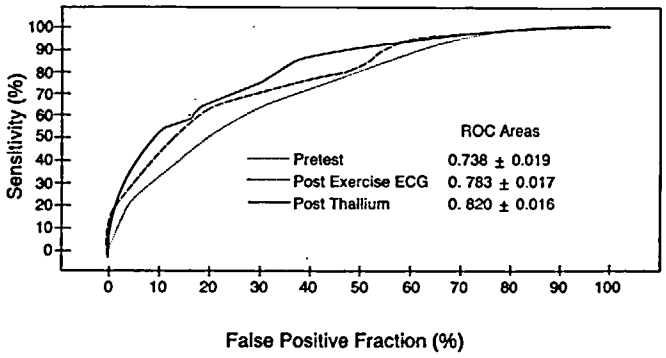


Figure 1. Receiver operating characteristic curves (ROC). Each curve was generated from probabilities derived from the incremental algorithms noted in Table 4. Each of the 778 patients used to generate these curves had sufficient clinical, exercise electrocardiographic (ECG) and thallium study data to satisfy each incremental equation. See text for further discussion.

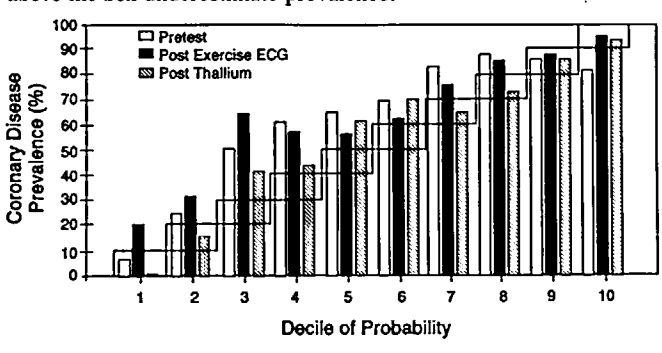
older and had a higher percentage of coronary disease, men and abnormal test results. The exercise ECG predictive values also reflect the differences in prevalence.

The lower half of Table 5 also reflects these differences in the algorithm results. The first group of receiver operating characteristic curve areas is meant to be compared horizontally, reflecting the results of all available patients at each institution with available data to calculate probabilities. The second set of curve areas is meant to be compared in a vertical fashion, reflecting the true incremental behavior at each institution. These comparisons are all consistent with the conclusion that the algorithm was more absolutely and incrementally discriminative in the lower prevalence group from Cleveland.

Discussion

Previous validation of other multivariate approaches has demonstrated reasonable diagnostic accuracy (5,7,8,14,20, 25–31). Different statistical methods have been used including bayesian (that is, independence-assuming) methods (10,15), linear regression or discriminant function analysis (2,3,4,6,11,14,17,18) and, as in the present study, logistic regression analysis (5,7–9,12,13,19,20). As the present study did not use independence-assuming methods, the subsequent discussion will deal only with the other multivariate approaches. A considerable number of previous validation studies have assessed such approaches in separate testing groups (5,7,8,14,20,25,26,28–30); however, only one study (25) used multiple validation institutions (with separate analyses of each) as did the present study.

Figure 2. Calibration (reliability). Data used to derive this graph are from the same 778 patients in Figure 1. **Vertical bars** represent the actual prevalence of coronary disease for each group of patients with a defined range of pretest, postexercise electrocardiography (ECG) and postthallium probability. Ascending stairsteps beginning at the lower left represent the predicted probability range for each decile. **Bar tops below the box** overestimate prevalence and **bar tops above the box** underestimate prevalence.



Characteristics of reported studies. Table 6 summarizes the characteristics of 16 previous studies that utilized multivariate methods to evaluate clinical or exercise variables, or both, concerning the diagnosis of coronary disease. Table 7 displays the variables that were evaluated in each of these studies as well as the independence of each as predictors of coronary disease presence. For the sake of comparison, the present study is represented in these two tables.

Gender. Table 6 indicates that nine previous studies evaluated results in both men and women and seven in either men or women. In studies of both men and women, >70% of subjects were male. Of nearly 12,000 patients evaluated, only 24% were women and 80% of these women were from two studies. Only 3 of 10 studies with data from women had at least 100 women in the sample. In addition, only four of nine studies analyzed data separately for men and women. Therefore, women are underrepresented in most of these previous studies.

Prevalence of disease. The prevalence of disease was >50% in 12 and >60% in 10 of 15 studies with available data. However, the definition of disease (percent stenosis on angiography) differed in many studies, that is, 70% to 75% versus 50% lumen diameter narrowing. The Coronary Artery Surgery Study (32) found that the prevalence of disease was 5% in both men and women who had lesions with 50% to 69% stenosis. Therefore, the prevalences in the studies using 70% lumen narrowing would not have been greatly increased with the more inclusive definition. Similarly, other studies have shown that the practical consequences of differing criteria on both the operating characteristics of exercise electrocardiography (33) and probability estimates of coronary disease (34) are small. All previous studies (except one in women only) that validated algorithms in separate populations used populations with a high prevalence of disease (≥55%).

We raise this issue of disease prevalence because of its relatively low prevalence (41%) in our derivation population. Although this prevalence is relatively low, it is likely to be higher than the prevalence in the clinical group of interest: patients with suspected coronary artery disease who come to the exercise laboratory before undergoing angiography. When we applied our algorithm to a large group of >4,500 patients from the Cedars-Sinai Medical Center, the mean posttest probability was 43%. When we applied another method of estimating posttest probability (10) to our own referral group of 1,888 patients, the mean posttest probabil-

Table 6. Comparison of Studies Using Multivariate Analysis of Clinical and Exercise Test Variables Pertaining to the Diagnosis of Coronary Artery Disease

Study (ref no.)	Pub Year	Sample Size	% Men	Analysis by Gender	% Stenosis	Prevalence	Validation Ref No.	Incremental
2	1979	405	?	N	50	72	—	N
3	1980	199	71	Y	70	57	26	N
4	1981	1,351	100	Y	70	65	—	Y
5	1982	3,840	71	N	70	65	5	Y
6	1982	141	82	N	70	78	—	N
7	1983	3,627	70?	Y	75	66	7	Y
8	1983	351	80	Y	—	—	8	N
9	1983	105	100	Y	50	62	—	Y
11	1984	92	0	Y	70	30	—	Y
12	1985	147	71	N	75	69	—	Y
13	1985	171	100	Y	70	68	—	Y
14	1985	253	100	Y	50	81	14	Y
17	1986	303	68	N	50	46	25	Y
18	1987	558	100	Y	50	56	—	N
19	1988	295	75	Y	50	70	28,29	Y
20	1991	135	0	Y	50	41	20	N
Present	1992	590	55	Y	50	41	Present	Y

(—) = data not available; N = no; Pub = publication; Ref No. = reference number; Y = yes; ? = not known.

ity was 27% in the entire group and only 18% in the group that did not undergo angiography. Our own data indicate that although the overall discriminant accuracy of our algorithm was good, it was better in the lower prevalence group from Cleveland. These considerations raise two questions: 1) Do previous studies using groups with a higher prevalence of disease adequately represent the real clinical group? 2) Should groups with greatly differing prevalences be compared with each other? Results of the present study indicate that both answers would be no.

Incremental consideration. The interpretation of exercise test data within the context of other known clinical pretest data is an important and often underappreciated consideration (22). Therefore, incremental evaluations are necessary to determine the true clinical relevance of exercise test variables. Table 6 indicates that 11 of 16 previous studies

Table 7. Comparison of Variables Evaluated in Studies Using Multivariate Analysis as It Pertains to the Diagnosis of Coronary Disease

Study (ref no.)	Age	Gender	Sym	Smoke	DM	Chol	Rest ECG	mm ST	ST Slope	Ex Ang	WL	PHR	RPP	TL
2	N	Y	—	—	—	—	—	Y	Y	N	—	—	—	
3	N	M	—	N	N	—	N	Y	N	N	Y	N	—	—
	Y	W	—	N	N	—	N	N	N	Y	N	N	—	—
4	Y	—	Y	Y	N	—	N	Y	—	Y	N	Y	Y	—
5	Y	Y	Y	Y	N	Y	N	Y	—	N	N	N	N	—
6	N	Y	Y	—	—	—	—	—	Y	N	N	N	Y	Y
7	Y	Y	Y	Y	Y	Y	Y	—	—	—	—	—	—	—
8	Y	M	—	—	—	—	N	Y	—	N	Y	Y	N	—
	N	W	—	—	—	—	N	N	—	N	Y	Y	N	—
9	N	—	Y	—	—	Y	—	Y	—	N	Y	N	N	—
11	N	W	Y	—	—	—	—	N	—	N	N	N	N	Y
12	Y	Y	Y	—	—	—	—	Y	—	—	—	—	—	Y
13	N	M	Y	—	—	—	—	Y	—	N	N	Y	N	Y
14	—	M	Y	—	—	—	—	Y	Y	—	Y	Y	N	—
17	Y	Y	Y	—	—	—	—	Y	—	—	—	—	—	Y
18	Y	M	Y	—	—	—	—	Y	Y	Y	Y	Y	Y	—
19	Y	M	Y	Y	N	Y	—	Y	—	Y	Y	—	—	—
	Y	W	Y	Y	N	Y	—	Y	—	Y	Y	—	—	—
20	—	W	—	—	—	—	—	Y	Y	Y	Y	Y	Y	—
Present	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	N	Y

Chol = cholesterol; DM = diabetes mellitus; ECG = electrocardiogram; Ex Ang = exercise angina; M = men only; mm ST = ST depression; N = not a significant predictor; PHR = peak heart rate; Ref No. = reference number; RPP = rate-pressure product; Sym = symptoms; TL = thallium; W = women only; WL = workload; Y = significant predictor; (—) = variable not evaluated.

could be considered incremental. As in our study, most exercise test variables were good bivariate predictors of disease (Table 2). However, after multivariate analysis with an incremental adjustment, some variables were not independent predictors (Table 3). Notable among these was METs. Clinicians routinely use METs to assess exercise capacity. However, the consideration of METs appears to have little value in the diagnostic evaluation of suspected coronary disease. In contrast, the value of exercise capacity for assessing prognosis is well supported by published data (35).

The method of incremental analysis varied among studies. Most studies (4,9,11–14,19) performed separate stepwise analyses of pretest and pretest plus exercise variables by using the best variables from each analysis for their final regression equation. We chose to force all of the pretest variables into the exercise multivariate analysis by using a cluster variable. One previous study (36 reported only as an abstract) compared the results generated from raw variables with those of probability estimates based on those raw variables and found some slight but significant loss of accuracy. Although the use of the cluster variable may have resulted in some loss of discriminant accuracy, it reduced the complexity of the multivariate analysis and the final algorithm (Table 4).

Only 5 of the 11 incremental studies indicated a positive incremental value for exercise ECG variables. However, 3 of the 11 did not address this question.

Variables evaluated (Table 7). *Age.* Age was a good predictor in 10 of 15 studies. However, it was an independent predictor in all five studies with a sample size >500.

Gender, symptoms, cholesterol. When these three variables were evaluated, there was universal agreement that they were independent predictors, irrespective of sample size, definition of disease or total number of variables studied.

Smoking. Smoking was a good predictor in four of six studies and was an independent predictor in the three largest studies. In the present study smoking was not an independent predictor in either the bivariate or the multivariate evaluations. The definition of smoking differed in all studies. Our study used the Framingham definition (only current cigarette smoking) (37). In retrospect, this definition would exclude many former smokers and its use may explain why smoking was not a significant predictor in our study.

Diabetes. Diabetes was a good predictor in only two of six studies. Of the three largest studies, two failed to show that diabetes was an independent predictor and the third study showed a low impact on the final model (7). The definition of diabetes in most studies, including the present, was fairly simple, that is, the presence or absence of diabetes in the patient's clinical history. Others (10,37) have used glucose intolerance (defined by serum glucose levels) rather than history. Nevertheless, the significance of diabetes to the model in the present study appears to be considerable. The prevalence and definition of coronary disease did differ in our study from those of the other studies that evaluated diabetes. Additional studies in groups with a lower prevalence of disease using a more refined definition of diabetes and glucose intolerance may reveal the real status of diabetes as an independent variable.

ST depression. Among the exercise ECG variables, ST depression was a good predictor of coronary disease in 15 of 16 studies. As in the present study, two of three studies that evaluated ST depression separately in men and women found that ST depression was not a predictor in women. However, two other studies limited to women were divided on this issue.

ST slope. ST slope was a good predictor in six of seven studies. However, the definition of this variable was not uniform and only two studies included incremental considerations. ST slope was considered a categoric variable in the present study as in several earlier studies (2,3), whereas it was considered a continuous variable in others (6,14,18,20).

Non-ST exercise variables. The remaining non-ST exercise variables were good predictors in about 50% of studies, and incremental considerations did not improve on this ratio. Exercise capacity, usually defined as exercise duration or METs achieved, was found to be a good predictor in only three of eight studies with incremental considerations. Owing to a lack of uniformity in definition, patient groups and other variables considered, the incremental diagnostic importance of these exercise variables is unclear.

Thallium scintigraphy. All six studies (five of six incremental) that evaluated thallium scintigraphy found it to be a good predictor. The thallium test variable in the present study was a simple categoric one based on the presence of reversible, fixed or no defects. We did not consider other thallium test variables such as the number of segments with defects, the intensity of a defect, lung/heart activity ratio or transient dilation for reasons stated earlier. The thallium studies used in our derivation group represent the gamut of thallium technology from planar to single-photon emission computed tomographs with and without quantitation. Despite this limited variable information, there was a significant incremental increase in discrimination for the entire validation group. It is possible that with the consideration of more thallium variables, the incremental increase might have been greater.

Table 5 reveals that individual validation subgroups did not enjoy this same incremental increase. The worst receiver operating characteristic area was noted for the Long Beach group with only 58 patients. The Long Beach and Cedars groups combined had an 87% rate for abnormal thallium studies, whereas the Cleveland group had a rate of only 41%, similar to that of the derivation group (45%).

Summary. This study demonstrates that an incremental multivariate algorithm developed in a group with a lower prevalence of coronary disease is valid in a separate group with a higher prevalence. In addition, exercise electrocardiography added important diagnostic information even after the pretest data were considered. Similarly, the thallium data added important diagnostic information to the pretest and exercise ECG data. Each institutional analysis revealed that

discriminant accuracy was lower in those groups that had a significantly higher prevalence of disease than that of the derivation group, whereas it was high in a group with a similar prevalence. It is not unrealistic to conclude that the accuracy of an algorithm depends largely on the group from which it was derived.

Of equal clinical relevance is how well the derivation group represents the clinical group of interest. In 16 previous studies that involved multivariate analysis most of the derivation groups had a percentage of men and disease prevalence that are unlikely to adequately represent the clinical group of interest. Although referral bias was present in all of these catheterization groups, it was greater in the groups with a higher prevalence of disease. In other words, the degree of referral bias was probably greater in the two subgroups in this study in which discriminant accuracy was relatively lower. This bias was evidenced by the larger proportion of abnormal exercise ECGs and thallium tests in addition to the higher prevalence of disease. Although the prevalence of disease in our derivation group was higher than that we estimate for the group presenting for diagnostic exercise testing, it was lower than that in any published study. Unless all patients who come to the exercise laboratory undergo angiography, it is unlikely that prevalence rates <40% will be reported in future studies.

Comparison of 16 prior studies with our study reveals general agreement on the variables that are independent predictors of the presence of coronary disease, thus emphasizing that the present study yielded conclusions about these variables that are consistent with well reported data.

Clinical relevance. For the individual clinician, the practical clinical relevance of our study depends on one question: Is decision making improved when probability-estimating methods are used in conjunction with exercise (or any) testing? Although it is not our intent in this report to develop a complete response to this question, we firmly believe that the generation of probabilities both before and after exercise testing can be a useful adjunct in decision making. At the very least, a posttest probability offers a single numeric result that incorporates information from both clinical and test variables. It expresses quantitatively how close one is to 0 or 100%, that is, no disease or disease. We hesitate to draw specific (yet arbitary) decision thresholds and would leave it to the clinician to determine how low or high the probability must be before a change in management is undertaken. Previous surveys (38) show that there are physicians who want and use the information that these algorithms generate.

It is important to generate both pretest and posttest probabilities before the exercise test is actually performed. Estimating the potential effect of a negative test result on the current probability may allow a clearer appreciation of whether the exercise test can realistically answer the diagnostic question being addressed. The clinician also needs a clear understanding of whether the exercise test is being considered for diagnosis or prognosis in an individual patient. Pretest probability can help clarify this question by indicating whether the patient is in a low, intermediate or high pretest probability group. Noninvasive testing for diagnosis is best limited to patients with a low or intermediate probability of disease. These patients have the best chance of having a low posttest probability if the test results show a decrease in probability of disease. Similarly, patients with a high pretest probability should undergo noninvasive testing only to assess prognosis. These patients have little to no chance of having a low posttest probability even with completely negative test results. Three articles (22,38,39) discuss these considerations more extensively. We also caution that the present study investigated the incremental impact on diagnostic questions only; it did not address the impact on questions that relate to the extent of coronary disease or to the effect of disease on prognosis.

For those who want to use the method developed in this study, the Appendix provides details on how to generate probabilities. Clinicians should decide whether the derivation group used in this study is similar to their local group and evaluate the algorithm for themselves.

Appendix

Instructions for Evaluating the Logistic Regression Algorithm Presented

For those who wish to evaluate the logistic regression algorithm presented in this study, we offer the following instructions. The basic logistic model equation is as follows:

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

where a = intercept, b = beta estimate, x = variable value. The values for a and b are found on Table 4. Seven equations can be generated depending on the incremental level, the gender and the variables used. The pretest and posttest probability variables require prior calculation and are numbers between 0 and 1. The following is a list of entry variable values:

Continuous variables
- Age (yr)
- Total cholesterol (mg/dl)
- ST depression (mm)
- Peak heart rate (beats/min)
- METs
- Probability (0.0–1.0)
  - Pretest
  - Posttest

Categoric variables
- Gender (male = 0, female = 1)
- Symptoms (Ref 21)
  - Typical angina = 4
  - Atypical angina = 3
  - Nonanginal pain = 2
  - No pain = 1
- ST slope
  - Upsloping = 1
  - Horizontal = 2
  - Downsloping = 3
- Thallium defect
  - None = 1
  - Fixed = 2
  - Reversible = 3
- Dichotomous (yes = 1, no = 0)
  - Diabetes
  - Exercise angina

References

1. Berman JL, Wynne J, Cohn PF. A multivariate approach for interpreting treadmill exercise tests in coronary artery disease. Circulation 1978;58: 505–12.
2. Cohn K, Kamm B, Feteih N, Brand R, Goldschlager N. Use of treadmill score to quantify ischemic response and predict extent of coronary disease. Circulation 1979;59:286–96.
3. Greenberg PS, Cangiano B, Leamy L, Ellestad MH. Use of the multivariate approach to enhance the diagnostic accuracy of the treadmill stress test. J Electrocardiol 1980;13:227–36.
4. Fisher LD, Kennedy JW, Chaitman BR, et al. Diagnostic quantification of CASS clinical and exercise test results in determining presence and extent of coronary artery disease. Circulation 1981;63:987–1000.
5. Goldman L, Cook EF, Mitchell N, et al. Incremental value of the exercise test for diagnosing the presence or absence of coronary artery disease. Circulation 1982;66:945–53.
6. McCarthy DM, Sciacca RR, Blood DK, Cannon PJ. Discriminant function analysis using thallium-201 scintiscans and exercise stress test variables to predict the presence and extent of coronary artery disease. Am J Cardiol 1982;49:1917–26.
7. Pryor DB, Harrell FE, Lee KL, Califf RM, Rosati R. Estimating the likelihood of significant coronary artery disease. Am J Med 1983;75:771–80.
8. Kansal S, Roitman D, Bradley EL, Sheffield LT. Enhanced evaluation of treadmill tests by means of scoring based on multivariate analysis and its clinical application: a study of 608 patients. Am J Cardiol 1983;52:1155–60.
9. Currie PJ, Kelly MJ, Harper RW, et al. Incremental value of clinical assessment, supine exercise electrocardiography, and biplane exercise radionuclide ventriculography in the prediction of coronary artery disease in men with chest pain. Am J Cardiol 1983;52:927–35.
10. Diamond GA, Staniloff HM, Forrester JS, Pollack BH, Swan HJC. Computer-assisted diagnosis in the noninvasive evaluation of patients with suspected coronary artery disease. J Am Coll Cardiol 1983;1: 444–55.
11. Hung J, Chaitman BR, Lam J, et al. Noninvasive diagnostic test choices for the evaluation of coronary artery disease in women: a multivariate comparison of cardiac fluoroscopy, exercise electrocardiography and exercise thallium myocardial perfusion scintigraphy. J Am Coll Cardiol 1984;4:8–16.
12. Weintraub WS, Barr-Alderfer VA, Seelaus PA, et al. A sequential approach to the diagnosis of coronary artery disease using multivariate analysis. Am Heart J 1985;109:999–1005.
13. Hung J, Chaitman BR, Lam J, et al. A logistic regression analysis of multiple noninvasive tests for the prediction of the presence and extent of coronary artery disease in men. Am Heart J 1985;110:460–9.
14. Detry JMD, Robert A, Luwaert RJ, 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.
15. Melin JA, Wijns W, Vanbutsele RJ, et al. Alternative diagnostic strategies for coronary artery disease in women: demonstration of the usefulness and efficiency of probability analysis. Circulation 1985;71:535–42.
16. Hollenberg M, Zoltick JM, Go M, et al. Comparison of a quantitative treadmill exercise score with standard electrocardiographic criteria in screening asymptomatic young men for coronary artery disease. N Engl J Med 1985;313:600–6.
17. Detrano R, Leatherman J, Salcedo EE, Yiannikas J, Williams G. Bayesian analysis versus discriminant function analysis: their relative utility in the diagnosis of coronary disease. Circulation 1986;73:970–7.
18. Pruvost P, Lablanche JM, Beuscart R, 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.
19. Deckers JW, Tijssen JGP, Rensing BJ, Azar AJ, Simoons ML. Strategies for the diagnosis of coronary artery disease based on logistic and Bayesian analysis of clinical and exercise variables. In: Deckers JW, ed. Diagnostic and Prognostic Implications of Exercise Testing. Rotterdam: Erasmus University, 1988, 75–90.
20. Robert AR, Melin JA, Detry JR. Logistic discriminant analysis improves diagnostic accuracy of exercise testing for coronary artery disease in women. Circulation 1991;83:1202–9.
21. Diamond GA. A clinically relevant classification of chest discomfort. J Am Coll Cardiol 1983;1:574–5.
22. Diamond GA. Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. J Am Coll Cardiol 1989;14: 12A–22A.
23. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Diagn Radiol 1982;143:29–36.
24. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–43.
25. Detrano R, Janosi A, Steinbrunn W, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am J Cardiol 1989;64:304–10.
26. Greenberg PS, Bible M, Ellestad MH. Prospective application of the multivariate approach to enhance the accuracy of the treadmill stress test. J Electrocardiol 1982;15:143–8.
27. Morise AP, Duval RD. The estimation of post-test probability of coronary disease following exercise testing using the sequential application of two Bayesian methods. Am Heart J 1990;120:1292–7.
28. Deckers JW, Rensing BJ, Simoons ML, Roelandt JRTC. Diagnostic merits of exercise testing in females. Eur Heart J 1989;10:543–50.
29. Deckers JW, Rensing BJ, Tijssen JGP, Vinke RVH, Azar AJ, Simoons ML. A comparison of methods of analysing exercise tests for diagnosis of coronary artery disease. Br Heart J 1989;62:438–44.
30. Greenberg PS, Ellestad MH, Clover RC. Comparison of the multivariate analysis and CADENZA systems for determination of the probability of coronary artery disease. Am J Cardiol 1984;53:493–6.
31. Guppy KH, Detrano R, Abbassi N, Janosi A, Sandhu S, Froelicher V. The reliability of probability analysis in the prediction of coronary artery disease in two hospitals. Med Decis Making 1989;9:181–9.
32. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation 1981;64:360–7.
33. Sketch MH, Mohiuddin SM, Lynch JD, Zencka AE, Runco V. Significant sex differences in the correlation of electrocardiographic exercise testing and coronary angiograms. Am J Cardiol 1975;36:169–73.
34. Diamond GA. Probability of CAD. Circulation 1982;65:641–2.
35. Morris CK, Ueshima K, Kawaguchi T, Hideg A, Froelicher VF. The prognostic value of exercise capacity: a review of the literature. Am Heart J 1991;122:1423–31.
36. Pollack BH, Ferguson JG, Diamond GA. How important is the raw data? (abstr). Clin Res 1987;35:314A.
37. American Heart Association. Coronary Risk Handbook. Dallas, Texas, 1973:4.
38. Diamond GA, Pollock BH, Work JW. Clinical decisions and computers. J Am Coll Cardiol 1987;9:1385–96.
39. Patterson RE, Horowitz SF. Importance of epidemiology and biostatistrics in deciding clinical strategies for using diagnostic tests: a simplified approach using examples from coronary artery disease. J Am Coll Cardiol 1989;13:1653–65.