PROGNOSTIC VALUE OF EXERCISE TESTING FOR ISCHAEMIC HEART DISEASE*

R. J. SHEPHARD, MD, PhD

*School of Physical and Health Education, Dept. of Preventive Medicine and Biostatistics
Institute of Medical Sciences, University of Toronto and Toronto Rehabilitation Centre*

"The augmented two-step test, when negative, completely excludes the presence of coronary artery disease" (Master, 1972).

In some millennial age, clinical scientists may devise a true prognostic test, or battery of tests. At the conclusion of our investigations, we may be able to advise a patient with a reasonable degree of confidence that he or she will live in good health for a specified period such as 13½ years. But at present, we are still struggling to make a simple bimodal decision. Is a patient who shows no clinical manifestations of disease healthy, or (in the context of our conference) is there an increased risk of myocardial infarction and sudden death? Will the patient who has already sustained a non-fatal infarction follow a normal clinical course, or is there an increased risk of reinfarction and death?

The process can be illustrated with respect to one of the most commonly employed prognostic tests of the cardiologist — the appearance of a deep horizontal or downward sloping depression of the ST segment of the electrocardiogram during exercise. The prognostic ideal is illustrated in the upper panel of Fig. 1. Even during maximum effort, normal individuals would develop a depression averaging only 0.5 mm, and the scatter in this response would be sufficiently small to avoid any overlap with that group of patients where the prognosis is poor. Unfortunately, the actual situation is all too commonly as represented in the lower panel. A proportion of subjects with a normal prognosis show substantial ST segmental depression, and a proportion of patients

with a poor prognosis show an apparently normal response. A substantial fraction of those who are tested thus become wrongly classified.

Fig. 1: To illustrate ideal and actual separation of normal and abnormal.

Simoons and Hugenholz in 1977 have argued that it would be better to present results as a likelihood ratio, interpreting ECG changes relative to heart-rate specific curves. In fact, one probably needs age, sex and heart rate specific curves. However, it is unlikely this approach will prove acceptable to either the patient or the family physician. Both wish for a simple bimodal answer — is there, or is there not latent coronary disease that will affect prognosis adversely?

The epidemiologist may not be overly concerned by some misclassification of patients. If he wishes to specify

Correspondence to:

Dr. Roy Shephard
320 Huron Street
Toronto, Ontario
M5S 1A1

*Based in part on a paper presented to the Second World Congress on Cardiac Rehabilitation, Jerusalem, December, 1981.

the prevalence of abnormality in any given environment, the precision of his estimate can be up-graded by increasing the sample size. Other factors being equal, the accuracy of the estimate increases as the square root of the number of observations that are made. However, the clinician must advise the individual patient rather than the community, and he thus lacks the recourse of an increase in sample size. His main option is to adjust the prognostic criterion in such a way as to maximise both the sensitivity and the specificity of the test procedure.

SENSITIVITY

The *sensitivity* measures the completeness with which abnormal individuals are detected. The number of true positive tests is expressed as a percentage of the sum of true positive and false negative tests (Fig. 2). In the context of ST depression, the sensitivity can be increased by reducing the severity of the prognostic criterion. For example an ST depression of 1 mm rather than 2 mm can be accepted as abnormal. However, the penalty of such an increase in sensitivity is a parallel rise in the proportion of false positive tests results.

$$\% \text{ Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100$$

$$\% \text{ Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \times 100$$

Fig. 2: To illustrate calculation of sensitivity and specificity.

SPECIFICITY

The *specificity* measures the precision with which the normal is recognised. The number of true negative tests is expressed as a percentage of the sum of true negative and false positive tests (Fig. 2).

CHANGING THE TEST CRITERIA

The specificity of a test can be increased by setting a more severe criterion of abnormality, but the penalty is then a loss of sensitivity. For example, Martin and McConahay found in 1972 that an increase of the prognostic threshold from 1.0 to 1.5 mm of ST depression increased the specificity of the stress electrocardiogram from 89 to 100%, but at the same time the sensitivity was decreased from 62 to 48%. A nice judgement must thus be used in setting an appropriate dividing line between a normal and an abnormal test result.

The choice of test sensitivity should be influenced by such factors as the cost of a false positive diagnosis, the seriousness of a false negative diagnosis, the intended frequency of screening and the prevalence of disease

within the community. A false positive electrocardiogram is bad news! If explained injudiciously to the patient, it can turn a healthy person into a cardiac cripple, with much anxiety and needless restriction of activity. It may also expose the patient to an unnecessary coronary angiogram — a costly procedure with a significant mortality. How dangerous is a false negative stress test? It may encourage a patient to undertake excessively strenuous physical activity, with an increased risk of ventricular fibrillation and sudden death. It may also weaken the impact of lifestyle advice such as the need to reduce body fat or stop smoking. However, a moderate frequency of false negative tests is acceptable, particularly if the prognostic screening is to be repeated. If a middle-aged adult has an annual check-up, abnormalities missed at the first test will likely be detected at a subsequent examination. The prevalence of cardiac disease in the general population is quite low; the frequency of abnormal ECGs rises from near zero at age 30 to about 20% at age 65. Thus, a fairly high test sensitivity is needed in order to yield a useful number of positive results when ECG screening is applied to the general population.

REFERENCE CRITERIA FOR EXERCISE TESTS

The assessment of prognostic value implies some absolute reference criterion against which a test can be judged. Unfortunately, there is no simple and unequivocal indicator of a normal cardiac prognosis. Many of the published statistics relate the sensitivity and specificity of the exercise electrocardiogram to obstruction of the coronary arteries as seen at angiography rather than to the clinical prognosis as revealed by prospective study. Coronary angiograms are themselves somewhat fallible. Intra-observer variation from one reading to another is 8 to 37% (Detre et al, 1975), with an average reader to reader variability of 21% (Zir et al, 1975). One must thus suspect that some "false positive" results are a reflection upon the angiogram rather than the ECG (Blackburn, 1977). In a proportion of "false positive" diagnoses, the angiogram shows a minor degree of coronary obstruction (30-50%); this is not usually regarded as significant disease, but nevertheless it may be sufficient to cause ischaemia in some regions of the myocardium (Ellestad and Halliday, 1967). Finally, because of the risks of angiography, it is not always possible to check the validity of a negative exercise ECG response. In general, angiograms overstate the prognostic value of the ECG relative to assessments based upon a moderate period of clinical observation (Table I). This is perhaps to be expected. Even if there is severe three vessel disease at angiography, a patient can still die from an extraneous cause such as a traffic accident. If the period of clinical observation is extended, a certain number of false positive results are converted to true positives, but unfortunately there is a rather larger increase in the number of false negative tests.

TABLE I

Sensitivity and specificity of exercise electrocardiogram, measured relative to coronary arteriogram, and to subsequent development of ischaemic heart disease (for details of references not cited in bibliography, see Shephard (1982))

(a) Coronary Arteriogram

Author	Sensitivity %	Specificity %
Fletcher & Cantwell (1974)	74	90
McHenry & Morris (1976)	83	95
Goldschlager et al (1976)	64	93
Simoons (1976)+ (1977 - visual)	85 (51)	91 (95)
Dagenais & Irshad (1978)	64	90
Ellestad & Halliday (1977)**	89	37
Martin & McConahay (1972) 0.5 (1.0)	65 ^x (35)	83 ^x (100)
Bartel et al (1974)	65	91
Piessens et al (1974)	65	82
Linhart & Turnoff (1974)	85***	100
Rios & Burwitz (1974)	83***	90
Cohn (1972) 0.5 (1.0)	86 ^x (77)	73 ^x (88)
Average	72	86

(b) Onset of Ischaemic Heart Disease*

Author	Sensitivity %	Specificity %
Bruce (1969)	60	86
Aronow (1973)	75	77
Kattus (1971)	100	67
Cumming (1975)	58	75
Froelicher (1977) Normals	(29)	(98)
All	61	92
Bruce (1975)	29	91
Average	64	81

x Master test

* Sensitivity varies with duration of follow-up (2.5 to 7 years)

+ Computer-assisted readings 20 and 80 msec after J-point, Frank leads

** Population where 74% of patients had significant angiographic disease

*** Relatively small samples (57, 50 patients respectively)

Some authors have suggested that relatively good prognostic information can be obtained from the use of a simple Master two-step test. For instance, in 1972 Cohn and his associates reported a test sensitivity of 86% and a test specificity of 73% for a criterion ST depression of 0.5 mm. However, this fairly satisfactory result may have been due to the elimination of patients who failed to reach a heart rate of 110 beat/min. while stepping. Both the sensitivity and the specificity of the

stress ECG are improved by (i) carrying out near maximum rather than low intensity exercise, (ii) the use of multiple rather than single ECG leads and (iii) recording the electrocardiogram during and following exercise. Additional prognostic information can be derived according to Ellestad and Halliday in 1967 from the intensity of exercise at which ST depression is provoked, the contour of the ST segment, persistence of ST depression into the recovery period, and concordance of exercise and recovery records (Blackburn, 1977). However, even if a 12-lead ECG is recorded during exercise and recovery, with the test carried to 85% of maximum effort, the sensitivity relative to angiography is under 80%, while the specificity is around 90%.

Let us suppose that such a test is applied to a population of 1000 middle-aged adults, with a 5% prevalence of ischaemic heart disease. The results will be as illustrated in Table II. As 95 patients have a false positive test, and 10 a false negative, more than 10% of all patients are misclassified, and more than 70% of warnings of an adverse prognosis are incorrect. Surprisingly, this statement is quite compatible with reports that a positive stress test indicates a 2-15 fold worsening of prognosis as reported by Ascoop (1977); Cumming (1978) and Ellestad and Halliday (1977). Indeed, in the example of Table II, the average outlook is 25 times worse for those individuals with a positive stress test. Nevertheless, to avoid the heavy toll of false positive results it would be necessary to develop a test with a very much greater specificity. Even if we could develop a test with a specificity of 98%, at least a third of ostensibly healthy middle-aged patients would be warned wrongly of impending cardiac disease.

TABLE II

Likely results of testing 1000 middle-aged adults (prevalence of ischaemic heart disease 5%, test sensitivity 80%, test specificity = 90%)

Response to Stress Test	Normal 950 Cases	Abnormal 50 Cases	Total Sample 1000 Cases
Positive	95	40	135
Negative	855	10	865

$$\text{Prognostic Ratio} = \frac{\text{Positive}}{\text{Negative}} = \frac{29.6}{1.17} = 25.3$$

EXERCISE TESTS AND THE TEST POPULATION

A more practical alternative is to restrict the application of exercise stress tests to patients with a high-risk of ischaemic heart disease. In 1977 Paffenbarger found an 8.5 fold increase in the incidence of fatal heart attacks when he identified individuals with a combination of low daily energy output, heavy cigarette smoking and

high resting systolic blood pressure. If the prevalence of cardiac disease in the test population were increased 8.5 times through such a simple screening procedure, the sensitivity and specificity of the exercise electrocardiogram would become as illustrated in Table III. The total percentage of patients who are misclassified is increased slightly (to 14.3%), but less than 15% of adverse prognoses are now erroneous. Because of this effect of disease prevalence, the stress test is more effective in the elderly than in younger populations (Froelicher et al, 1973). It is particularly successful in identifying critical lesions, picking out as many as 92% of patients with main left stem disease (Dagenais and Irshad, 1978).

$$S_1 = \frac{TP}{TP + FN} \quad S_2 = \frac{FP}{TN + FP}$$

$$S_3 = \frac{TN}{TN + FP} \quad S_4 = \frac{FN}{TP + FN}$$

$$\text{Predictive Power of Positive Test} = \frac{S_1 (\text{Prevalence})}{S_1 (\text{Prevalence}) + S_2 (1 - \text{Prevalence})}$$

$$\text{Predictive Power of Negative Test} = \frac{S_3 (1 - \text{Prevalence})}{S_4 (\text{Prevalence}) + S_3 (1 - \text{Prevalence})}$$

TP = true positive, FP = false positive, TN = true negative, FN = false negative

Fig. 3: The probability of disease, calculated by Bayes theorem for a positive and for a negative test result.

TABLE III

Likely results of testing 1000 middle-aged adults (prevalence of ischaemic heart disease 42.5%, test sensitivity 80%, test specificity 90%)

Response to Stress Test	Normal 575 Cases	Abnormal 425 Cases	Total Sample 1000 Cases
Positive	58	340	398
Negative	517	85	602

BAYES THEOREM

A more formal analysis of the prognostic value of the exercise ECG is based on Bayes theorem (Fig. 3). This theorem allows calculation of predictive values; the predictive value of a positive test is that percentage of subjects with a positive result who are in fact abnormal, while the predictive value of a negative test is the percentage of subjects with a negative test who are normal. Notice that in populations with a low prevalence of ischaemic heart disease, the tossing of a coin is superior to exercise testing (Froelicher et al, 1973; Goldschlager

et al, 1976). As the disease prevalence in a sample increases, the information content of a positive test improves, but that of a negative test deteriorates (Table IV). With a 20% prevalence of disease, the baseline expectation of abnormality is 20%, so that the information yield of a positive response is 46.7%, while that of a negative response is 14.7%.

TABLE IV

Influence of disease prevalence upon the predictive value of positive and negative tests (in this example, we have assumed a sensitivity of 80% and a specificity of 90%). The numbers shown are calculated by an application of Bayes Theorem (Fig. 3)

Disease Prevalence %	Predictive Value of Positive Test %	Predictive Value of Negative Test %
1	7.5	99.8
2	14.0	99.5
5	29.6	98.8
10	47.1	97.6
20	66.7	94.7
50	88.9	81.8
75	96.0	60.0

One flaw in the direct application of Bayes theorem is that the sensitivity and specificity of the test are assumed to be independent of the severity of disease. However, in practice, the sensitivity of a stress ECG increases substantially if two or three vessels rather than a single vessel are affected by the disease process (Table V).

TABLE V

Influence of severity of coronary arterial disease upon sensitivity of exercise ECG. Based on data collected by Blackburn

Author	Sensitivity and Number of Vessels Involved		
	One Vessel Disease %	Two Vessel Disease %	Three Vessel Disease %
McHenry & Morris (1976)	60	91	100
Kaplan et al (1973)	46	73	87
Martin & McConahay (1972)	35	67	86
Average	47	77	91

POST INFARCTION

Among patients who have already sustained a non-fatal myocardial infarction, the appearance of the ST segment during exercise still has some prognostic value, although

the sensitivity and specificity of the test are even poorer than before infarction. Over an average follow-up of 36.5 months, Kavanagh et al in 1979 reported that a horizontal or downward sloping ST depression ≥ 2 mm at 75% of aerobic power had a sensitivity of 41% and a specificity of 78% in the prediction of future fatal and non-fatal cardiac episodes. The actual prevalence of recurrent infarction was 7.2% in their sample, so that 18.1% were misclassified on the basis of exercise electrocardiography. A negative test interpretation was correct in 94% of instances, but a positive test was incorrect in almost 88% of patients (Table VI).

TABLE VI

Likely results of testing 1000 patients following myocardial infarction (prevalence of recurrence 7.2%, test sensitivity 41%, test specificity 78%)

Response to Stress Test	Normal 928 Cases	Abnormal 72 Cases	Total Sample 1000 Cases
Positive	127	18	145
Negative	801	54	855

TEST CONDITIONS

These discouraging statistics refer to optimal test conditions. Unless considerable care is taken in the conduct of stress tests, the sensitivity and specificity will be much poorer. A combination of tachycardia and hyperventilation can cause a positive stress test response in patients who are not habituated to both the laboratory and the required procedures (Smokler et al, 1973). Lack of proper warm-up can also give rise to temporary ST depression in up to two thirds of middle-aged men who show a normal electrocardiographic response to steady-state treadmill exercise. Negative test results may also be erroneous if the required intensity of effort is not attained due to poor motivation, clinical concerns (excessive breathlessness, significant arrhythmia, falling systolic pressure), or extracardiac symptoms (such as leg pain or fatigue).

WOMEN

The statistics discussed thus far refer to men, and in women the proportion of false positive tests is much higher, for reasons discussed elsewhere (Shephard, 1978). Cumming, 1978 reported the frequency of "positive" ECGs in women as 6% at 20-29 years, 19% at 30-39 years, 27% at 40-49 years, 42% at 50-59 years and 61% at 60-69 years. Since relatively few women develop ischaemic heart disease, we must presume that the majority of those with "significant" ST depression had false positive results.

FALSE POSITIVES

Among potential reasons for a false positive test (Table

VII), we may list treatment with digitalis, procaine amide, quinidine, antihypertensives (reserpine, guanethidine) or antidepressants (phenothiazine), potassium loss due to use of diuretics, cigarette smoking, glucose and carbohydrate loading, resting ST segmental depression, conduction abnormalities such as a left ventricular bundle-branch block and the Wolf-Parkinson White Syndrome, an abnormal increase of left ventricular stress (for example, early hypertension, increased sympathetic tonus, vasoregulatory asthenia or a combination of vigorous rhythmic exercise and isometric stress), left or right ventricular hypertrophy, valvular heart disease, congestive cardiomyopathy, and asymmetric septal hypertrophy.

TABLE VII

Factors leading to false positive and false negative exercise electrocardiographic changes

False Positive Results

Inadequate criterion (e.g. angiography)
Lack of habituation
Eating
Inadequate warm-up
Hyperventilation
Treatment with digitalis, procaine amide, quinidine, antihypertensives (reserpine, guanethidine) or antidepressants
Potassium loss due to diuretics
Cigarette smoking
Glucose and carbohydrate loading
Resting ST depression
Conduction abnormalities such as left ventricular bundle branch block and WPW syndrome
Abnormal stress on left ventricle (increased sympathetic tonus, altered myocardial contractility, vasoregulatory asthenia, isometric effort, early hypertension)
Left or right ventricular hypertrophy
Valvular disease
Congestive cardiomyopathy
Asymmetric septal hypertrophy
Myocarditis and pericarditis
Pectus excavatum
Inadequate equipment of lead system

False Negative Results

Inadequate exercise intensity
Use of propranolol, nitroglycerine or vasodilator drugs
Resting ST elevation, old infarction
Abnormalities of left or right ventricular conduction

FALSE NEGATIVES

A false negative test can arise from the use of propan-

olol, nitroglycerine and other vasodilator drugs, resting ST elevation, abnormalities of ventricular conduction, and an insufficient intensity of exercise (Pressens et al, 1974; Zohman and Kattus, 1977).

REPEATABILITY AND OBSERVER VARIATION

Further problems can arise from a lack of reliability in the ST segmental measurements. Bailey et al (1974) found that 4% of subjects who had been cleared for exercise by means of a standard stress test showed an abnormal ECG when the same observer repeated his observations a few days later. Cumming, 1978 found 18% of abnormal records disappeared with repeat testing, while in 1975 Bruce et al set the proportion of resolved abnormalities as high as 58%. Variability of interpretation is even greater if several observers are involved. Blackburn, 1969, handed a standard set of exercise ECG records to 14 expert cardiologists, and found the percentage of positive diagnoses that were reported ranged from 5-55%. Interestingly, technical personnel were more consistent readers of a given electrocardiogram than were the cardiologists. The technicians achieved complete inter-observer agreement in 85% of tracings, and in patients with more than 1 mm of ST depression they were unanimous concerning the presence or absence of this sign. The paramedical workers were equipped with magnifying lenses, and simple but standardised rules for interpreting the ECG records. Reliability of reporting is further improved if the observer checks his performance periodically against known ECG records, and a proportion of current tracings are sent to a reference laboratory for re-evaluation.

EXERCISE TESTS COMPARED WITH CLINICAL ASSESSMENT

Given the litany of difficulties limiting the reliability and validity of the exercise electrocardiogram, one must enquire whether any alternative procedure gives better prognostic information. Many of the available statistics are only crude estimates, because there is reluctance to carry out the criterion measurement (angiography) on apparently normal patients (Table VIII). Some of the authors cited have accepted as significant disease angiographic narrowing of less than 75%, or have regarded ST changes of < 1 mm as evidence of an ischaemic ECG response. Bearing in mind these limitations, clinical judgement often compares quite favourably with stress testing (Cumming 1978; Erikssen et al, 1981; Redwood and Epstein, 1972). Given a typical history of angina, the clinical report has a 90-95% agreement with coronary angiography, compared with 70-75% for the exercise ECG. If the angina is atypical, neither history nor exercise ECG are correct in more than 60% of instances. The exercise ECG is particularly likely to be negative in patients with Prinzmetal's variant angina (Fernandez et al, 1973). If the pain fails to meet clinical

criteria for angina, the clinician achieves 90-95% agreement with angiography, while the stress test has an accuracy of about 80%. With a more randomly selected population, the clinical assessment has a sensitivity of 81-91%, but a specificity of only 76%. Relative to the exercise ECG, clinical evaluation shows an increase of sensitivity at the expense of some loss of specificity. This has the disadvantage of reducing the predictive value of a positive test result.

TABLE VIII

Sensitivity and specificity of various alternative methods of evaluating ischaemic heart disease (angiographic criterion). Based in part on data collected by Zohman and Kattus

Test	Sensitivity %	Specificity %	Author
Clinical judgement	88-91	76	Kelemen et al (1973)*
Resting ECG ⁺ x	50	69	Cohn et al (1972)
Holter monitoring	91	78	Stern et al (1975)*
Atrial pacing	73	40	Kelemen et al (1973)*
Hypoxic ECG ⁺ x	73	33	Kassebaum et al (1968)*
Myocardial lactate production	55	87	Dagenais & Irshad (1978)
Scintigraphy ⁺ **	93	89	Berger et al (1979)
Scintigraphy ⁺	96	97	McCarthy et al (1979)
Scintigraphy	79	85	Bodenheimer et al (1979)
Scintigraphy	65-95	90	Johnstone et al (1980)
Scintigraphy	78-95	86-100	Berman et al (1977)

* See Zohman & Kattus for details of references

+ Less than 75% narrowing of coronary vessels accepted as a positive criterion

x ECG changes other than horizontal ST segmental depression also regarded as an ischaemic response

** Complex discriminant function used

OTHER METHODS

Other possible methods of detecting latent myocardial ischaemia include examination of the resting ECG, Holter monitoring, ECG recording during isometric hand-grip or hypoxia, atrial pacing, assessment of myocardial lactate production, Thallium-201 scintigraphy and radionuclide angiography. The most promising of these alternatives to date seem scintigraphy and radionuclide angiography, both being carried out immediately following vigorous exercise (Bailey et al, 1977; Berger et al, 1979; Johnstone et al, 1980). Although some reports have cited relatively high values for the sensitivity and specificity of these two tests, there remains a need to extend observations to larger, randomly selected samples of patients; to date, most of the results have been collected on patients with a high probability of coronary vascular disease.

Following myocardial infarction, the simple procedures available to the clinician are mostly of quite limited value in assessing prognosis (Table IX). The one observation that has a fair prognostic value is exercise compliance. Subjects who are unwilling to comply with an exercise-centred post-coronary rehabilitation programme have a 23 fold increase in the risk of recurrent disease.

TABLE IX

Sensitivity and specificity of various alternative methods of evaluating prognosis in ischaemic heart disease following a non-fatal myocardial infarction (criterion — 36.5 month incidence of fatal or non-fatal recurrence).

Based on data collected by Kavanagh et al

Test	Sensitivity %	Specificity %
Resting ECG abnormalities	98	3
Exercise non-compliance	52	99.5
Continued smoking	45	85
Angina (initial test)	43	69
Antiarrhythmic medication	41	70
Diuretic medication	27	84
Digitalis	25	88
Hypertension	23	77

IMPROVING EXERCISE TESTING — SERIAL TESTS AND DISCRIMINANT ANALYSIS

One technique that has been suggested to improve the sensitivity and specificity of pulmonary function testing is to use the patient as his own control; observations should begin while the patient is in good health, with repetition at intervals of perhaps one year. Any sudden deterioration of function is then brought to light. Such an approach could have some merit in exercise testing (Bruce, 1970; Doyle and Kinch, 1970), although dividends are likely to be smaller than in the pulmonary laboratory, because within subject variation is a major source of difficulty, and coronary narrowing often has a gradual rather than a sudden onset.

A further possible approach is to carry out a discriminant analysis, seeking by appropriate multivariate techniques an optimum combination of observations that will maximise the distinction between patients with normal and abnormal prognoses. We have already noted some suggestions that the sensitivity and specificity of exercise tests can be improved in this fashion. Simoons and Holgenholz in 1977 found only a small improvement from a single computer-assisted 80 msec post-J reading (90% sensitivity, 78% specificity) to multiple readings (optimum combination 20 and 80 msec post-J, 85% sensitivity, 91% specificity). Berman et al (1980) tested patients with chest pain; in this population, a

criterion based on summed R waves and ST depression had a predictive power of 98.6%, compared with 89.1% for ST depression alone. Siegel (1978) found that a combination of anginal pain, atrial gallop and ST depression had a specificity of 99%. Blackburn recommended that the discriminant function should include any fall of heart rate or blood pressure, and the concordance of ST segmental appearances with symptoms. Froelicher (1978) noted that serum cholesterol, resting blood pressure, serum triglycerides and maximum heart rate each helped in distinguishing true from false positive responses. Nevertheless, it still has to be established how far the ECG response (even with added variables) increases the prognostic yield relative to a consideration of age and such simple risk factors as cigarette smoking and obesity. In the "seven countries" study, the exercise ECG made a marginal additional contribution ($P < 0.05$) in only 1 of the 7 samples (Blackburn et al, 1970). Following infarction, specificity can be increased by a consideration of other risk factors such as hypertension and an elevation of serum cholesterol (Table X), but since relatively few patients demonstrate the required combination of findings, the sensitivity of a multivariate test falls to a very low level.

TABLE X

The influence of combinations of findings upon prognosis. Data for 610 patients sustaining a non-fatal myocardial infarction and followed for an average of 36.5 months

Variable	Sensitivity %	Specificity %
ST depression alone	41	78
ST depression + serum cholesterol elevation	9.0	98.2
ST depression + hypertension	6.8	93.1
Serum cholesterol elevation + hypertension	2.3	97.9

INDICTMENT

Cumming (1978) summarised his indictment of exercise testing by noting over 40 different systems of ECG leads, 50 different exercise protocols, a 40% inter-observer disagreement on test results, a list of at least 10 pathologies that yielded a positive test with 40% false positive and 30% false negative responses, a cost of at least \$50 per test, a time commitment of at least 30 minutes from a physician and a specialist technician, some mortality and a risk of litigation. Against this formidable list of criticisms, he found little evidence that the test result was influencing clinical decisions.

JUSTIFICATION

Ellestad and Halliday (1977) listed five possible justifications for exercise testing (Fig. 4). On the data we have

Discovery of latent disease
 Differential diagnosis of chest pain
 Prediction of exercise tolerance
 Prediction of coronary risk
 Motivation

Fig. 4: Justification for exercise ECG (Ellestad).

discussed above, their first suggestion (the detection of latent coronary arterial disease) must be rejected. Irrespective of the test outcome, the patient is likely to be told to take more exercise, to reduce body fat, to limit cholesterol intake, and to cease smoking. The exercise ECG also provides little rational basis for predicting future coronary events among groups such as symptomless airline pilots. It is highly debatable whether an exercise ECG helps materially in the differential diagnosis of chest pain. The use of a stress test to define a "safe" level of exercise where the patient is not exposed to deep ST segmental depression and multiple premature ventricular contractions is more widely accepted. Nevertheless, this indication can be questioned on two counts: (i) the criticisms of poor test reliability, false positive and false negative findings apply to exercise

prescription as much as to diagnosis and prognosis, and
 (ii) even in patients with a true positive response, the intensity of effort inducing this change varies with incidental details of the exercise undertaken (for example, environmental temperature, muscle groups activated and the relative proportions of rhythmic and isometric work).

It remains to be proven how far this last objection can be answered by relating exercise prescriptions to cardiac work-rate, as estimated from the pulse-pressure product (Ellestad and Halliday, 1977).

There remains the mystique of the laboratory stress test, with its associated motivational effects upon the physician and the patient. This aspect of medicine has received little formal examination. Some doctors may find in an exercise ECG a psychological prop to a shaky clinical diagnosis. For some patients, laboratory exercise may provide motivation to a change of lifestyle. Whether an equal psychological effect could be achieved by the use of a simpler, cheaper and safer protocol such as the "Canadian Home Fitness Test" has yet to be evaluated.

Finally, it must be stressed that both the computer-averaging of records and the multivariate approach to prognosis is in its infancy. Further application of such procedures may yet give to exercise testing the predictive value that was anticipated by pioneers in this field.

REFERENCES

- Ascoop, C. A., 1977 "What is an abnormal ischemic ECG response to exercise?" pp. 143-154. In: *Coronary Heart Disease, Exercise Testing and Cardiac Rehabilitation*. Ed. W. E. James and E. A. Amsterdam, Symposia Specialists: Miami, Fla.
- Bailey, D. A., Shephard, R. J., Mirwald, R. L. et al, 1974 "Current levels of Canadian cardiorespiratory fitness". *Canad. Med. Assoc. J.* 111: 24-30.
- Bailey, I. K., Griffith, L. S. C., Rouleau, J. et al, 1977 "Thallium-201 myocardial perfusion imaging at rest and during exercise". *Circulation* 55: 79-87.
- Bartel, A. G., Behar, V. S., Peter, R. H., Orgain, E. S. and Kong, Y., 1974 "Graded exercise stress tests in angiographically documented coronary artery disease". *Circulation* 49: 348-356.
- Berger, B. C., Watson, D. D., Taylor, G. J., Craddock, G. B., Martin, R. P. and Bellar, G. A., 1979 "Sensitivity of quantitative Thallium-201 scintigraphy following nondiagnostic exercise stress". *Circulation* 59/60 Suppl. II: 172.
- Berman, D. S., Amsterdam, E. A. and Mason, D. T., 1977 "Detection of myocardial ischemia by rest and exercise Thallium-201 scintigraphy". pp. 245-263. In: *Exercise in Cardiovascular Health and Disease*. Ed. E. A. Amsterdam, J. H. Wilmore and A. N. DeMaria. Yorke Books: New York.
- Berman, J. L., Wynne, J. and Cohn, P. F., 1980 "Multiple-lead QRS changes with exercise testing: diagnostic value and hemodynamic implications". *Circulation* 61: 53-61.
- Blackburn, H., 1969 "The exercise electrocardiogram: technological, procedural and conceptual developments". In: *Measurement in Exercise Electrocardiography*. The Ernst Simonson Conference. Ed. H. Blackburn, C. C. Thomas: Springfield, Ill.

Blackburn, H., Taylor, H. L. and Keys, A., 1970 "CHD in seven countries: the electrocardiographic prediction of 5 year CHD incidence among men aged 40-59". *Circulation* 41 Suppl. 1: 154.

Blackburn, H., 1977 "The exercise electrocardiogram in diagnosis". *Cardiology* 62: 190-205.

Bodenheimer, M. M., Banka, V. S. and Helfant, R. H., 1979 "Limitation of exercise Thallium in evaluation of patients with chest pain". *Circulation* 59/60 Suppl. II: 171.

Bruce, R. A., Pas, Y. L., Ting, N. et al, 1975 "Seven year follow-up of cardiovascular study and maximal exercise of Chinese men". *Circulation* 51: 890-901.

Cohn, P. F., Vokonas, P. A., Most, A. S. et al, 1972 "Diagnostic accuracy of two-step post-exercise electrocardiogram: results in 305 subjects studied by coronary arteriography". *J.Amer.Med.Assoc.* 220: 501-506.

Cumming, G. R., 1978 "Resolve that the exercise electrocardiogram is of limited diagnostic value". pp. 13-29. In: *Sports Medicine*. Ed. F. Landry and W. A. R. Orban, Symposia Specialists: Miami, Fla.

Dagenais, G. R. and Irshad, F., 1978 "The diagnostic value of exercise electrocardiogram: a five year experience". pp. 31-39. In: *Sports Medicine*. Ed. F. Landry and W. A. R. Orban. Symposia Specialists: Miami, Fla.

Detre, K. M., Wright, E., Murphy, M. L. et al, 1975 "Observer agreement in evaluating coronary angiograms". *Circulation* 52: 979-986.

Doyle, J. T. and Kinch, S. H., 1970 "The prognosis of an abnormal electrocardiographic stress test". *Circulation* 41: 545-553.

Ellestad, M. H. and Halliday, W. K., 1977 "Stress testing in the prognosis and management of ischemic heart disease". *Angiology* 28: 149-159.

Erikssen, J., Enge, I., Forgang, K. and Storstein, O., 1981 "False positive diagnostic tests and coronary angiographic findings in 105 presumably healthy males". *Circulation* 54: 371-376.

Fernandez, D., Rosenthal, J. E., Cohen, L. S. et al, 1973 "Alcohol-induced Prinzmetal variant angina". *Amer.J.Cardiol.* 32: 238-239.

Fletcher, G. F. and Cantwell, J. D., 1971. *Exercise and coronary heart disease*". C. C. Thomas. Springfield, Ill.

Froelicher, V. F., Yanovitz, F. G., Thompson, A. J. and Lancaster, M. C., 1973 "The correlation of coronary arteriography and the electrocardiographic response to maximal treadmill testing in 76 asymptomatic men". *Circulation* 48: 597-604.

Froelicher, V. F., 1977 "Use of the exercise electrocardiogram to identify latent atherosclerotic heart disease". pp. 188-208. In: *Exercise in Cardiovascular Health and Disease*. Ed. E. A. Amsterdam, J. H. Wilmore and A. N. DeMaria. Yorke Books: New York.

Goldschlager, N., Selzer, A. and Cohn, K., 1976 "Treadmill stress tests as indicators of presence and severity of coronary artery disease". *Ann.Intern.Med.* 85: 277-286.

Johnstone, D. E., Sands, M. J., Berger, H. J., Reduto, L. A., Lachman, A. S., Wackers, F. J. Th., Cohen, L. S., Gottschalk, A., Zaret, B. L. and Pytlik, L., 1980 "Comparison of exercise radionuclide angiography and Thallium-201 myocardial perfusion imaging in coronary artery disease". *Amer.J.Cardiol.* 45: 1113-1119.

Kaplan, M. A., Harris, C. N., Aronow, W. S., Parker, D. P. and Ellestad, M. H., 1973 "Inability of the submaximal treadmill stress test to predict the location of coronary disease". *Circulation* 47: 250-256.

Kavanagh, T., Shephard, R. J., Chisholm, A. W., Qureshi, S. and Kennedy, J., 1979 "Prognostic indices for ischemic heart disease patients enrolled in an exercise - centred rehabilitation programme". *Amer.J.Cardiol.* 44: 1230-1240.

- Linhart, J. W. and Turnoff, H. B., 1974 "Maximal treadmill exercise test in patients with abnormal control electrocardiograms". *Circulation* 49: 667-672.
- Martin, C. M. and McConahay, D. R., 1972 "Maximal treadmill exercise electrocardiography". *Circulation* 46: 956-962.
- Master, A. M., 1972 "Exercise testing for evaluation of cardiac performance". *Amer.J.Cardiol.* 30: 718-721.
- McCarthy, D. M., Sciacca, R. R., Blood, D. K. and Cannon, P. J., 1979 "Derivation and validation of a combined treadmill-scintiscan score to predict the presence of coronary artery disease". *Circulation* 59/60 Suppl. II: 174.
- McHenry, P. L. and Morris, S. N., 1976 "Exercise electrocardiography. Current state of the art". In: Schlant and Hurst: *Advances in Electrocardiography*. New York: Grune and Stratton.
- Paffenbarger, R., 1977 "Physical activity and fatal heart attack. Protection or selection?" In: *Exercise in Cardiovascular Health and Disease*. Ed. E. A. Amsterdam and J. H. Wilmore. Yorke Medical Books: New York.
- Piessens, J., Van Mieghem, W., Kestelost, H. and De Geest, H., 1974 "Diagnostic value of clinical history, exercise testing and atrial pacing in patients with chest pain". *Amer.J.Cardiol.* 33: 351-356.
- Redwood, D. R. and Epstein, S. E., 1972 "Uses and limitations of stress testing in the evaluation of ischemic heart disease". *Circulation* 46: 1115-1131.
- Rios, J. C. and Hurwitz, L. E., 1974 "Electrocardiographic responses to atrial pacing and multistage treadmill exercise testing: correlation with coronary arteriography". *Amer.J.Cardiol.* 34: 661-666.
- Shephard, R. J., 1982. *Physiology and Biochemistry of Exercise*. Praeger: New York.
- Shephard, R. J., 1978. *Physical Activity and Aging*. Croom Helm: London.
- Siegel, W., 1978 "New perspectives in clinical exercise testing". pp. 3-12. In: *Sports Medicine*. Ed. F. Landry and W. A. R. Orban. Symposia Specialists: Miami.
- Simoons, M. L., 1976. *Computer Assisted Interpretation of Exercise Electrocardiograms*. Bronder-Offset B. V.: Rotterdam.
- Simoons, M. L. and Hugenholz, P. G., 1977 "Estimation of the probability of exercise-induced ischemia by quantitative ECG analysis". *Circulation* 56: 552-559.
- Smokler, P. E., MacAlpin, R. N., Alvaro, A. et al, 1973 "Reproducibility of a multi-stage near-maximal treadmill test for exercise tolerance in angina pectoris. *Circulation* 48: 345-351.
- Zir, L. M., Miller, S. W., Dinsmore, R. E. et al, 1975 "Interobserver variability in coronary angiography" (Abstr.). *Circulation* 51/52 Suppl. II: 27.
- Zohman, L. R. and Kattus, A. A., 1977 "Exercise testing in the diagnosis of coronary heart disease: a perspective". *Amer.J.Cardiol.* 40: 243-250.