A critical evaluation of stress testing in the diagnosis of peripheral vascular disease

Kenneth Ouriel, M.D.,* Anne E. McDonnell, B.S., Charles E. Metz, Ph.D., and Christopher K. Zarins, M.D., Chicago, Ill.

We studied 218 patients (372 limbs) and 25 normal subjects (50 limbs) with resting ankle index (RAI), treadmill exercise (TE), and postocclusive reactive hyperemia (PORH) to determine whether diagnostic accuracy is improved through the use of stress testing. In addition, we studied 10 patients with stable claudication (20 limbs) to determine the reproducibility of the three measures. RAI was the most reproducible measure, with the smallest variance between testing days (P < 0.001). RAI differentiated between arteriographically diseased and normal limbs with a sensitivity of 97% and a specificity of 100%, whereas the corresponding values for TE were 97% and 96% and for PORH 89% and 96%. Recovery to baseline index was prolonged in the diseased group compared with normal (P < 0.001 for both TE and PORH), but this was of limited discriminative value. Receiver-operating characteristic curve analysis documented that RAI was as diagnostically useful as TE and that both were more valuable than PORH (P < 0.02). However, the routine addition of stress testing increased diagnostic yield by only 1.6% and cost $1100 for each limb correctly diagnosed through the addition of stress testing. RAI is a simple, accurate, and reproducible test. Routine stress testing is not cost effective, adding little diagnostic information to RAI, and it should be reserved for the small subset of symptomatic patients with normal RAI.

From the Departments of Surgery and Radiology, University of Chicago, Chicago, Ill.

RESTING ANKLE INDEX (RAI) is well accepted as a useful diagnostic measure to objectively assess the presence of peripheral vascular disease.21,25 Stress testing has been advocated by many to detect early disease in patients with normal pulses and resting measurements.† In such cases the narrowing of the vessels has not reached the so-called critical stenosis,12,13 and it is only during the poststress hyperemic state that the lesions become hemodynamically significant. The measurement of ankle index during these periods of increased blood flow to the extremity has been assumed to increase the sensitivity of the noninvasive diagnostic examination.

Presently, the most widely implemented mode of stress testing is treadmill exercise (TE). A second method is postocclusive reactive hyperemia (PORH), which involves the measurement of hemodynamic changes following a period of temporary limb ischemia through total circulatory arrest. Although many investigators recommend the use of stress testing as an adjuvant to resting determinations, few data exist regarding the comparative accuracy of RAI with and without poststress determinations. The purposes of this study were to document the accuracy and reproducibility of RAI, TE, and PORH and to determine if any increase in diagnostic accuracy is obtained through addition of stress testing to resting measurements alone.

METHODS

Patient selection. During a 3-year period, 218 consecutive patients with clinically evident lower extremity peripheral vascular disease were evaluated with stress testing in the vascular noninvasive laboratory at the University of Chicago (Fig. 1). Patients deemed suitable for treadmill testing were tested in this manner (100 patients), whereas patients unfit for exercise because of cardiopulmonary disease, amputation, or cerebrovascular accident underwent PORH


Reprint requests: Christopher K. Zarins, M.D., Department of Surgery, University of Chicago, 950 East 59th St., Chicago, IL 60637.

*Current address: Department of Surgery, University of Rochester, Rochester, N.Y.

†See references 1, 4, 8-10, 14, 19, 22, 24.
testing (118 patients). Arteriographic evaluation of the aorta and iliac and lower limb vessels was done in 70 (33%) of the patients, providing a standard for statistical analysis.

Patients that had previously undergone limb surgery were excluded from the study. In total, 372 limbs were included in the data analysis. These were divided into two groups on the basis of the presence or absence of symptoms. There were 305 symptomatic and 67 asymptomatic limbs. The symptomatic group was subdivided into two smaller groups on the basis of the severity of symptoms. There were 247 limbs that were the source of claudication complaints and 58 limbs with severe disease evidenced by rest pain, ischemic ulceration, or gangrene (RP/U/G).

**Normal subjects.** Data were collected from 25 normal control subjects (50 limbs), and TE and PORH were administered to each subject. The order of the tests was randomly assigned, and a 30-minute rest period was interposed between the tests. In an effort to assure disease-free vasculature, all control subjects were younger than 30 years and without history of smoking, diabetes, or lower limb trauma. Directional Doppler velocity waveforms were triphasic (normal) in all controls.

**Reproducibility.** To determine the reproducibility of the tests, 20 limbs of 10 stable claudication patients were studied prospectively. Each person was evaluated with RAI, TE, and PORH on five consecutive testing days spaced 30 days apart. There was a 30-minute rest period between the administration of TE and PORH. RAI, treadmill walking distance, and minimum poststress index measurements were made by the same observer, with the same equipment, and at the same time of day.

**Technique.** All patients underwent baseline arm and ankle systolic pressure measurements utilizing a unidirectional Doppler ultrasound velocity detector. Brachial pressures were determined for each arm, and the higher of the two was used. A 12 cm pressure cuff was placed just proximal to the malleoli, and measurements were made for the dorsalis pedis and posterior tibial vessels. When the values of the two vessels differed, the higher of the two was used. All ankle indices were calculated as ankle systolic pressure divided by brachial systolic pressure.

TE was performed at 1.5 mph with a 7% grade. Patients were instructed to walk as long as possible, to a maximum of 10 minutes. At the termination of exercise, ankle and brachial pressure measurements were taken immediately and at 2-minute intervals for a total of 15 minutes. The minimum postexercise ankle index was tabulated, and the amount of time taken to recover to 50%, 70%, and 90% of the baseline ankle index was calculated. The 90% value was chosen rather than the 100% value after observing many limbs, including some normal controls, that did not recover to the 100% mark during the 15-minute period, despite having reached the 90% level earlier.

PORH testing has been used by previous investigators to measure the postocclusion arterial velocity, distal pulse reappearance time, or postocclusion distal arterial pressure. In the present study we chose to evaluate the latter. PORH testing was accomplished with wide pneumatic cuff placement at the upper calf and inflation for 4 minutes at 100 mm Hg higher than resting ankle systolic pressure, to a maximum of 300 mm Hg. Ankle and brachial pressure measurements were taken immediately after occlusion and then at 20-second intervals for a total of 3 minutes. The minimum postocclusion ankle index was tabulated, and the 50%, 70%, and 90% recovery times were calculated as previously described.

**Statistical analysis.** All bivariate data were analyzed with the Student's t test. Linear regression was determined utilizing the method of least squares, and correlations were accomplished through calculation of Pearson's coefficient.

Normal ranges for RAI, TE, and PORH were determined by using data from the normal subjects and
then calculating the 95% one-tailed confidence interval. Sensitivities were determined by calculating the percentage of limbs with arteriographic evidence of disease that fell outside the normal range. Specificities were defined through calculation of the percentage of control limbs that fell within the normal range.

The reproducibility of the tests was determined by analyzing the intrasubject variance of each test over successive testing days. This variance was expressed as the relative standard deviation of the test. The relative standard deviation is an index whereby if a test is repeated for a given individual many times, 95% of the outcomes may be expected to fall within that percentage of the mean outcome equal to twice the relative standard deviation.

The comparative diagnostic accuracy of RAI, TE, and PORH was assessed with receiver-operating characteristic (ROC) curve analysis. This statistical test evolved as dissatisfaction arose with the standard tests of diagnostic accuracy—sensitivity and specificity. These two parameters describe the accuracy of a test at only one specific cutoff point between what is to be declared "normal" and what is to be "abnormal." The ROC curve is generated by plotting the percentage of false positive results (abscissa) versus the percentage of true positive results (ordinate) and describes a test over the spectrum of individual cutoff values. The area underneath the curve is large whenever the percentage of true positive results is large for the corresponding percentage of false positive results. Thus, the area beneath the ROC curve is an index of the diagnostic decision quality of the test. Although it is difficult to define statistical significance between two sensitivities or specificities, a test has been developed for the significance of differences between ROC curves. Two ROC curves may be compared and a probability value generated for rejecting the null hypothesis that the tests are of the same diagnostic value.

RESULTS

Group comparison. No significant differences were detected with respect to sex and RAI between the TE and PORH groups. However, the PORH-tested patients were slightly older than the TE-tested patients (66 versus 62 years, \(P < 0.05\)). This is probably a result of the method of selection, since the older patients were less fit for TE.

Resting ankle index. The RAI (Fig. 2) was lowest...
in the RP/U/G group (0.46 ± 0.02, mean ± SEM), significantly higher in the claudication group (0.62 ± 0.01), and highest in the normal group (1.10 ± 0.01, \( P < 0.001 \)).

Walking distance. All control subjects completed the full 10-minute exercise period, walking 1320 feet. The claudication patients walked 650 ± 40 feet, and the RP/U/G patients walked 660 ± 220 feet, both significantly less than control (\( P < 0.001 \)), but neither differing from each other (\( P > 0.40 \)). A weak but significant correlation existed between walking distance and RAI (\( r = 0.23, P < 0.05 \)).

Treadmill exercise. Minimum postexercise ankle indices followed the same pattern as the RAI (Fig. 2). The mean of the RP/U/G limbs (0.22 ± 0.04) was significantly lower than that of the claudication limbs (0.37 ± 0.02), and the mean of the control limbs (1.01 ± 0.01) was higher than that of either the claudication or the RP/U/G group (\( P < 0.001 \)). There was a good correlation between minimum postexercise index and RAI (\( r = 0.78 \)), which was significant at the 0.001 level. Recovery to the baseline RAI following exercise was rapid in the control limbs, and although the recovery times for the claudication group and the RP/U/G group were longer than that of the control

<table>
<thead>
<tr>
<th>Test</th>
<th>Normal range</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI</td>
<td>&gt; 0.97</td>
<td>97</td>
<td>100</td>
</tr>
<tr>
<td>TE</td>
<td>&gt; 0.85</td>
<td>97</td>
<td>96</td>
</tr>
<tr>
<td>PORH</td>
<td>&gt; 0.63</td>
<td>89</td>
<td>96</td>
</tr>
</tbody>
</table>

group (\( P < 0.001 \)), they did not differ significantly from each other (Fig. 3).

Postocclusive reactive hyperemia. Minimum postocclusion ankle indices were lowest in the RP/U/G group (0.20 ± 0.02), significantly higher in the claudication group (0.38 ± 0.02), and highest in the control group (0.82 ± 0.02, \( P < 0.001 \)) (Fig. 2). The correlation coefficient of 0.81 between PORH minimum index and RAI was significant at the 0.001 level. The recovery times for both the claudication and RP/U/G groups were significantly longer than that of the normal group (Fig. 3), but in contrast to the recovery after exercise, the postocclusion recovery times were longer in the RP/U/G group than in the claudication group (\( P < 0.001 \)). The failure to attain statistical significance in the treadmill-tested limbs was
probably a result of the small population size of the TE-tested RP/U/G group.

Reproducibility. The reproducibility of RAI, TE, and PORH was determined through calculation of the relative standard deviation. RAI was the most reproducible test \( (P < 0.001) \), with a relative standard deviation of 9.5%. Walking distance was the least reproducible test \( (P < 0.001) \), with a relative standard deviation of 43.9%. TE and PORH minimum indices were equally reproducible measures \( (P > 0.50) \), with

Accuracy of the tests. The 95% confidence interval for RAI defined the normal range as higher than 0.97 (Table I). This was associated with a sensitivity of 97% and a specificity of 100%. TE demonstrated a sensitivity of 97% and a specificity of 96%, while PORH was associated with a sensitivity of 89% and a specificity of 96%. The normal ranges, sensitivities, and specificities for the poststress recovery times are shown in Table II, illustrating that recovery time determinations after TE or PORH are specific but insensitive measures.

Several investigators have expressed poststress data as the percentage of decrease in ankle index, and in this way the patient has also served as the control.9,24 Despite the theoretical advantages, when our data were expressed in this fashion (Fig. 4), sensitivities decreased to 72% for TE and 64% for PORH, while specificities remained high at 94% for both tests. This loss of sensitivity reflects the observation that many individuals with arterial occlusive disease and abnormal RAI have small decreases in ankle index after exercise or arterial occlusion.

The ROC curves for RAI, TE, and PORH are plotted in Figs. 5 and 6. When the tests were used to distinguish arteriographically diseased limbs from normal limbs, RAI was as diagnostically useful as TE \( (P > 0.60) \), and both were significantly more valuable than PORH \( (P < 0.02) \) (Fig. 5). When the tests were used to define severity of disease (i.e., predict whether a
Table II. Normal range, sensitivity, and specificity for the 50%, 70%, and 90% recovery to baseline ankle index after TE and PORH

<table>
<thead>
<tr>
<th>% Recovery</th>
<th>Normal range (mm)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Normal range (sec)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>&lt; 1.0</td>
<td>43</td>
<td>100</td>
<td>&lt; 10</td>
<td>44</td>
<td>100</td>
</tr>
<tr>
<td>70</td>
<td>&lt; 1.0</td>
<td>63</td>
<td>100</td>
<td>&lt; 17</td>
<td>87</td>
<td>84</td>
</tr>
<tr>
<td>90</td>
<td>&lt; 3.3</td>
<td>74</td>
<td>94</td>
<td>&lt; 57</td>
<td>67</td>
<td>92</td>
</tr>
</tbody>
</table>

symptomatic limb was the source of claudication complaints only or whether RP/U/G was present), no test was significantly more valuable than another (P > 0.40) (Fig. 6).

Data on two groups of limbs with early disease were closely examined to ascertain if the addition of stress testing to RAI increased the sensitivity of RAI determination alone. The first group consisted of 67 asymptomatic limbs with contralateral symptomatic disease. An abnormal resting index was detected in 53 (79%), and the addition of stress testing did not increase the sensitivity. Of 46 limbs undergoing TE, an abnormal postexercise index was noted in 33 (72%), and 13 (62%) of the remaining 21 limbs tested with PORH were detected as abnormal.

The second group comprised nine limbs with claudication symptoms and a normal RAI. Two (22%) of these limbs were defined as diseased through use of stress testing, one with TE and one with PORH. In the 372 total limbs studied, the addition of TE or PORH to RAI testing accounted for an increase in diagnostic yield of only 6 of 372, or 1.6%.

Cost of the tests. At the time of this study, RAI cost the patient $25 for both limbs, TE cost an additional $45, and PORH cost $25. Overall, in the 218 patients studied, the addition of stress testing to RAI alone cost $6620. Since six of the limbs may have been incorrectly diagnosed without the information obtained by stress testing, routine use of stress testing cost $1100 per limb correctly diagnosed.

DISCUSSION

It has been shown that physiologic measurements can be important aids in the evaluation of arterial disease. Historically, resting ankle pressure was the first parameter examined, and it was found to be a good indicator of the severity of disease.21 The concepts of exercise and ischemia-induced hemodynamic changes are old ones. Andre-Thomas, Levy-Valensi, and others have documented a diminution of the peripheral pulse of claudication patients after exercise.9,14 Ejrup6 and Winsor22 were able to quantitate pressure changes distal to an arterial stenosis after exercise, and Sumner and Strandness15 described the use of TE testing as a means of increasing the sensitivity of the noninvasive arterial examination.

Reactive hyperemia has long been recognized, with early descriptions of this phenomenon appearing in the writings of Lister.11 Several investigators have implemented PORH to show differences in blood flow, velocity, and pulse pressure between patients with intermittent claudication and normal subjects.8,10,17 Baker1 and Hummel et al.9 compared TE with PORH and found good correlation between the pressure decreases induced by the two tests. However, to date there have been few data published regarding the value of these tests in the definition of the presence and extent of disease in a large number of patients.

In the present study we set out to determine the usefulness of stress testing over resting measurements in the diagnosis of lower limb arterial occlusive disease. We evaluated two essential characteristics of the tests—reliability and discrimination. Reliability refers to how consistent the results of a test are or how replicable the test is.30 The prospective segment of this study illustrated that when a limb was tested with RAI on successive occasions, almost identical results were obtained. Therefore, RAI is reliable on the basis of its repeatability. TE and PORH were significantly less reliable in this regard, whereas treadmill walking distance was an imprecise, unreliable measure.

Discrimination refers to how accurate a test is in distinguishing disease states.30 In order to evaluate the discrimination of resting versus poststress testing, we compared data on limbs with arteriographic evidence of disease with data of the normal controls; we found resting determinations to be highly sensitive and specific for the presence of occlusive disease. However, the subset of arteriogramed patients generally represents a select group with advanced disease, a population distinct from the one for which stress testing was conceived. Therefore, we chose to evaluate a second group
of patients with early disease, comprising asymptomatic limbs with contralateral symptomatic disease. Since atherosclerosis is a bilateral disorder, the vast majority of these limbs had at least mild disease. Stress testing added little to resting index determinations in this subpopulation as well, accounting for only a 6% increase in diagnostic yield. This suggests that RAI determination alone is an adequate screening procedure for the detection of mild arterial stenosis.

Several studies have focused on the value of measuring the time it takes for an ankle index to return to resting levels after TE or PORH. All have shown that recovery time is prolonged in occlusive disease. The present study illustrated that TE and PORH recovery time measurements were poor indicators of disease. Although an abnormal recovery time was generally associated with an abnormal extremity, a normal recovery time did not exclude disease.

The distance a patient is able to walk on the treadmill has previously been considered a useful indicator of severity and progression of disease. Although the walking distances of both the claudication and RP/U/G groups were shorter than that of the control group, neither differed from the other. Walking distance was not a reproducible measure; in addition, there was a weak correlation between the walking distance and the severity of disease as assessed by RAI. This suggests that walking distance is not a useful indicator of the severity of disease, and more accurate information is obtained with less subjective parameters.

Resting index determination is a fast, simple, and painless procedure. In contrast, TE and PORH have many technical disadvantages. TE requires a motorized treadmill, and many patients require simultaneous cardiac monitoring. Thus, it is difficult to perform at the bedside. Certain patients, for example individuals with severe pulmonary or cardiac disease, cannot safely be tested in this manner. The amount of exercise is limited by the leg with more severe disease, and it is not possible to obtain an accurate assessment of the less involved leg in patients with bilateral disease. Breathlessness and other factors unrelated to vascular disease determine the stopping point and may not allow the patient to continue walking long enough to obtain representative pressure decreases. PORH is an uncomfortable test for the patient to undergo, and some patients do not tolerate the occlusive period. Neither TE nor PORH is ideal for the postoperative patient. Incisional pain interferes with treadmill walking, and occlusion of a distal graft during reactive hyperemic testing may cause graft injury and thrombosis.

In summary, our data show that the population that may benefit from stress testing is very small. Alone, resting determinations are sensitive and specific in defining the presence of disease. Overall, only six additional limbs were defined as diseased with the addition of stress testing, accounting for a 1.6% increase in diagnostic yield. We conclude that RAI is a simple, accurate, and reproducible test. Routine stress testing is not cost effective, adding little diagnostic information to resting determinations. Stress testing in the form of TE may be worthwhile in the small subset of symptomatic patients with a normal resting index, but it should not be used routinely for all patients. PORH is not as accurate as exercise testing, but it is safer and may be desirable for the patient for whom TE is risky.

REFERENCES
11. Lewis T, Grant R. Observations upon reactive hyperemia in man. Heart 12:73-120, 1926


