Reliability of Single-Lead and Multiple-Lead Electrocardiography during and after Exercise*

Michael H. Sketch, M.D.; Chandra K. Nair, M.D.; Dennis J. Esterbrooks, M.D.; and Syed M. Mohiuddin, M.D.

To define the number, type, and criteria for interpretation of electrocardiographic leads yielding the most reliable results during and after exercise, 203 men were studied. Twelve-lead exercise electrocardiograms (including lead CM5) and coronary arteriograms were correlated. Depreciation of the S-T segment that was present only during exercise was of doubtful significance. Analysis of ECGs after exercise revealed that only leads V5, V6, and CM5 had acceptable sensitivity (61, 53, and 60 percent, respectively). Analysis with multiple leads increased the sensitivity of exercise tests to 65 percent. Lead V5 correctly identified 39 percent (17) of the 44 patients with single-vessel disease and 74 percent (49/66) of those with multiple-vessel disease. The predictive values of interpretations based on lead V5 and multiple leads were not significantly different (84 percent [67/80] and 82 percent [71/88], respectively). The highest predictive values were in lead 2 (88 percent; 37/42) and lead V4 (88 percent; 45/51). Sensitivity was inversely related to the degree of ST-segment depression required for abnormality. Specificity was directly related. Predictive value was directly related in all except leads V4 and CM5. A predictive value of 100 percent was manifested by the various leads with the following amounts of ST-segment depression: lead V3, ≥2 mm; leads V4, V5, and aVF, ≥3 mm; and lead V6, ≥4 mm. Even with ST-segment depression of 4 mm or more, lead CM5 had only 83 percent predictive value. Using analysis with multiple leads, all patients with abnormal findings on maximal exercise tests of less than six minutes in duration had significant disease.

McHenry1 has observed that “the number, the torso placements and the type of ECG leads which will yield the most reliable results in exercise electrocardiography remain to be defined.” For the most part, the selection of leads for exercise electrocardiograms is based on the classic work of Blackburn et al2 published in 1967. That report ranked 15 different systems of leads based on their sensitivity in recognizing exercise-induced ischemic depression of the S-T segment. Correlation of lead-specific ST-segment depression and the results of coronary arteriographic studies was not evaluated. Hence, in order to better define the reliability of different systems of electrocardiographic leads and their criteria for interpretation, the 12-lead ECG (including lead CM5) was correlated with angiographic results in 203 men.

MATERIALS AND METHODS

Patients

Two hundred and three men who were referred for evaluation of pain in the chest were studied. The mean age

of these subjects was 50 ± 9 years (± SD). Patients suffering from cardiac valvular disease, myocarditis, pericarditis, cardiomyopathy, anemia, thyroid disease, and electrolyte imbalance were excluded. No patient received digitalis during the two weeks prior to exercise testing. One hundred and ninety (94 percent) of the 203 men had normal resting ECGs. Thirteen men (6 percent) had nonspecific changes in the T wave on their resting ECGs.

Exercise Testing

All patients performed continuous multistage exercise tests according to the protocol of Bruce. Forty-five patients (22 percent) discontinued exercise when they attained a heart rate between 85 and 95 percent of the maximum predicted for their age3 but were not limited by symptoms or the ECG. These men were considered to have performed nearly maximal stress tests. The remaining 158 patients (78 percent), whose exercise was discontinued because of symptoms or electrocardiographic changes or because they had attained their maximal predicted heart rate, were considered to have performed maximal exercise tests. The mean duration of exercise was 7.5 ± 2.8 minutes. The mean maximal heart rate achieved was 145 ± 22 beats per minute.

Electrocardiographic tracings of high quality were obtained using a dental burr4 to prepare the skin for placement of electrodes and by using an elastic rib belt to prevent excessive motion of the wires of the electrocardiographic leads. Placement of electrodes was according to the recommendation of Mason and associates.5 Electrocardiographic tracings were recorded using a Viagraph system. Two-tenths of one minute was required to record each 12-lead ECG. The recording sys-

*From the Department of Cardiology, Creighton University School of Medicine, Omaha.

- From the Department of Cardiology, Creighton University School of Medicine, Omaha.
- From the Department of Cardiology, Creighton University School of Medicine, Omaha.
- From the Department of Cardiology, Creighton University School of Medicine, Omaha.
- From the Department of Cardiology, Creighton University School of Medicine, Omaha.
- From the Department of Cardiology, Creighton University School of Medicine, Omaha.

394 Sketch ET AL

CHEST, 74: 4, OCTOBER, 1978
iment was modified to delete lead aVR. It was replaced by lead CM$_5$ in 110 patients. Twelve-lead ECGs were recorded in both the supine and standing positions prior to exercise. During exercise, ECGs were recorded at the end of each three-minute state. Simultaneously with stopping the treadmill, ECGs were recorded in the supine position. Subsequently, ECGs were recorded in the supine position at one-half, one, two, three, four, five, and six minutes after exercise.

Electrocardiographic Analysis

All electrocardiographic tracings were interpreted by three of us independently. If independent interpretations did not agree, the recordings were reviewed by us in conference. If agreement could not be reached, the case was discarded. It was necessary to discard seven cases. To be classified as abnormal, exercise-induced ST-segment depression had to be at least 0.05 second in duration and flat or downward sloping. If the resting ECG manifested non-specific changes in the T wave at rest, exercise-induced flat or downward ST-segment depression of 1 mm or more was also considered to be abnormal. Each electrocardiographic lead was evaluated independently and was graded as manifesting (1) no ST-segment depression or less than 1 mm of depression, (2) 1 to 1.9 mm of ST-segment depression, (3) 2 to 2.9 mm of ST-segment depression, (4) 3 to 3.9 mm of ST-segment depression, or (5) 4 mm or more of ST-segment depression. These grades were applied independently to the following subgroups: (1) ECGs during exercise, (2) ECGs immediately after exercise (standing, simultaneously with stopping the treadmill), and (3) all those recordings during the subsequent period of recovery. Each subgroup was graded according to the ECG which manifested the greatest ST-segment depression.

Coronary Arteriograms

Selective coronary arteriographic studies were performed on all patients according to the method previously described. Each coronary arteriogram was reviewed by a minimum of three of us, and the agreed consensus was accepted at the final interpretation. A coronary artery was considered to demonstrate significant disease if it manifested a stenotic lesion equal to or greater than 75 percent of the diameter of the artery in any projection. The severity of coronary arterial disease was determined by calculating the "severity index," as previously described. All lesions were classified as single or multiple. Each stenosis was classified according to the percentage by which it narrowed the diameter of the vessel's lumen (less than 50 percent, 50 to 74 percent, 75 to 94 percent, or 95 percent or greater [including total occlusion]).

Statistical Analysis

All data were sorted and analyzed by standard statistical techniques using a computer (IBM System 370). Sensitivity was defined as the percentage of patients with disease who had an abnormal test, i.e., true positives/(true positives + false negatives) × 100. Specificity was defined as the percentage of patients without disease who had an abnormal test, i.e., true negatives/(true negatives + false positives) × 100. Predictive value was defined as the percentage of patients with an abnormal test who had disease, i.e., true positives/(true positives + false positives) × 100.

RESULTS

Coronary Arteriograms

Eighty-three men (41 percent) had normal coronary arteriograms or failed to reveal stenosis equal to or greater than 50 percent of the diameter of the artery involved. Ten men (5 percent) had stenosis

---

Figure 1. Resultant sensitivities, specificities, and predictive values using only ECGs recorded one-half to six minutes after exercise (A), using ECGs immediately after exercise and all subsequent ECGs (B), and using all ECGs during and after exercise (C). Probabilities are based on exercise-induced ST-segment depression of 1 mm or more and coronary arterial stenosis of 75 percent or more being abnormal.
Leads with ST-Segment Depression

All abnormal exercise tests were correctly identified by the presence of exercise-induced ST-segment depression in one or more of leads 2, aVF, and V₃ to V₆.

Accuracy Related to Time

Figure 1 reveals that the highest specificities and predictive values were obtained when exercise tests were judged to be abnormal or normal on the basis of those electrocardiographic tracings recorded one-half to six minutes after exercise (A). In general, when electrocardiographic tracings recorded immediately after exercise were included for interpretation (B), the specificities and predictive values fell.

When interpretation of the exercise tests was based on all electrocardiographic tracings recorded during and after exercise (C), the specificities and predictive values fell further.

Table 1 reveals that although the incidence of exercise tests that were abnormal only during and immediately after exercise was small, the percentages of these tests that were true positives (ie, predictive values) were low. In contradistinction to specificity and predictive value, the sensitivity of the exercise tests increased when the electrocardiographic tracings immediately after exercise and during exercise were included with the ECGs obtained one-half to six minutes after exercise for interpretation.

Comparative Accuracy of Single and Multiple Leads

From Figure 2 and Table 2, it is apparent that when exercise tests were judged abnormal or normal by analysis of a single lead, only leads V₅, V₆, and CM₅ manifested acceptable sensitivities and that when exercise tests were judged abnormal or normal from analysis of electrocardiographic leads in combination, sensitivity was highest (65 percent). Hence, the more leads reviewed for analysis, the greater the sensitivity. In contradistinction, analyses of single leads resulted in the highest specificities and predictive values (Fig 2). Of the individual leads that manifested the highest sensitivity (leads V₅, V₆, and CM₅), lead V₅ had the highest specificity (86 percent; 80/93) and highest predictive value (84 percent; 67/80), although leads V₆ and CM₅ had acceptable values. The highest specificities were in lead V₅ (94 percent; 87/93), lead V₆ (94 per-
cent; 87/93), lead 2 (95 percent; 88/93), and lead aVF (94 percent; 87/93). The highest predictive values were in lead 2 (88 percent; 37/42) and lead V4 (88 percent; 45/51).

Sensitivity and Single vs Multiple Stenoses

Table 3 illustrates that all leads manifest a statistically significant difference in their ability to identify coronary arterial disease of single vs multiple vessels (P < 0.05).

ST-Segment Depression and Predictive Probabilities

Figure 3 reveals the sensitivities, specificities, and predictive values that resulted from using increasing degrees of ST-segment depression as the cutoff for normality (i.e., 1, 2, 3, and 4 mm). It is readily apparent that in all leads, sensitivity rapidly falls to unacceptable levels as the degree of ST-segment depression required for abnormality is increased. In all leads, specificity progressively increased as the degree of ST-segment depression required for abnormality was increased. When any electrocardiographic lead, except lead CM5, manifested 4 mm or more of exercise-induced ST-segment depression, the resultant predictive value was 100 percent, but even at this degree of ST-segment depression, lead CM5 manifested a predictive value of only 83 percent. In those patients who had 3 mm or more of exercise-induced ST-segment depression in leads V5, V4, V6, 2, and aVF, the resultant predictive values were 100 percent. Lead V2 was the only electrocardiographic lead that manifested 100 percent predictive value at 2 mm of ST-segment depression. It is important to note that by increasing the cutoff for normality from 1 mm to 2 mm of exercise-induced ST-segment depression, there was a fall in the resultant predictive values in leads V4 and CM5. In fact, lead CM5 manifested a progressive fall in predictive value as the degree of required ST-segment depression was increased to 3 mm.

ST-Segment Depression and Severity of Disease

Figure 4 reveals that in general, there was a direct relationship between the severity of disease and the

<table>
<thead>
<tr>
<th>Lead</th>
<th>Single-Vessel</th>
<th>Multiple-Vessel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead CM1</td>
<td>44 (11/25)</td>
<td>69 (29/42)</td>
</tr>
<tr>
<td>Lead V1</td>
<td>14 (6/44)</td>
<td>21 (14/66)</td>
</tr>
<tr>
<td>Lead V2</td>
<td>20 (9/44)</td>
<td>55 (36/66)</td>
</tr>
<tr>
<td>Lead V3</td>
<td>39 (17/44)</td>
<td>74 (49/66)</td>
</tr>
<tr>
<td>Lead V4</td>
<td>25 (11/44)</td>
<td>71 (47/66)</td>
</tr>
<tr>
<td>Lead 2</td>
<td>16 (7/44)</td>
<td>45 (30/66)</td>
</tr>
<tr>
<td>Lead aVF</td>
<td>14 (6/44)</td>
<td>39 (26/66)</td>
</tr>
</tbody>
</table>

*Abnormal exercise test had 1 mm or more of ST-segment depression.
**No. of patients with 75 percent or more stenosis over those with abnormal ECGs.
degree of ST-segment depression in all electrocardiographic leads, even though the standard deviations were large. The differences between the severity indices in those patients with 1 to 1.9 mm of ST-segment depression and those with 3 mm or more of ST-segment depression were statistically significant in leads V₅, V₆, V₇, 2, and aVF (P < 0.05). The same relationship existed between ST-segment depression and the mean number of vessels with stenosis of 75 percent or more. Although between some leads, there appear to be differences in the mean severity indices and the mean number of vessels with stenosis that were associated with a given degree of ST-segment depression (especially in those patients with 3 to 3.9 mm of ST-segment depression), the differences were not statistically significant (P < 0.05). Hence, a given degree of ST-segment depression appears to have the same clinical significance (as to severity of disease) in no matter which electrocardiographic lead it is present.

**Duration of Exercise**

Table 4 illustrates the results of exercise testing and coronary arteriograms in the 158 patients who underwent maximal exercise testing. There is an indirect relationship between the duration of exercise and both the severity and the incidence of disease. Sensitivity, specificity, and predictive value were also indirectly related to the duration of exercise. The sensitivity of exercise tests was very high in those patients who exercised less than six minutes and was very low in those who exercised for nine minutes or longer. All patients who exercised for less than six minutes and had an abnormal exercise test.
and reported that virtually all ischemic responses to exercise could be identified in leads 2, aVF, and V_4 to V_6. Of these, lead V_4 was shown to be the best single lead. It identified 89 percent of the ischemic results.2

In order to further improve the sensitivity of exercise testing, the use of other leads has been proposed. Probably the most popular single lead has been lead CM_5, as proposed by Ellestad11 and others. Most recommendations for the use of different systems of single and multiple leads in exercise ECGs are based on the ability of a lead or leads to identify an ischemic response and not on the ability to identify significant coronary arterial disease. This is probably why McHenry1 recently pointed out that “the number, the torso placements and the type of ECG leads which will yield the most reliable results in exercise electrocardiography remain to be defined.” Our retrospective study correlated the response of the S-T segment in various systems using single and multiple leads (both during and after exercise) with coronary arteriographic results.

Of the predictive probabilities evaluated, the sensitivity and the predictive value are most clinically significant. Sensitivity permits an estimation of the percentage of patients with diseases in a similar population who would be identified by the test. Predictive value permits an estimation of the likelihood that a given patient with an abnormal test has disease; however, it must be recognized that predictive value is dependent upon the incidence of disease in the population.

From Figure 1, it is evident that the sensitivity of the exercise tests was highest when their interpretation was based on all electrocardiographic recordings during and after exercise. Although the incidence of exercise tests that are abnormal only during exercise is small, the incidence of false-positive results in this group is high (Table 1). All electrocardiographic leads manifest highest specificity in the recordings after exercise (Fig 1). This suggests that in clinical practice, when the purpose of exercise testing is screening, all ECGs during and after exercise should be evaluated; however, when the purpose of exercise testing is to aid in the diagnosis of significant coronary arterial disease, the interpretation of exercise ECGs as performed in our study should be based on recordings after exercise only. The prevalence of false-positive results was high in patients whose ECGs were abnormal during exercise only; however, the number of patients manifesting this phenomenon was small, precluding definite conclusions, but was sufficient to suggest the need for further study of this subgroup.

Virtually all exercise-induced ischemic responses could be identified in leads 2, aVF, and V_4 to V_6.

### Table 4—Duration of Exercise Related to Severity of Coronary Arterial Disease and Predictive Probabilities

<table>
<thead>
<tr>
<th>Data</th>
<th>Duration of Exercise, min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-5.9</td>
</tr>
<tr>
<td>No. of patients</td>
<td>49</td>
</tr>
<tr>
<td>Patients with more than 75 percent stenosis (percent)</td>
<td>43 (88)</td>
</tr>
<tr>
<td>Mean severity index (± SD)</td>
<td>6.4 ± 2.6</td>
</tr>
<tr>
<td>Mean No. of vessels stenosed (± SD)</td>
<td>2.1 ± 0.7</td>
</tr>
</tbody>
</table>

*Leads 2, aVF, and V_4 to V_6.*

(i.e., 1 mm or more of exercise-induced ST-segment depression) had at least one major coronary artery with stenosis of 75 percent or more. Although the predictive value of abnormal tests in patients who exercised 6 to 8.9 minutes is acceptable (eg, in lead V_4, predictive value of 87 percent), this value was very low in patients who exercised for nine minutes or more. At least partially, this is due to the low incidence of coronary arterial stenosis in these patients.

### Nonspecific Changes in T Wave

Exclusion of the 13 men with nonspecific changes in the T wave on their resting ECGs from the group under study failed to significantly alter prevalences, the results of statistical analysis, or conclusions.

### Discussion

As early as 1931, Wood and Wolferth9 reported that the precordial leads manifested greater sensitivity in detecting ST-segment depression than did the standard leads of Einthoven. Master and Rosenfeld8 proposed that electrocardiographic leads V_4 to V_6 and 2 be used to evaluate the “ischemic” response of the S-T segment to exercise. In 1964, Blackburn and associates10 evaluated 100 abnormal exercise tests
confirming the work of Blackburn et al.\textsuperscript{1,10} Lead V\textsubscript{6} was found to be best among unipolar or bipolar leads. This is in agreement with the observation of Froelicher et al.\textsuperscript{12} Using multiple leads for interpretation increased the sensitivity of the exercise tests. This improvement was only slightly nullified by a minimal decrease in predictive value, and this could be improved upon by observing which leads manifested significant ST-segment depression; for instance, when the tracing from lead V\textsubscript{6} was abnormal, the predictive value was 90 percent. Hence, it appears that there is a definite advantage in using ECGs from multiple leads.

It has been suggested that criteria for significant ST-segment depression in exercise testing are not directly interchangeable between different electrocardiographic leads.\textsuperscript{12} Experience has shown that by visual analysis, it is difficult to judge what is significant when the degree of exercise-induced ST-segment depression is less than 1 mm, especially in leads with low amplitude of the QRS complex. It would appear, therefore, that any modification of presently accepted criteria would have to consist of increasing the required degree of ST-segment depression, rather than decreasing it. In all leads studied, increasing the degree of ST-segment depression required for abnormality resulted in an unacceptable decrease in sensitivity in all leads (Fig 3). Because of this, the presently accepted minimal criteria of 1 mm of ST-segment depression appears best.

When interpretation of exercise tests was based on ST-segment depression in leads V\textsubscript{3}, V\textsubscript{5}, V\textsubscript{6}, and aVF, there was a direct relationship between the degree of ST-segment depression and the probability that an abnormal test was associated with significant coronary arterial stenosis. This direct relationship was poor in leads V\textsubscript{5} and CM\textsubscript{5} (Fig 3). Hence, although 1 mm of ST-segment depression appears to be the best minimal criterion for abnormality in all leads, by noting the degree of ST-segment depression in multiple leads, the clinician can improve the predictive accuracy of exercise testing.

Our data confirm the observations of Dietry,\textsuperscript{13} Bruce,\textsuperscript{14} Bartel et al.\textsuperscript{15} and others that marked ST-segment depression is, in general, associated with severe coronary arterial disease (Fig 4), although lesser degrees of ST-segment depression are not necessarily predictive of the severity of disease. Although between some leads, there appeared to be differences in the severity of coronary arterial disease associated with a given degree of ST-segment depression, these differences were not statistically significant (Table 4). Hence, it appears that the clinician cannot infer that a patient who has a given degree of ST-segment depression in one lead has either more or less severe disease than another patient who has the same degree of ST-segment depression in another lead.

Table 4 reveals an inverse relationship between the duration of exercise and both the incidence and the severity of coronary arterial stenosis. The accuracy of the exercise tests was also inversely related to the duration of exercise (Table 4). Hence, by observing the duration of a maximal exercise test, it appears that the clinician can more accurately predict the probability that a given test is associated with significant coronary arterial stenosis.

Chaitman et al.\textsuperscript{16} recently reported improved efficiency of treadmill exercise testing using multiple leads. Superficially, the improvement shown in that study is greater than in this report; for instance, in the study by Chaitman et al.\textsuperscript{16} analysis of 11 standard leads resulted in a sensitivity of 76\% (50/66), compared to 56\% (37/66) using lead V\textsubscript{3} and 68\% (45/53) using lead CM\textsubscript{5}. From our data, these sensitivities were 65 percent (72/110), 61 percent (67/110), and 60 percent (40/67), respectively; however, Chaitman et al.\textsuperscript{16} recognized slow upward sloping of the S-T segment as an abnormal test. This could explain their higher sensitivities, because when slow upward sloping was excluded in the study of Chaitman et al.,\textsuperscript{16} sensitivities were 67 percent using 11 standard leads for analysis, 40 percent when using lead V\textsubscript{3}, and 54 percent when using lead CM\textsubscript{5}. The large difference demonstrated between analysis with multiple leads and analysis with lead V\textsubscript{3} or lead CM\textsubscript{5} was due to the low sensitivity obtained from leads V\textsubscript{3} and CM\textsubscript{5} (40 percent and 54 percent, respectively). The predictive values reported in the study by Chaitman et al.\textsuperscript{16} were higher than ours; however, the incidence of disease was greater in that population than in our population (66 percent vs. 54 percent), as was the prevalence of multivessel disease (45 percent vs 33 percent), and this would increase the predictive values of Chaitman et al.\textsuperscript{16} per se. Despite these differences, the results of our study confirm the conclusions of Chaitman et al.\textsuperscript{16}

Finally, it should be recognized that many of the conclusions drawn from this study apply only to visual analysis. They cannot be inferred to apply to computerized analysis, which may result in better correlations when using leads other than those evaluated in this study.\textsuperscript{17} Furthermore, some of the conclusions reflect the incidences of disease in our population and are not necessarily applicable to other subsets, such as women, men with atypical angina, and asymptomatic patients.
REFERENCES


4 Sketch MH: Exercise tests: What to look for. Primary Cardiol, Feb 1976, p 16-19


8 Wood FC, Wolferth CC: Angina pectoris: The clinical and electrocardiographic phenomena of the attack and their comparison with the effects of experimental temporary coronary occlusion. Arch Intern Med 47:339, 1931


