Exercise-Induced ST Depression in the Diagnosis of Coronary Artery Disease

A Meta-Analysis

Renato Gianrossi, MD, Robert Detrano, MD, PhD, Daniel Mulvihill, MD,
Kenneth Lehmann, MD, Paul Dubach, MD, Antonio Colombo, MD,
David McArthur, PhD, and Victor Froelicher, MD

To evaluate the variability in the reported diagnostic accuracy of the exercise electrocardiogram, we applied meta-analysis to 147 consecutively published reports comparing exercise-induced ST depression with coronary angiography. These reports involved 24,074 patients who underwent both tests. Population characteristics and technical and methodologic factors, including publication year, number of electrocardiographic leads, exercise protocol, use of hyperventilation, definition of an abnormal ST response, exclusion of certain subgroups, and blinding of test interpretation were analyzed. Wide variability in sensitivity and specificity was found (mean sensitivity, 68%; range, 23–100%; SD, 16%; and mean specificity, 77%; range, 17–100%; SD, 17%). The four study characteristics found to be significantly and independently related to sensitivity were the treatment of equivocal test results, comparison with a “better” test such as thallium scintigraphy, exclusion of patients on digitalis, and publication year. The four variables found to be significantly and independently related to specificity were the treatment of upsloping ST depressions, the exclusion of subjects with prior infarction or left bundle branch block, and the use of preexercise hyperventilation. Stepwise linear regression explained less than 35% of the variance in sensitivities and specificities reported in the 147 publications. There is wide variability in the reported accuracy of the exercise electrocardiogram. This variability is not explained by information reported in the medical literature. (Circulation 1989;80:87–98)

The diagnosis of coronary artery disease in patients with chest pain syndromes is a major application of the exercise electrocardiogram (ECG). Although several variables have been reported as useful in the analysis of electrocardiographic results, depression of the ST segment below baseline has received the most attention. Experts agree concerning the importance of ST depression in the diagnosis of coronary disease. However, disagreements occur concerning the quantitation of ST segment depression measurements.1–3 Furthermore, exercise ECG protocols differ markedly as to number of leads used, type of exercise (e.g., bicycle, treadmill), computerization of results, and treatment of equivocal or nondiagnostic tests.

To determine the accuracy of the exercise ECG for the prediction of coronary disease, clinical investigators have performed numerous studies comparing exercise-induced ST depression with results of coronary angiography. Though excellent research has been accomplished, little agreement has been reached concerning the optimum protocol and method of interpretation or even the accuracy of this test.

Philbrick and colleagues4 reviewed this subject in 1980 and found wide variability in reported accuracies. They concluded that the failure to conform to methodologic standards of research design could explain this variability. Such methodologic deficiencies would lead to a spurious increase in reported accuracy and false expectations for the exercise ECG. Meta-analysis is a quantitative analysis of the variability of results from numerous reported studies.5–7 This technique has been applied in medicine8–10 but to our knowledge has not yet been used for evaluating diagnostic testing. We applied meta-analysis to the international literature on the diagnostic accuracy of exercise-induced ST depres-
TABLE 1. Variables Abstracted From Exercise Electrocardiogram Literature

<table>
<thead>
<tr>
<th>Not used in multivariate analysis because &gt;50% of data missing</th>
<th>Technical factors (5 variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population characteristics (7 variables)</strong></td>
<td><strong>Percent patients achieving &quot;adequate heart rate&quot;</strong></td>
</tr>
<tr>
<td>Percent of study group with:</td>
<td>Mean work load achieved</td>
</tr>
<tr>
<td>typical angina pectoris</td>
<td>Mean heart rate achieved</td>
</tr>
<tr>
<td>left ventricular hypertrophy</td>
<td>Mean double product achieved</td>
</tr>
<tr>
<td>right bundle branch block</td>
<td>Time interval between exercise test and coronary angiogram</td>
</tr>
<tr>
<td>mitral valve prolapse</td>
<td></td>
</tr>
<tr>
<td>resting repolarization abnormalities</td>
<td></td>
</tr>
<tr>
<td><strong>Percent of study group taking:</strong></td>
<td></td>
</tr>
<tr>
<td>β-receptor–blocking agents</td>
<td></td>
</tr>
<tr>
<td>long-acting nitrates</td>
<td></td>
</tr>
<tr>
<td><strong>Used as covariates in multivariate analysis</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Population characteristics (11 variables)</strong></td>
<td><strong>Technical factors (11 variables)</strong></td>
</tr>
<tr>
<td>Mean age</td>
<td>Publication year</td>
</tr>
<tr>
<td>Percent men</td>
<td>Continent of study center</td>
</tr>
<tr>
<td>Were patients with these conditions</td>
<td>Hyperventilation used before exercise (yes/no)</td>
</tr>
<tr>
<td>excluded from the study? (yes/no)</td>
<td>Exercise protocol (Bruce, Ellestad, other treadmill, upright bicycle, supine bicycle, Master step test)</td>
</tr>
<tr>
<td>Prior myocardial infarction</td>
<td>Number of leads</td>
</tr>
<tr>
<td>Left ventricular hypertrophy</td>
<td>Smallest amount of ST depression deemed abnormal (&lt;0.1, 0.1, &gt;0.1 mV)</td>
</tr>
<tr>
<td>Right bundle branch block</td>
<td>Point in time when measurement was made</td>
</tr>
<tr>
<td>Left bundle branch block</td>
<td>Upsloping ST segments considered abnormal (yes/no) (see text)</td>
</tr>
<tr>
<td>Mitral valve prolapse</td>
<td>ST depressions adjusted for heart rate (yes/no) (see text)</td>
</tr>
<tr>
<td>Resting repolarization abnormalities</td>
<td>Computer algorithm used to analyze ST segment (yes/no)</td>
</tr>
<tr>
<td>Were patients taking these medications</td>
<td>Angiographic definition of disease (50% vs. 70%)</td>
</tr>
<tr>
<td>excluded from the study? (yes/no)</td>
<td></td>
</tr>
<tr>
<td>β-Receptor–blocking agents</td>
<td></td>
</tr>
<tr>
<td>Long-acting nitrates</td>
<td></td>
</tr>
<tr>
<td>Digitalis</td>
<td></td>
</tr>
</tbody>
</table>

sion to determine the expected value and variability in this test's accuracy and determine which technical and methodologic factors independently affect the reported accuracy and its variability.

**Methods**

**Literature Review**

The Bibliography Retrieval Service and Medlars were used to search the National Library of Medicine data base for reports published after 1967 on the diagnostic accuracy of the exercise ECG when compared with coronary angiography. Though both exercise testing and angiography were used before 1967, few reports of their use in the same patients had appeared before that date. The search terms used were "exercise electrocardiogram" and "coronary artery disease." The bibliographies of three major textbooks on the subject were also searched.1-3 The bibliographies of review articles published between 1984 and 1987 and retrieved from the computer search were also scanned to complete the meta-analytic data base.

The 325 publications resulting from these searches were screened for studies involving groups of 50 patients or more undergoing exercise electrocardiography with ST segment measurement and coronary angiography; for the latest report when more than one report from the same institution and same time period (within 3 years) was found, unless it could be verified that duplication of data was not involved; and for reports that did allow the calculation of sensitivity, specificity, or both for the prediction of any significant angiographic obstruction. From the original list, 147 reports were retained for review,11-157 Three of these studies reported results of two separate study samples that had been based on differing case definitions or protocols; these study samples were treated as independent observations in the statistical calculations. There were, therefore, 150 study groups.

**Recording of Independent Variables**

Two investigators reviewed each of the 147 publications and recorded the values of the 40 independent variables listed in Table 1. The 29 reports in foreign languages were first translated into English by a physician fluent in both languages.
Variables Studied

The variables recorded for each publication included population characteristics, technical factors, and methodologic factors. Table 1 displays these variables.

Population characteristics. The first column of Table 1 lists the seven population characteristics for which the frequency of insufficient data were too high (>50%) to be considered eligible for candidacy in the multivariate analysis. The other 10 population variables with sufficient data for the multivariate analysis are listed in the third column of Table 1.

Technical factors. These factors regard only the technical performance and analysis of the exercise ECG. They are listed in the second and fourth columns of Table 1. The five technical variables in the second column were not used in the multivariate analysis because of their high frequency of missing data (>50%). The 11 variables in the fourth column were included in the multivariate regression. Three of these, concerning the point in time when the ST segment was measured, the interpretation of the upsloping ST segments, and heart rate adjustments of ST measurements, require some explanation.

Though most cardiologists measure the ST depression 80 msec after the inflection point between the QRS and ST segment (J point), some use the J point and a few use other points. When information about the point used to measure ST depression was available, it was recorded.

Controversy exists as to whether upsloping ST depressions should be considered as abnormal test results. When a report included a statement to the effect that only flat or downsloping depressions were considered abnormal, this was noted. Otherwise, the assumption was made that upsloping depressions were considered abnormal by the authors of that report.

Because the ischemic ST response is related to myocardial work and the latter is related to heart rate, heart rate adjustments of ST segment depression have been proposed to improve accuracy in the diagnosis of coronary artery disease. Because several publications reported on this topic, heart rate adjustment was included as a variable.

Methodologic factors. These factors involve the comparison of the exercise ECG with a “better” test, patient selection criteria, bias in reviewing either the exercise electrocardiogram or the angiogram, and the treatment of equivocal test results. The fifth column of Table 1 divides methodologic factors into three categories.

The first category is the comparison of the exercise-induced ST depression with another exercise test such an exercise thallium scintigraphy or exercise echocardiography, which was claimed by the authors to be of superior accuracy. Investigators assessing such “superior” exercise tests might expend less effort toward the interpretation of the exercise ECG and thereby decrease its reported accuracy.

The second category is the adherence to three methodologic standards proposed by Philbrick et al. The first of these, the avoidance of workup bias, was considered fulfilled if the study included a statement that the results of the exercise ECG did not affect the decision to perform coronary angiography. The second and third standards concerning the blind reading of the coronary angiogram and the exercise ECG were considered fulfilled if there was a clear statement that these tests were read in a blinded fashion.

And the third category is the treatment of equivocal test results. Investigators commonly report some test results as equivocal or nondiagnostic. These “equivocals” involve either results that are difficult to interpret because they are close to the cutoff of abnormality or they involve patients who had normal exercise ST segments but failed to achieve a target level of cardiac stress. In calculating test accuracies, these equivocal results are often excluded from the analysis or included but considered as normal results. Some investigators do not mention equivocal results or state that there were no such results. In these cases, it was assumed that such results, if any, were assigned either a normal or abnormal value based on some arbitrary criterion.

Recording of Test Accuracy

Sensitivity is the percentage of patients with angiographic disease who are correctly classified by an observed exercise-induced ST depression exceeding a predefined cutpoint value. Specificity is the percentage of patients without disease correctly identified by not having the above exercise result. These quantities were recorded from the publications reviewed.

Treatment of Missing Data

Missing data were frequently problematic. Figure 1 shows the frequency of missing data for the 18 variables for which some data were missing.

Because multivariate regression procedures require that all observations (study groups) have complete data, either potentially important continuous variables with frequently missing data must be deleted from the calculations or substituted by variables with complete data, or an estimate of the missing values made.

The second of these strategies (substitution) was applied for percent of patients with left ventricular hypertrophy, right bundle branch blocks, mitral prolapse, or resting repolarization abnormalities and percent of patients on β-receptor blockers or nitrates. This was done by assuming that there were patients with a characteristic (i.e., left ventricular hypertrophy) when it was so stated in a report or when no mention was made of the absence of same. The characteristic could then be assumed present
elicit bivariate relations between the variables and sensitivity and specificity.

**Multivariate Regression Analysis**

The variables listed in the last three columns of Table 1 were entered in a weighted, multivariate, linear regression analysis (SAS stepwise).\(^1\)\(^6\) This method calculates \(F\) statistics for each independent variable and begins by choosing the variable with the highest \(F\) statistic as long as they exceed a criterion value, corresponding to \(p=0.05\). It then proceeds to calculate and examine the \(F\) statistics for each of the remaining variables and chooses one for entry by the same criterion. After entering a variable, the method reexamines those already entered and deletes those that produce \(F\) statistics lower than the criterion. The recorded sensitivities and specificities were entered as dependent variables. The weights assigned to these calculations were the numbers of patients with angiographic coronary disease for sensitivity and the numbers without disease for specificity.\(^1\)\(^6\)

**Results**

Eighty of the 147 reports resulted from research done in North America; 59 were from Europe, five from Asia, and three from Australia. There were 150 study groups treated by the 147 reports. Three reports involved two study groups each. The 150 study groups were composed of 24,074 patients. Of these, 15,893 had angiographic coronary artery disease (defined as \(>50%\) diameter occlusion of at least one major vessel), whereas 8,181 did not.

Table 1 shows the variables considered in the analysis. The items in the first two columns of the table are the variables that contained so much missing data (more than 50% of the study groups) that they were not included in the multivariate analysis. The items in the last three columns of the table were included among the candidates for the multivariate analysis.

Figure 1 shows the percentage of missing data for those items in Table 1 that have data missing for any of the study groups. Only the six variables at the bottom of the figure (mean age, percent men, protocol used, number of leads, time of measurement, and angiographic disease definition) were candidates for the multivariate analysis because less than 35% of the data were missing for these variables. Of the items included in the multivariate analysis, mean age had the maximum percentage of missing data (32%).

Sensitivity could be determined for 144 of the study groups and specificity for 132. Figures 2 and 3 show the distributions of sensitivity and specificity in the groups. The weighted mean sensitivity was 68\(\pm\)16\% (SD); the weighted mean specificity was 77\(\pm\)17\%. The distributions illustrated in Figures 2 and 3 show the wide variability of the sensitivities and specificities reported in the literature.

**PERCENT MISSING DATA**

FIGURE 1. Frequency of missing data for all variables in Table 1 for which some observations were missing. Angina, history of typical angina pectoris; LVH, left ventricular hypertrophy; RBBB, right bundle branch block; MVP, mitral valve prolapse; rest ST, ST abnormalities at rest; intertest interval, number of days between exercise test and angiogram; time of measurement, number of msec after J point at which ST depression was measured; angio definition, 50% vs. 70% luminal narrowing.

(-1) or absent (-1), thus creating a dichotomous variable for all cases.

Other variables were deleted from the calculation when data were missing for more than 50% of the study groups. This was done for the percent of the study group with typical angina, percent achieving an adequate heart rate, mean work load, mean heart rate, and mean double product achieved, and time interval between the exercise test and the angiogram.

When data were available for at least 50% of the study groups, multivariate linear regression on all other independent variables in the last three columns of Table 1 was used\(^1\)\(^6\) to estimate the missing values. This imputation was performed for the mean age, percent men in study group, exercise protocol, number of ECG leads used for measurement, time of ST measurement (J point versus 80 msec), and angiographic definition of disease.

In this way, all observations for which sensitivity and specificity were available could be analyzed using the variables in the third, fourth, and fifth columns of Table 1.

**Bivariate Analysis**

Nonpaired \(r\) test was applied to the dichotomous variables, and Pearsonian correlation coefficients were calculated for the continuous variables to
Bivariate Analysis

Table 2 displays the results of the bivariate analysis. Because multiple comparisons were done, only results with conservative \( p \) values of 0.01 were used. Because none of the continuous variables showed a significant correlation with either sensitivity or specificity \((p \leq 0.01)\), only the results of the nonpaired \( t \) test are shown. The results of the bivariate analysis are seen to be similar to those of the multivariate analysis.

To be certain that the lack of correlation between some of the continuous variables and the sensitivity and specificity reported by the studies was not an artifact of the imputation used for the missing data, the bivariate analysis was done for these variables (mean age, percent men, number of leads, and time of measurement of ST segment) both before and after application of the imputation. With or without the imputation, these variables did not show a significant correlation with the sensitivity and specificity.

Some cardiologists might argue that studies that include equivocal exercise tests, compare the exercise-induced ST segment depression to a purportedly “better” test, and include patients taking digitalis glycosides should be excluded from the analysis. These three variables were found to be significantly related to sensitivity. However, when they were excluded from the analysis, the mean sensitivity in the 37 remaining study groups increased only to 72% and the SD remained constant at 16%. Similarly, when studies that included patients with prior myocardial infarction, left bundle branch block, or both were excluded, only 18 studies remained. The mean sensitivity actually decreased to 69%, and the SD increased to 24%.

Multivariate Analysis

Table 3 displays the results of the multivariate linear regression analyses for sensitivity and specificity. The second column of the table shows the partial \( R^2 \) or the percentage of the total variance in sensitivity and specificity accounted for by the independent variable listed in the first column. For example, considering equivocal results as normal accounted for 22% of the variance in sensitivity. The percentage of the variance explained by the entire model can be calculated by simply adding the numbers for sensitivity and specificity in the second column of the table. Thus, the multivariate models created by stepwise procedure accounted for 33% of the variance in sensitivity and 22% of the variance in specificity.

The regression coefficients in the third column of Table 3 show the relative weighting of the variables that enter the model. For example, including equivocal results at a given center would decrease sensitivity by about 7.7% according to the model for
TABLE 3. Variables Independently Associated With Sensitivity and Specificity as Determined by Stepwise Linear Regression

<table>
<thead>
<tr>
<th></th>
<th>Partial $R^2$</th>
<th>Regression coefficient</th>
<th>$p$ for entry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were equivocal results included and considered as normal? (yes: 1, no: -1)</td>
<td>0.219</td>
<td>-0.077</td>
<td>0.0001</td>
</tr>
<tr>
<td>Was the exercise ECG being compared with a “better” test? (yes: 1, no: -1)</td>
<td>0.068</td>
<td>-0.047</td>
<td>0.0003</td>
</tr>
<tr>
<td>Were patients on digitalis excluded? (yes: 1, no: -1)</td>
<td>0.023</td>
<td>0.033</td>
<td>0.008</td>
</tr>
<tr>
<td>Publication year</td>
<td>0.019</td>
<td>0.0061</td>
<td>0.047</td>
</tr>
<tr>
<td>Intercept</td>
<td>-11.467</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was upsloping ST depression considered abnormal? (yes: 1, no: -1)</td>
<td>0.094</td>
<td>-0.044</td>
<td>0.005</td>
</tr>
<tr>
<td>Were patients with prior infarction excluded? (yes: 1, no: -1)</td>
<td>0.060</td>
<td>-0.037</td>
<td>0.002</td>
</tr>
<tr>
<td>Were patients with left bundle branch block excluded? (yes: 1, no: -1)</td>
<td>0.042</td>
<td>0.032</td>
<td>0.007</td>
</tr>
<tr>
<td>Was hyperventilation used before exercise? (yes: 1, no: -1)</td>
<td>0.028</td>
<td>-0.064</td>
<td>0.040</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.692</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

sensitivity; comparing the exercise ECG with a “better” test decreases sensitivity by about 4.79%, and sensitivity increased by 0.61% a year since 1967. Similarly, considering upsloping ST depressions as abnormal would decrease specificity by about 4.4%, according to the model for specificity, and excluding patients with previous infarctions would decrease specificity by 3.79%. The last column of the table shows the $p$ value for entry, which is the significance level at which variables entered the model. Variables that did not achieve a $p$ value for entry of less than 0.05 were not allowed to enter. In general, the multivariate models correlated poorly with sensitivity and specificity.

Discussion

A patient’s probability of coronary disease after exercise electrocardiography depends on clinical characteristics and the sensitivity and specificity of the test result. For this reason, accurate estimates of sensitivity and specificity are required to accurately predict disease probability. If sensitivity and specificity remained relatively constant from one institution to another, a reasonable approach to obtaining accurate estimates would be to pool literature results and compute average values for these parameters. Figures 2 and 3 show that such an approach would result in erroneous sensitivities and specificities at many of the institutions because of the wide variability from one institution to another.

The present literature review fails to explain most of this variability in accuracy. Though measurement error and random variation due to differences in sample size have their effects, we propose other reasons for this failure. First, it is possible that the authors of the 147 reports did not disclose information that might affect sensitivity and specificity. This contention is supported by the fact that 33% of the data concerning the variables in Table 1 were missing in the reports. Underreporting may be responsible for the failure to explain this variability in accuracy. A second possible explanation is that unsuspected technical, methodologic, or clinical variables affect sensitivity and specificity by poorly understood mechanisms.

Table 3 lists the factors that were independently associated with sensitivity and specificity; these can be divided into methodologic and technical factors. Equivocal or nondiagnostic tests were considered as normal results in 16% of the reports, whereas 18% stated that such test results were totally excluded from the analysis. Usually, these “equivocal” results involved patients who failed to achieve a certain target heart rate, although some reports also included borderline or unclear tracings as equivocal. Because many of these patients undoubtedly had coronary heart disease, it is not surprising that sensitivity decreased when their results were counted as normal.

The results of this meta-analysis show that the sensitivity was significantly and adversely affected by comparison with tests that were reported as superior. When the exercise ECG is compared with a test purported to be superior in accuracy, more attention may be paid to the reading of the “better” test than is paid to the ECG. It is also likely that comparative reports of differences between the accuracy of a new technology and that of the exercise ECG are more favorably reviewed and more easily published. Though such investigator and journal reviewer bias may play a part in determining differences in reported accuracies, it may also be that the better test is applied to subjects unable to achieve high levels of cardiac stress or to subjects in whom the exercise ECG is difficult to interpret (e.g., patients with bundle branch block). Support of this contention is derived from the fact that reports involving such a comparison were more likely to include patients with left bundle branch block ($p=0.09$), right bundle branch block ($p=0.02$), resting ST abnormalities ($p=0.1$), and left ventricular hypertrophy ($p=0.2$). Investigator and journal reviewer bias as well as inclusion of equivocal ECGs can affect the decline in accuracy when such comparisons are made.

Table 3 reveals that exclusion of patients taking digitalis was associated with an improvement in mean sensitivity. Sensitivity in the 73 study groups without any digitalis ingestion was 70% as compared with 63% in the 77 groups in which some digitalis ingestion was present ($p=0.0003$). The
specificity in the groups not taking digitalis was not significantly different from the groups using digitalis. This finding of decreased sensitivity with little change in specificity is not what one would expect.\textsuperscript{162} One possible explanation is that patients taking digitalis have poorer left ventricular function and, therefore, have lower exercise capacity than those not taking such medication. For this reason, their exercise tests are liable to be terminated before ST depressions occur, thus resulting in false-negative tests and decreased sensitivity. Another possible explanation is that the investigator interpreting the test had knowledge of patient’s medication intake and was biased toward negative test interpretation to avoid false-positive ones. We have noted a similar bias in the interpretation of the exercise thallium scintigrams of women.\textsuperscript{163}

An increase in sensitivity and decrease in specificity as years go by is expected due to a tendency toward increased workup bias with time: As clinicians become more familiar with a test, they increasingly trust its results and allow these results to influence the decision to perform angiography. A similar pattern has been noted for radionuclide angiography\textsuperscript{164} and thallium scintigraphy.\textsuperscript{165}

Controversy exists as to whether upsloping ST depressions should be considered as abnormal test results.\textsuperscript{166} Upsloping ST depression was considered an abnormal finding in 58 study groups. The classification of upsloping results as abnormal lowered specificity significantly—73\% versus 80\% \textit{(p} = 0.005). Table 3 shows that this factor independently contributed to a lower specificity, which is in accord with the discretion commonly used when interpreting upsloping ST depressions.

The exclusion of subjects with previous myocardial infarction was associated with decreased specificity. Because exercise testing for the diagnosis of coronary artery disease is of little use in patients with previous infarction, the large number of groups (106) including such subjects is puzzling. This may be partially explained by the fact that the research often had other, nondiagnostic aims such as determining accuracy in predicting disease severity or prognosis. However, in only 17\% of the reports involving study groups with previous infarction was it possible to derive separately the test accuracy in those with and those without this condition. The absence of this information makes the interpretation of test accuracy difficult. Actual numbers with infarction were reported for only 66 out of the 106 groups, and in only 17 groups was it possible to separately calculate sensitivity for individual subjects with prior infarction. Because there were no significant differences in sensitivity in subjects with or without previous myocardial infarction in these 17 groups and no difference in specificity between the 106 groups including previous infarction and the 44 groups excluding it, the significant difference in specificity is even more enigmatic.

Increased specificity when patients with left bundle branch block are excluded is in accord with the findings of other investigators,\textsuperscript{167,168} but the association of hyperventilation with decreased specificity appears paradoxical at first glance. Hyperventilation is frequently used before exercise testing to unveil "labile" ST segment changes that might be associated with false-positive results.\textsuperscript{169} If hyperventilation were to cause respiratory alkalemia during exercise, it could decrease the threshold of ischemia\textsuperscript{170} in subjects with borderline angiographic disease and even in those with normal coronary arteries,\textsuperscript{171} which might decrease rather than increase specificity. The results of this meta-analysis support this concept. Careful research needs to be done concerning the persistence of the metabolic and physiologic effects of hyperventilation in subjects undergoing exercise testing.

\textbf{Study Limitations}

Although the multivariate stepwise regression analysis used in this investigation is useful for determining the independent predictors of accuracy and their effects, two limitations of this analysis need to be emphasized. The first involves the large quantity of missing data in the scientific literature reviewed. The second, mathematically related problem involves certain realities of stepwise regression. Because each step of this statistical technique always searches for the strongest association with the dependent variable (highest \textit{F} value), small differences in \textit{F} value for two variables that are associated with each other will cause one of them to be ignored. Thus variables might be rejected by the regression algorithm because of an insignificant difference in \textit{F} statistics.

\textbf{Conclusions}

The wide variability in reported sensitivity and specificity of exercise-induced ST depression cannot be explained by the data presented in the international medical literature, which is largely due to incomplete reporting of potentially important data involving population and technical factors. Reviewers and editors should carefully consider the documentation of important data such as the percent of patients with various chest pain syndromes and the workload and heart rate achieved when making decisions for publication of reports.

Defects in research methodology and selection bias may also play a role in the wide range of accuracies reported from different centers. Certain minimal standards regarding research methodology should be applied in making decisions for publication. Standardization of research methods and completeness of recording should be encouraged.

\textbf{Acknowledgment}

The editorial assistance of Ms. Maggie Meyer is deeply appreciated.
exercise radionuclide ventriculography in the prediction of coronary artery disease in men with chest pain. Am J Cardiol 1983;52:927–935


62. Gaslin VS, Lupanov VP, Mazaev VP, Gadzhaeva FU: Criteria of myocardial ischemia in bicycle ergometry of patients with stenosing arteriosclerosis of the cardiac coronary arteries depending on the initial ECG at rest. Kardiologiya 1980;20:68–72


75. Hollenberg M, Go M, Massie BM, Wisneski JA, Gertz EW: Influence of R-wave amplitude on exercise-induced ST depression: Need for a “gain factor” correction when
76. Hsu TS, Chiang CW, Lee YS, Hung JS: Treadmill exercise testing in patients with significant coronary artery disease. Taiwan I Hsueh Hui Tsai Chih 1984;83:155–166


152. Vincent Weiner


**KEY WORDS** • exercise testing • meta-analysis • coronary artery disease • coronary angiography
Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis.
R Gianrossi, R Detrano, D Mulvihill, K Lehmann, P Dubach, A Colombo, D McArthur and V Froelicher

_Circulation._ 1989;80:87-98
doi: 10.1161/01.CIR.80.1.87

_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1989 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/80/1/87