Exercise Echocardiography or Exercise SPECT Imaging?
A Meta-analysis of Diagnostic Test Performance

Kirsten E. Fleischmann, MD, MPH; Maria G. M. Hunink, MD, PhD; Karen M. Kuntz, ScD; Pamela S. Douglas, MD

Context.—Cardiac imaging has advanced rapidly, providing clinicians with several choices for evaluating patients with suspected coronary artery disease, but few studies compare modalities directly.

Objectives.—To review the contemporary literature and to compare the diagnostic performance of exercise echocardiography (ECHO) and exercise single-photon emission computed tomography (SPECT) imaging in the diagnosis of coronary artery disease.

Data Sources.—Studies published between January 1990 and October 1997 identified from MEDLINE search; bibliographies of reviews and original articles; and suggestions from experts in each area.

Study Selection.—Articles were included if they discussed exercise ECHO and/or exercise SPECT imaging with thallous chloride TI 201 (thallium) or technetium Tc 99m sestamibi for detection and/or evaluation of coronary artery disease, if data on coronary angiography were presented as the reference test, and if the absolute numbers of true-positive, false-negative, true-negative, and false-positive observations were available or derivable from the data presented. Studies performed exclusively in patients after myocardial infarction, after percutaneous transluminal coronary angioplasty, after coronary artery bypass grafting, or with recent unstable coronary syndromes were excluded.

Data Extraction.—Clinical variables, technical factors, and test performance were independently extracted by 2 reviewers on a standardized spreadsheet. Discrepancies were resolved by consensus.

Results.—Forty-four articles met inclusion criteria. In pooled data weighted by the sample size of each study, exercise ECHO had a sensitivity of 85% (95% confidence interval [CI], 83%-87%) with a specificity of 77% (95% CI, 74%-80%). Exercise SPECT yielded a similar sensitivity of 87% (95% CI, 86%-88%) but a lower specificity of 64% (95% CI, 60%-68%). In a summary receiver operating characteristic model comparing exercise ECHO performance to exercise SPECT, exercise ECHO was associated with significantly better discriminatory power (parameter estimate, 1.18; 95% CI, 0.71-1.65), when adjusted for age, publication year, and a setting including known coronary artery disease for SPECT studies. In models comparing the discriminatory abilities of exercise ECHO and exercise SPECT vs exercise testing without imaging, both ECHO and SPECT performed significantly better than exercise testing. The incremental improvement in performance was greater for ECHO (3.43; 95% CI, 2.74-4.11) than for SPECT (1.49; 95% CI, 0.91-2.08).

Conclusions.—Exercise ECHO and exercise SPECT have similar sensitivities for the detection of coronary artery disease, but exercise ECHO has better specificity and, therefore, higher overall discriminatory capabilities as used in contemporary practice.
overall has been hampered by significant variability in case mix and pretest probability of disease, differences in the positivity criterion used, and the lack of reliable techniques to combine reports on the performance of diagnostic tests. Therefore, we undertook a meta-analysis of the contemporary literature on exercise SPECT imaging and exercise ECHO using a new methodologic tool, summary receiver operating characteristic (SROC) analysis.10–12

METHODS

Data Extraction

A literature review was performed to identify articles published from January 1980 to October 1997 in the English-language literature on exercise SPECT imaging and exercise ECHO.12 We focused on this period because both exercise ECHO and exercise SPECT were in widespread clinical use. Also, these years postdated several potentially important developments in myocardial perfusion imaging, such as SPECT techniques and the introduction of technetium Tc 99m sestamibi as an imaging agent. Articles were obtained through a MEDLINE search using the key words coronary disease and exercise test in conjunction with either echocardiography, thallium or thallium isotopes, or sestamibi and restricted by the term human; from bibliographies of original and review articles; and in response to suggestions by experts in each area.

Articles3,5,8,14–16 were included in the analysis if (1) exercise ECHO and/or exercise SPECT imaging were performed with thallous chloride Tl 201 (thallium) or technetium Tc 99m sestamibi to detect and/or evaluate coronary artery disease (if data on exercise ECG results in the same patient population were also presented, these data were extracted, although a comprehensive search of the literature concerning standard exercise testing was not performed); (2) data on coronary angiography were presented as the reference test; and (3) the absolute numbers of true-positive, false-negative, true-negative, and false-positive observations were available or derivable from the data presented. Studies performed exclusively in patients after myocardial infarction (MI), after percutaneous transluminal coronary angioplasty, after coronary artery bypass grafting (CABG), or with unstable coronary syndromes were excluded. One stress ECHO article was excluded because of its substantially different positivity criterion, namely displacement of the atroventricular plane. In cases in which authors had published several reports in quick succession, we were able to contact them and to determine whether multiple analyses were reported on a substantially overlapping patient population. If this was the case, only 1 report was used in the analysis to avoid undue influence by results in one cohort of patients. A radiologist (M.G.M.H.) and cardiologist (K.E.F.) independently extracted data on clinical variables, technical factors, and test performance by means of a standardized spreadsheet. Discrepancies were resolved by consensus. Variables extracted included identifying information (first author, journal, city, and institution), publication year, mean age, percentage of male patients, clinical indication or setting and percentage of patients with previous MI, percutaneous transluminal coronary angioplasty, or CABG surgery. In addition, we recorded such test factors as the test used, the type of exercise performed, type of radionuclide used (if applicable), the positivity criterion used, the reference test used, how authors defined significant coronary artery disease (eg, ≥50% stenosis or ≥70% stenosis), and the presence of possible verification bias. The number of patients with and without disease, number of true positives and false positives, and percentage of diagnostic tests (patients reached at least 85% of their predicted maximal heart rate) were extracted as well as the percentage of uninterpretable tests and morbidity and mortality associated with each test, if provided.

If the article reported sensitivity and specificity data for important patient subsets, such as patients with multivessel disease, these data were extracted separately. If several types of tests were assessed in the same report (ie, exercise ECG, exercise ECHO, and exercise SPECT), data on test performance were extracted for each. An assessment of the likelihood of verification bias was made by noting whether the decision to proceed to the reference test, in this case contrast angiography, was in part dependent on the results of the noninvasive test. Since positive noninvasive test results are more likely to be followed up by an invasive test, this tends to increase the chance of detecting a true positive relative to a false negative and tends to increase the chance of detecting a false positive relative to a true negative. Therefore, sensitivity may appear falsely elevated and specificity falsely depressed in the verified sample.

Weighted Pooled Analysis

In our analysis, pooled results weighted by the sample size of each study were first calculated using a fixed-effects model.10 Pooled results in the analysis were also generated for important patient subsets, such as patients without prior MI and patients with severe coronary artery disease as defined in each report. We also explored the effect of varying positivity criteria on pooled results. For the analysis, positivity criteria were grouped according to whether an abnormal test result required that an abnormality manifest after exercise vs either at rest or after exercise. If not fully specified, a less stringent criterion of abnormality at rest or after exercise was used. In addition, exercise ECHO studies were grouped as requiring either new regional wall motion abnormalities (RWMAs) or new or worse RWMAs. All results are presented as percentage sensitivity and specificity with 95% confidence intervals (CIs).

SROC Analysis for Each Diagnostic Test

Receiver operating characteristic curves graphically represent the true-positive and false-positive ratios for a diagnostic test when the threshold for a positive test result is varied.11,12 The area under the curve summarizes the overall diagnostic performance of the test, with larger areas corresponding to a more discriminating test. An area of 0.5 corresponds to a test whose results are no better than chance, while an area of 1.0 denotes a perfect test. SROC analysis assumes that part of the variability in test performance reported in the literature is due to the use of different cut points or positivity criteria and adjusts for these differences. For example, a study of exercise ECHO may use a more stringent criterion such as development of a new RWMa to signal an abnormal test result or a more lenient criterion such as the absence of a hyperkinetic contractile response to exercise resulting in a different reported sensitivity and specificity. SROC analysis also allows adjustment for important clinical covariates and for comparison between different types of tests.11,12 In this last case, a variable corresponding to the type of exercise testing used (eg, exercise ECHO or exercise SPECT) is entered into the regression equation. The b-coefficients for the variable give a measure of the difference in diagnostic performance of the 2 tests, with positive coefficients indicating better discriminatory power and negative coefficients corresponding to reduced discriminatory ability.

SROC curves were generated separately for each test. We assessed the influence of variables including publication year; proportion of patients with prior MI; indication or setting for the test; mean age; proportion of men in the study population; presence of verification bias; the angiographic definition of disease (50% stenosis vs 70% or 75%); type of exercise (treadmill vs other); positivity criterion; and in the case of exercise SPECT, type of nuclide (thallium
vs technetium Tc 99m sestamibi) and type of analysis (visual vs quantitative) on test performance. We used a significance level of $P < .05$, implying that β regression coefficients whose 95% CIs do not cross zero are statistically significant. Since multiple variables were tested, one would normally adjust $P$ values for multiple comparisons, eg, with a Bonferroni correction. However, to minimize the chance of missing potential confounding factors, all explanatory variables with $P < .05$ were included in subsequent multivariate analyses. The Bonferroni-adjusted $P$ values in our analyses were approximately .005; therefore, explanatory variables with .005 < $P < .05$ should be interpreted with caution.

Data for some explanatory variables were not specified in all articles. For example, the proportion of prior percutaneous transluminal coronary angioplasty and CABG and the percentage of uninterpretable tests were excluded from all analyses due to missing data. Also, in the 1 article in which mean age was not given, in the 1 case in which the proportion of men was absent, in the 1 article in which the clinical setting for testing was not fully specified, and in the 2 articles in which the proportion of patients with MI was not specified, we used estimates for these variables based on a best subset regression analysis, so the widest range of literature could be included. Sensitivity analyses using weighted mean values for the missing data or excluding these articles from the analysis did not change the results substantively. Therefore, the base case including all articles is presented.

We then assessed which explanatory variables were significant in multivariate analysis for each test using stepwise forward regression analysis with backward elimination including all variables. All variables with a $P < .05$ were kept in the model. Variables with .05 < $P < .10$ were kept in the model if they substantially increased the explanatory power of the model.

**Comparison SROC Analyses Between Diagnostic Tests**

SROC curves were performed comparing exercise ECHO to exercise SPECT, exercise ECHO to exercise ECG, and exercise SPECT to exercise ECG. In each case, all explanatory variables were once again tested along with the variable indicating the test comparison. We then performed multivariate stepwise forward regression analysis with backward elimination including all variables with $P < .05$ in the step above as well as all variables significant in multivariate analysis for each individual test. To develop a parsimonious model with similar explanatory variables across all comparisons, we included in the final models significant variables with similar effects for all tests as well as interaction terms to account for significant variables with effects specific to a particular diagnostic test. Final models were tested for heterogeneity by comparing the values predicted by the model to the 95% CI of the actual values for each article. Similar SROC analyses were also performed in the following important subsets: studies comparing exercise ECHO and exercise SPECT directly in the same patients, patients without prior MI, and patients with severe disease as defined in each report.

**RESULTS**

**Overview of Studies**

A total of 902 citations published between January 1990 and October 1997 were screened, and 44 articles met inclusion criteria.3-14,45 The clinical characteristics of the studies analyzed are outlined in Table 1. Overall, 24 studies reported exercise ECHO results in 2637 patients with a weighted mean age of 59 years. The patients were predominantly men (69%). Significant coronary artery disease, as defined in each report, was present in 66% of patients and prior MI was present in 20% of patients. Twenty-seven studies reported exercise SPECT data on 3237 patients, of whom 70% were male, with a mean age of 59 years. Significant coronary artery disease was present in 78%, with past MI in 33%. In 24 articles on exercise ECHO and/or exercise SPECT, data on results of the exercise ECG in the same patient population were also given. These encompassed 2456 patients with a mean age of 59 years, of whom 66% were men. The prevalence of coronary artery disease was 66% with past MI in 20%.

**Weighted Pooled Results**

Diagnostic test performance for individual studies is outlined in Table 1. In pooled data weighted by the number of patients with and without disease in each study, exercise ECHO had a sensitivity (true-positive ratio) of 85% (95% CI, 83%-87%) with a specificity (1 − false-positive ratio) of 77% (95% CI, 74%-80%). The SPECT analyses yielded a similar sensitivity of 87% (95% CI, 86%-88%) but a lower specificity of 64% (95% CI, 60%-68%). Weighted pooled data for exercise ECG revealed a significantly lower sensitivity of 52% (95% CI, 50%-55%) and a specificity of 71% (95% CI, 68%-74%).

When only patients without prior MI were considered, the sensitivity and specificity for exercise ECHO were similar to those of the cohort overall at 87% (95% CI, 84%-89%) and 84% (95% CI, 81%-88%). Weighted pooled results for SPECT in this subset were also similar to overall results with a sensitivity of 86% (95% CI, 83%-88%) and a specificity of 62% (95% CI, 55%-70%).

The criterion for severe disease varied from article to article but generally denoted multivessel disease. In patients with severe disease, the sensitivity for exercise ECHO rose to 92% (95% CI, 90%-94%) when any normal test result was considered positive, but the specificity also dropped significantly to 54% (95% CI, 50%-58%). In contrast, when only abnormalities in multiple coronary distributions were considered an abnormal test result, the sensitivity fell to 60% (95% CI, 53%-67%) while specificity rose to 88% (95% CI, 85%-92%). Similar trends were seen for exercise SPECT with a sensitivity of 92% (95% CI, 90%-94%) and a specificity of 21% (95% CI, 19%-24%) when any abnormality was considered a positive test result, but a sensitivity of 44% (95% CI, 38%-49%) and a specificity of 87% (95% CI, 85%-89%) when only abnormalities in multiple distributions were considered positive.

Pooled results were analyzed by positivity criteria as outlined in the "Methods" section and compared with the overall results. No significant differences in sensitivity were detected for either exercise ECHO or exercise SPECT. However, specificity in reports requiring an abnormality seen only after exercise rather than at rest or after exercise was significantly higher for exercise SPECT (76% [95% CI, 69%-83%] vs 61% [95% CI, 57%-65%]) and for exercise ECHO (84% [95% CI, 81%-87%] vs 70% [95% CI, 66%-74%]). Specificity was also higher for exercise ECHO reports requiring a new RWMA than for those using a new or worsened RWMA as a positivity criterion (80% [95% CI, 78%-82%] vs 75% [95% CI, 72%-78%]).

**SROC Analysis for Each Diagnostic Test**

In univariate analysis for exercise ECHO, age and publication year were significant predictors. The discriminatory power of the test diminished slightly with increasing mean age (β-coefficient, −0.22 per year; 95% CI, −0.31 to −0.12) and with later publication year (−0.41 per year; 95% CI, −0.58 to −0.24). Performance was better in reports requiring an abnormality seen only after exercise rather than at rest or after exercise (0.87; 95% CI, 0.065-1.74). Sex was not a significant predictor (0.04; 95% CI, −0.99 to 1.79). Similarly, increasing age (−0.15; 95% CI, −0.24 to −0.06) was a significant predictor of SPECT performance. Unlike exercise ECHO, settings that in-
cluded known coronary artery disease vs suspected coronary artery disease only were associated with better discriminatory performance (1.14; 95% CI, 0.59-1.70) as were higher proportions of men in the study population (1.89; 95% CI, 0.72-3.07). No predictors were significant for exercise ECG.

In multivariate analysis, increasing age (−0.16; 95% CI, −0.23 to −0.09) remained a significant predictor of performance for exercise ECHO, as did a setting including known coronary artery disease (−0.56; 95% CI, −1.00 to −0.12). For exercise SPECT, a setting including...
known or suspected coronary artery disease (1.19; 95% CI, 0.67-1.70) in the study population was associated with improved performance. For both tests, performance diminished with later publication year with a parameter estimate of −0.27 (95% CI, −0.40 to −0.14) for ECHO and −0.15 (95% CI, −0.28 to −0.012) for SPECT. No explanatory variables were significant for exercise ECG in multivariate analysis.

### Comparison SROC Analyses Between Diagnostic Tests

In a model comparing exercise ECHO performance to SPECT, exercise ECHO displayed significantly better discriminatory power (parameter estimate for incremental improvement, 1.18; 95% CI, 0.71-1.65), when adjusted for age, publication year, and a setting including known coronary artery disease for SPECT studies (Table 2). In models comparing the performance of exercise ECHO vs exercise ECG and exercise SPECT vs exercise ECG, both ECHO and SPECT performed significantly better than exercise ECG (Table 2). The incremental improvement was greater for ECHO than for SPECT. These results were presented graphically in Figures 1 and 2. In the subset of reports (n=6) comparing exercise ECHO and exercise SPECT directly in the same patients, exercise ECHO also displayed better discriminatory performance parameter estimate (1.06; 95% CI, 0.77-1.36) in a model adjusted for age.

In patients without prior MI, results were qualitatively similar to the overall analysis. When exercise ECHO and exercise SPECT were compared in this subgroup, exercise ECHO had better discriminatory power (1.21; 95% CI, 0.57-1.85). Both exercise SPECT and exercise ECHO performed significantly better than exercise ECG in patients without prior MI, but the incremental improvement for exercise ECHO was larger (2.35; 95% CI, 1.79-2.91) than for exercise SPECT (1.36; 95% CI, 0.58-2.15). Results were also similar for models assessing diagnostic performance in patients with severe coronary artery disease as defined by each report. Exercise ECHO performed significantly better than exercise SPECT (parameter estimate, 0.91; 95% CI, 0.26-1.56). This result should be interpreted with caution, however, given the varying definitions of severe coronary artery disease from study to study and the relatively few studies in which these data were reported.

#### Table 1.—Clinical Characteristics of the Studies Included in the Meta-analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Mean Age, y</th>
<th>Men, %</th>
<th>Previous MI, %</th>
<th>Exercise Used</th>
<th>Radionuclide Used</th>
<th>Verification Bias</th>
<th>Exercise ECG Results</th>
<th>True Positive</th>
<th>False Negative</th>
<th>True Positive</th>
<th>False Negative</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECHO and SPECT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hecht et al¹</td>
<td>58</td>
<td>86</td>
<td>24</td>
<td>Treadmill/ bicycle</td>
<td>TI No Yes</td>
<td>New or worsening RWMA at rest</td>
<td>46</td>
<td>5</td>
<td>16</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoffmann et al²</td>
<td>57</td>
<td>77</td>
<td>0</td>
<td>Bicycle</td>
<td>MIBI No Yes</td>
<td>ECHO new RWMA after exercise</td>
<td>40</td>
<td>10</td>
<td>14</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pozzoli et al³</td>
<td>52</td>
<td>87</td>
<td>19</td>
<td>Bicycle</td>
<td>MIBI Yes Yes</td>
<td>ECHO new RWMA after exercise</td>
<td>35</td>
<td>14</td>
<td>25</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salustri et al⁴</td>
<td>59</td>
<td>80</td>
<td>36</td>
<td>Bicycle</td>
<td>MIBI/TI Likely</td>
<td>ECHO new RWMA after exercise</td>
<td>35</td>
<td>14</td>
<td>25</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinones et al⁵</td>
<td>57</td>
<td>67</td>
<td>. . .</td>
<td>Treadmill</td>
<td>TI Yes No</td>
<td>ECHO new or worse RWMA at rest</td>
<td>64</td>
<td>22</td>
<td>23</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marwick et al⁶</td>
<td>59</td>
<td>70</td>
<td>0</td>
<td>Bicycle</td>
<td>MIBI No Yes</td>
<td>ECHO RWMA after exercise</td>
<td>49</td>
<td>7</td>
<td>24</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Mean Age, y</td>
<td>Men, %</td>
<td>Previous MI, %</td>
<td>Exercise Used</td>
<td>Radionuclide Used</td>
<td>Verification Bias</td>
<td>Exercise ECG Results</td>
<td>True Positive</td>
<td>False Negative</td>
<td>True Positive</td>
<td>False Negative</td>
<td>False Positive</td>
</tr>
<tr>
<td><strong>Comparison Variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise ECHO vs exercise SPECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHO vs SPECT</td>
<td>1.18 (0.71 to 1.65)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, per year</td>
<td>−0.12 (−0.17 to −0.06)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication year</td>
<td>−0.21 (−0.30 to −0.11)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPECT mixed setting</td>
<td>0.77 (0.26 to 1.29)</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise SPECT vs exercise ECG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPECT vs ETT</td>
<td>1.49 (0.91 to 2.08)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, per year</td>
<td>−0.047 (−0.093 to −0.0004)</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPECT publication year</td>
<td>−0.15 (−0.27 to −0.029)</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPECT mixed setting</td>
<td>1.01 (0.52 to 1.50)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise ECHO vs exercise ECG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHO vs ETT</td>
<td>3.43 (2.74 to 4.11)</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, per year</td>
<td>−0.066 (−0.11 to −0.021)</td>
<td>.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHO publication year</td>
<td>−0.35 (−0.50 to −0.21)</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ECHO indicates echocardiography; SPECT, single-photon emission computed tomography; ECG, electrocardiography; MI, myocardial infarction; MIBI, technetium Tc 99m sestamibi; TI, thallous chloride TI 201 (thallium); RWMA, regional wall motion abnormality; and PD, perfusion defect. Ellipses indicate that data is not available.

---

©1998 American Medical Association. All rights reserved.

Downloaded From: https://jama.jamanetwork.com/ by a Non-Human Traffic (NHT) User on 04/26/2019
The base case is adjusted to a population aged 59 years, publication year 1993, and a setting including known or suspected coronary artery disease (CAD) with exercise electrocardiography. In all panels, the horizontal axis represents the false-positive ratio (1 − specificity) and the vertical axis the true-positive ratio (sensitivity). The case is adjusted to a population aged 59 years, publication year 1993, and a setting including known or suspected CAD in all cases, and worst case, which is adjusted to age 65 years, publication year 1995, and a setting with suspected CAD only in all cases.

Final models were tested for heterogeneity by comparing the model predicted discriminatory power to the 95% CIs of the actual discriminatory power for each study. Analyses were homogeneous with the exception of the following: in the models comparing exercise echocardiography (ECHO), exercise single-photon emission computed tomography (SPECT), and exercise testing without adjunct imaging. The horizontal axis represents the false-positive ratio (1 − specificity) and the vertical axis the true-positive ratio (sensitivity).

Testing for Heterogeneity

Our data also show that both exercise SPECT imaging and exercise ECHO have significantly higher discriminatory capabilities than exercise ECG in these patient populations, although the incremental improvement in performance is greater for exercise ECHO. Detrano et al.56 have suggested in previous meta-analyses of exercise testing that the incremental value of imaging may be overestimated when data on exercise ECG are drawn from reports in which exercise ECG is compared to a “better” test. This may, in part, be related to study of cohorts enriched for patients with baseline electrocardiographic abnormalities or other conditions affecting accuracy of the exercise ECG. In addition, the data on exercise ECG in reports focusing on an imaging modality may not include detailed information on important variables such as exercise duration, symptoms, or degree of ST-segment changes. They do, however, represent contemporaneous data on exercise ECG results in the same clinical cohorts. Our estimates of the performance of exercise ECG lie within the 95% CIs generated in the meta-analysis by Detrano et al and are unlikely to give qualitatively varying results. In actual clinical practice, the performance of an imaging test such as SPECT or ECHO with exercise is not pitted against the performance of standard exercise testing but, rather, is used in conjunction with information from the exercise test to enhance its diagnostic and prognostic capabilities.57

Performance for exercise ECHO and exercise SPECT in the subset of patients without evidence of prior MI was similar to that for each test in the cohort overall. Interestingly, however, in multivariate analysis, a setting with known or suspected coronary artery disease (a proxy for pretest probability of disease) was associated with somewhat lower discriminatory power for exercise ECHO but led to improved discrimination for exercise SPECT. These findings may be related to the differing nature of exercise ECHO, exercise SPECT, and exercise ECG.
which relies on the functional response of the myocardium to exercise, and exercise SPECT, in which perfusion is gauged as normal or abnormal. For example, the baseline RWMA associated with prior infarction may make interpretation of exercise ECHO more difficult if a positivity criterion requiring development of new or worsening RWMA after exercise is used but may not influence interpretation of perfusion in an identical fashion. The findings may also relate to differences in positivity criteria, although the SROC methods used should adjust for this in large part.

Although these data suggest that exercise ECHO performs as well or better than exercise myocardial perfusion scintigraphy with SPECT imaging, an alternative interpretation should be considered. Several studies have reported that the specificity of diagnostic tests decreases with time as the test is used more widely in clinical decision making and is applied to a wider spectrum of patient populations. This is consistent with our finding that publication year is a significant predictor of the performance of each test and that discriminatory power decreases in more recent reports. Since myocardial perfusion scintigraphy has been in widespread clinical use since the early 1980s and as exercise ECHO has become popular more recently, the difference in specificity seen could be a function of this different experience. However, in the subset of studies comparing the 2 modalities directly in the same patient population, exercise ECHO performed significantly better, mitigating against this possibility. These direct comparisons may still be influenced by verification bias (also known as posttest referral bias) and the tendency to confirm noninvasive test findings more frequently in patients with abnormal test results, but the degree of bias should be similar for the 2 tests.

Several technical factors related to test performance were not significant predictors of discriminatory ability, although pairs of sensitivity and specificity can still be different in that they may be shifted along the same receiver operating characteristic curve. For example, the distinction between bicycle and treadmill exercise was not a significant predictor of exercise ECHO performance, suggesting that both can give similar results. Similarly, variables comparing thallium stress test with technetium Tc 99m sestamibi or distinguishing visual from quantitative or computer-based interpretation of exercise SPECT were also not significant in the multivariate analysis.

Some limitations of the literature reviewed deserve note. Information on the incidence of uninterpretable test results for either imaging modality was often sketchy or absent. In some reports, patients with inadequate imaging were excluded from the analysis, inflating diagnostic test performance and obviating meaningful adjustment for this parameter. Some insight may be gained from the large series comparing the 2 modalities head to head in the same patients by Quinones et al. In this series of 292 patients, 99% of ECHO and 100% of SPECT studies were adequate for interpretation. Information on other potentially important factors such as test performance in relation to sex or body size was also rarely given. Verification bias remains a concern since recent studies suggest substantial effects on reported sensitivities and specificities, although the likely presence or absence of verification bias was not a significant predictor for the comparison of exercise SPECT and exercise ECHO. Ongoing technological developments for both tests may also affect test performance.

For example, recent reports using gated SPECT imaging suggest this technique can enhance specificity for the diagnosis of coronary artery disease, particularly in patient subsets prone to artifact, such as women. Despite these limitations, an aggregate of the comparative literature likely represents the best factual information available regarding exercise imaging.

Our conclusions should also be viewed in light of study design considerations that could influence results. Meta-analysis of diagnostic tests may be subject to publication bias. That is, reports that present the diagnostic test under study in a favorable light may more likely be submitted and published than those showing little or no benefit. However, this bias affects the literature for both exercise SPECT and exercise ECHO and visual inspection of a plot of study size vs discriminatory power for each test indicated little systematic bias. Our analysis uses partially dependent data in that multiple tests were performed in some patients, but a sensitivity analysis randomly selecting 1 test per study yielded similar results. The analysis focuses on diagnostic test performance for coronary artery disease and does not incorporate potential differences in prognostic capabilities. Finally, these results cannot be fully generalized to a number of important subgroups, such as patients undergoing pharmacologic stress testing with adjunctive imaging or those undergoing exercise ECHO or exercise SPECT for other indications, such as risk stratification for recent angina or unstable MI. Because our search was conducted in English-language literature and reports were retrieved primarily from the United States and Canada, our conclusions may also be applicable primarily to practices in those countries.

Our analysis of the current literature shows that exercise ECHO and exercise SPECT have similar sensitivities for the detection of coronary artery disease; however, that exercise ECHO has better specificity and higher overall discriminatory capabilities as used in contemporary practice. Previous individual studies have had limited power to detect differences in the performance of the 2 tests, but in aggregate, these differences become significant. While exercise ECHO may, therefore, be preferable in most instances, the clinician’s ultimate choice of imaging modality must take into account other important factors such as local expertise and experience with each test, availability, and cost.

Kirsten E, Fleischmann, MD, MPH, is the recipient of a Clinical Investigator Development Award (1K08HL02964-01) from the National Heart, Lung and Blood Institute, Bethesda, Md. Dr Fleischmann and Ms Kuntz were also supported by a project grant from the American Society of Electrocardiography. Dr Hunink was supported by a Promotie Impuls voor Onderzoeksgroepen met Nieuwe Ideeën voor Excellente Research (PIONIER) award from the Netherlands Organization for Scientific Research, The Hague, Netherlands.

The authors thank Finn Manning, MD, for his thoughtful comments.

References


