Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis

Catherine Kim, MD, MPH,a Yeong S. Kwok, MD,c Patrick Heagerty, PhD,b and Rita Redberg, MD, MSc, FACCd Seattle, Wash, and San Francisco, Calif

Background Although noninvasive pharmacologic stress tests are widely used, their relative performance is not clear. We compared the performance of pharmacologic stress tests combined with echocardiography or nuclear imaging for the diagnosis of coronary disease.

Methods We performed a regression meta-analysis of published data. We included studies published between January 1975 and June 1999 in which subjects underwent echocardiographic or single-photon emission computed tomography (SPECT) stress testing with adenosine, dipyridamole, or dobutamine for diagnosis of coronary artery disease. All subjects also underwent coronary angiography. Two independent reviewers abstracted population characteristics, technical factors, methodologic factors, and results and calculated test sensitivity and specificity.

Results Eighty-two studies met the inclusion criteria. The sensitivity of dipyridamole SPECT imaging, 89% (95% CI, 84%-93%), was higher than that of dipyridamole echocardiography, but the specificity of dipyridamole SPECT imaging, 65% (95% CI, 54%-74%), was lower than that of dipyridamole echocardiography. Dipyridamole and adenosine tests had similar sensitivities and specificities. The sensitivity of dobutamine echocardiography, 80% (95% CI, 77%-83%) was similar to that of dobutamine SPECT imaging, but dobutamine echocardiography had a higher specificity, 84% (95% CI, 80%-86%) than dobutamine SPECT imaging did.

Conclusions The findings of our study can be used to guide the selection of the optimal pharmacologic stress test for each patient. Maximum sensitivity can be attained by use of a vasodilator combined with SPECT imaging. Maximum specificity can be attained by use of a vasodilator with echocardiography. The highest combination of sensitivity and specificity can be attained with dobutamine echocardiography. (Am Heart J 2001;142:934-44.)
Methods
Search for published data

We reviewed articles that studied noninvasive tests for CAD diagnosis using the inotropic agent dobutamine or the vasodilat-
ging agents dipyridamole or adenosine. We searched MEDLINE
for English language studies with human subjects for each of
these stressors using the search strategy “coronary disease/diag-
nosis AND (pharmacologic stress).” The searches covered a
time period of January 1975 to June 1999. We also reviewed
the reference lists of review articles and eligible studies and
consulted with experts to complete the data search.

Selection criteria

Entry criteria for studies were (1) all subjects underwent at
least one pharmacologic stress test with either echocardiogra-
phy or SPECT and coronary angiography, the reference stan-
dard for CAD diagnosis, and (2) data presentation in a manner
that allowed calculation of the sensitivity and specificity of
the tests. Excluded were (1) studies for post-myocardial infar-
ction risk stratification, post–coronary artery bypass grafting or
postangioplasty evaluation, or cardiac transplant evaluation,
(2) studies combining exercise with pharmacologic stress
test (except in the case of hand grip exercise), (3) studies using
oral pharmacologic agents, (4) studies in special subgroups
such as patients with significant chronic renal insufficiency or
aortic stenosis, and (5) studies that likely presented duplicate
data. These studies included data from previous publications,
usually presented in a different analysis. In these cases, only
the study with the largest number of subjects was selected for
inclusion.

Data collection

Two reviewers independently abstracted the eligible arti-
cles. Disagreements between reviewers were resolved by con-
ference. Information abstracted from each report included
population characteristics, technical factors, methodologic
factors, and results. For each study we recorded publication
year, type of test, total number of participants, number of
male and female participants, mean age, type of pharmaco-
logic stress, type of imaging modality, percentage of patients
with myocardial infarction (MI), angiographic definition of
coronary disease, and percentage of participants with CAD by
angiography (Table I). For studies of nuclear perfusion, we
recorded whether thallium 201 or technetium 99m sestamibi
was the nuclear isotope and whether all defects or only
reversible defects were considered “positive” tests.

For subgroup analysis, we abstracted data on test perfor-
ance by sex if the data were available. We compared the
performance of thallium versus sestamibi isotope. If available,
we also recorded test performance in multivessel disease. The
literature varied on how detection of multivessel disease was
defined. In some studies, detection of multivessel disease was
defined as any positive test. In others, the sensitivity and
specificity for multivessel disease was determined by abnor-
malities in 2 or more vascular territories. We calculated test
performance for both definitions and present them separately.
We rated methodologic quality on the basis of 3 criteria:
adequate description of the study group, potential for verifica-
tion bias, and potential for diagnostic and test review bias. An
adequate description entailed clearly defined participant
selection criteria and characteristics. Verification bias was
avoided if the results of the pharmacologic stress test did not
influence the decision to perform angiography. Diagnostic
and test review biases were absent if both the noninvasive
test and angiography were read blindly. We classified studies
“high quality” if they met all 3 methodologic criteria,
“medium quality” if they met 2 criteria, and “low quality” if
they met 1 or none of the criteria (Table I).

Statistical analysis

For each study we calculated sensitivity and specificity.
Using the sample size for the patients with and without CAD,
we calculated weighted average sensitivities and specificities
and constructed SROC curves for each pharmacologic stress
test. Use of an ROC curve allows simultaneous evaluation of
sensitivity and specificity and facilitates test comparison. An
ROC regression analysis also can determine the influence of
covariates on test accuracy.6

For SROC regression analysis, we performed a weighted
least-squares regression analysis using D = log (sensitivity) –
log (1 – specificity) as the response variable with S = log (sen-
sitivity) + log (1-specificity) as one of the predictor variables.6
With use of a model that specifies D = α + β · S defines an
ROC curve by the intercept α and the slope β. Covariates may
influence both the intercept or the slope. If a covariate, X,
only influences the SROC intercept and not the slope (ie, is a
main effect), then the ROC curves for different values of X
will not cross; hence X is purely associated with differences
in accuracy. We used weights for each observation that were
proportional to the inverse of the variance of D. We also used
SE estimates that are robust to misspecification of the variance
(ie, corrected for possible overdispersion) according to the
method of Huber.7

To determine whether certain study-specific covariates are
correlated with test accuracy, we used SROC regression mod-
els that included the additional variables: publication year, age
of subjects, methodologic quality (high, medium, low),
absence of verification bias, definition of coronary disease
(50% or 70% stenosis), coronary disease prevalence, percent-
age of subjects with history of myocardial infarction, and per-
centage of male subjects. For nuclear studies, we analyzed the
effect of choice of isotope (thallium or sestamibi), and the def-
ingition of a positive test. We included the factors with a uni-
variate P < .20 in a multivariate regression and then used
backward elimination to remove variables that did not achieve
a .05 level of significance. Subgroup analyses were carried out
for results for men and women separately and multivessel dis-
ease, if there were more than 3 studies that presented data for
each subgroup. All calculations were performed with S-PLUS
software.9

Results

Search for published data

Initial data searches yielded 605 titles for studies
using dipyridamole, 474 titles for studies using dobuta-
mine, and 297 titles for studies using adenosine, for a
total of 1379 studies. By reviewing titles and abstracts,
we excluded 1129 articles with no original data and
### Table I. Characteristics of studies included in meta-analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Test(s) studied</th>
<th>No.</th>
<th>Women (%)</th>
<th>With CAD (%)</th>
<th>Mean age (y)</th>
<th>With prior MI (%)</th>
<th>CAD definition*</th>
<th>Verification†</th>
<th>Quality‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aksut et al⁸</td>
<td>1995</td>
<td>AS</td>
<td>443</td>
<td>40</td>
<td>90</td>
<td>63</td>
<td>24</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Allman et al⁹</td>
<td>1992</td>
<td>AS</td>
<td>76</td>
<td>43</td>
<td>79</td>
<td>63</td>
<td>0</td>
<td>50%</td>
<td>No</td>
<td>Medium</td>
</tr>
<tr>
<td>Amanullah et al¹⁰</td>
<td>1993</td>
<td>AE, AS</td>
<td>40</td>
<td>20</td>
<td>85</td>
<td>61</td>
<td>30</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Amanullah et al¹¹</td>
<td>1997</td>
<td>AS</td>
<td>222</td>
<td>46</td>
<td>77</td>
<td>71</td>
<td>0</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Anthopoulos et al¹²</td>
<td>1997</td>
<td>AE, DE</td>
<td>128</td>
<td>39</td>
<td>76</td>
<td>73</td>
<td>30</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Anthopoulos et al¹³</td>
<td>1996</td>
<td>AE, DE</td>
<td>120</td>
<td>40</td>
<td>74</td>
<td>75</td>
<td>40</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Beer et al¹⁴</td>
<td>1989</td>
<td>PS</td>
<td>65</td>
<td>40</td>
<td>74</td>
<td>61</td>
<td>45</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Beleslin et al¹⁵</td>
<td>1999</td>
<td>DE, PE</td>
<td>168</td>
<td>18</td>
<td>76</td>
<td>51</td>
<td>50</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Burns et al¹⁶</td>
<td>1991</td>
<td>PS</td>
<td>16</td>
<td>31</td>
<td>50</td>
<td>63</td>
<td>6</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Cecil et al¹⁷</td>
<td>1996</td>
<td>PS</td>
<td>211</td>
<td>39</td>
<td>76</td>
<td>73</td>
<td>27</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁸</td>
<td>1993</td>
<td>DE</td>
<td>52</td>
<td>2</td>
<td>71</td>
<td>63</td>
<td>25</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁹</td>
<td>1997</td>
<td>AS</td>
<td>222</td>
<td>46</td>
<td>77</td>
<td>71</td>
<td>0</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Anthopoulos et al¹²</td>
<td>1997</td>
<td>AE, AS</td>
<td>40</td>
<td>20</td>
<td>85</td>
<td>61</td>
<td>30</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Anthopoulos et al¹³</td>
<td>1996</td>
<td>AE, DE</td>
<td>120</td>
<td>40</td>
<td>74</td>
<td>75</td>
<td>40</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Beer et al¹⁴</td>
<td>1989</td>
<td>PS</td>
<td>65</td>
<td>40</td>
<td>74</td>
<td>61</td>
<td>45</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Beleslin et al¹⁵</td>
<td>1999</td>
<td>DE, PE</td>
<td>168</td>
<td>18</td>
<td>76</td>
<td>51</td>
<td>50</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Burns et al¹⁶</td>
<td>1991</td>
<td>PS</td>
<td>16</td>
<td>31</td>
<td>50</td>
<td>63</td>
<td>6</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Cecil et al¹⁷</td>
<td>1996</td>
<td>PS</td>
<td>211</td>
<td>39</td>
<td>76</td>
<td>73</td>
<td>27</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁸</td>
<td>1993</td>
<td>DE</td>
<td>52</td>
<td>2</td>
<td>71</td>
<td>63</td>
<td>25</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁹</td>
<td>1997</td>
<td>AS</td>
<td>222</td>
<td>46</td>
<td>77</td>
<td>71</td>
<td>0</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Anthopoulos et al¹²</td>
<td>1997</td>
<td>AE, AS</td>
<td>40</td>
<td>20</td>
<td>85</td>
<td>61</td>
<td>30</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Anthopoulos et al¹³</td>
<td>1996</td>
<td>AE, DE</td>
<td>120</td>
<td>40</td>
<td>74</td>
<td>75</td>
<td>40</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Beer et al¹⁴</td>
<td>1989</td>
<td>PS</td>
<td>65</td>
<td>40</td>
<td>74</td>
<td>61</td>
<td>45</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Beleslin et al¹⁵</td>
<td>1999</td>
<td>DE, PE</td>
<td>168</td>
<td>18</td>
<td>76</td>
<td>51</td>
<td>50</td>
<td>50%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Burns et al¹⁶</td>
<td>1991</td>
<td>PS</td>
<td>16</td>
<td>31</td>
<td>50</td>
<td>63</td>
<td>6</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Cecil et al¹⁷</td>
<td>1996</td>
<td>PS</td>
<td>211</td>
<td>39</td>
<td>76</td>
<td>73</td>
<td>27</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁸</td>
<td>1993</td>
<td>DE</td>
<td>52</td>
<td>2</td>
<td>71</td>
<td>63</td>
<td>25</td>
<td>70%</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Cohen et al¹⁹</td>
<td>1997</td>
<td>AS</td>
<td>222</td>
<td>46</td>
<td>77</td>
<td>71</td>
<td>0</td>
<td>50%</td>
<td>No</td>
<td>Low</td>
</tr>
</tbody>
</table>

AS, Adenosine SPECT; AE, adenosine echocardiography; DE, dobutamine echocardiography; PS, dipyridamole SPECT; PE, dipyridamole echocardiography; DS, dobutamine SPECT.

*Luminal narrowing of ≥1 coronary artery on angiography required for diagnosis of CAD.
†Absence of verification bias.
‡Quality score; see Methods.
studies that did not include both angiography and stress testing. We reviewed the text of the remaining articles and further excluded articles using the above criteria, including 29 studies with likely duplicate data, although the objectives of the studies differed.91-117 Thus we identified 82 studies that met all the inclusion criteria.8-89

Some studies included data on more than one test. Overall, 6 studies presented data on adenosine echocardiography, 9 on adenosine SPECT imaging, 40 on dobutamine echocardiography, 14 on dobutamine SPECT imaging, 20 on dipyridamole echocardiography, and 21 on dipyridamole SPECT imaging (Table II).

SPECT imaging versus echocardiography

There were 3737 patients in 44 studies using SPECT imaging and 6448 patients in 66 studies using echocardiography. Vasodilator SPECT imaging offered higher sensitivity but lower specificity than did vasodilator echocardiography. Dobutamine SPECT imaging offered similar sensitivity but lower specificity than did dobutamine echocardiography (Table II).

SPECT imaging

Of the 3 pharmacologic stressors, dipyridamole was the most commonly combined with SPECT imaging. SPECT studies used either thallium or sestamibi as the nuclear isotope. There were 2112 patients in 23 studies with thallium and 1625 patients in 21 studies with sestamibi. The 2 isotopes had similar sensitivities and specificities for all pharmacologic stressors (results not shown). The majority of studies defined a positive test as any imaging abnormality; only 2 studies defined a positive test as a “reversible” defect.

Of the 30 studies of vasodilator SPECT imaging, most used thallium as the imaging agent. The sensitivity and specificity of dipyridamole SPECT imaging were not significantly different from those of adenosine SPECT imaging (Table II). In contrast, most of the 14 studies of dobutamine SPECT studies used sestamibi as the imaging agent. Compared with studies of vasodilator SPECT, dobutamine SPECT studies had a lower sensitivity but similar specificity (Table II).

Echocardiography

Of the 3 pharmacologic stressors, dobutamine was the most commonly combined with echocardiography. Dobutamine echocardiography had a higher sensitivity but a lower specificity than did dipyridamole or adenosine echocardiography. The 20 studies of dipyridamole echocardiography had a sensitivity and specificity that
were not significantly different from those of adenosine echocardiography (Table II).

**Sex-specific results**

Few studies presented sex-specific data (Table III). There were only enough studies for subgroup analysis for dobutamine echocardiography and dipyridamole echocardiography for men and dobutamine echocardiography for women. There were no statistically significant differences between the sex-specific results and the results for all patients because many of the studies included women but did not report results separately. The confidence intervals are broad, reflecting the paucity of studies and subjects that presented sex-specific data (Table III).

**Multivessel disease**

We examined the issue of multivessel disease in 2 ways. First, we looked only at whether the noninvasive test showed any imaging abnormality in the presence of multivessel disease on cardiac catheterization. In this instance we did not take into account the degree of abnormality or the number of vascular territories identified on the noninvasive imaging studies. Thus an echocardiographic study showing only a lateral wall motion abnormality in a subject with triple-vessel disease would count as a true-positive test. By this definition the sensitivity of finding any imaging abnormality can be calculated in patients with multivessel disease. Many studies used this definition (Table IV).

The second definition of multivessel disease required that noninvasive imaging studies show abnormalities in ≥2 vascular territories to be considered positive for the presence of multivessel disease. For example, an echocardiographic study showing only a lateral wall motion abnormality in a subject with triple-vessel disease would count as a negative test. At least 2 vascular territories would have to be abnormal on the echocardiogram for the test to be positive. Exact matching of the abnormal imaging areas to the diseased vessels was not required. This allows for the calculation of the sensitivity and specificity for the finding of imaging abnormalities in multiple vascular territories. Fewer studies presented data in this format (Table IV). As expected, the sensitivity calculated by use of this definition for the detection of multivessel disease is significantly lower than the less-exact first definition. However, the specificity of finding imaging abnormalities in >2 vascular territories compares favorably with the specificities for the detection of CAD in general. In particular, dipyridamole and dobutamine nuclear imaging studies have higher specificities for multivessel CAD detection than they do for general CAD detection. Thus the presence of multiple imaging abnormalities seems to be a specific but not a sensitive marker of multivessel disease.

**SROC analysis**

The shape of the ROC curve for dobutamine is different from the ROC curve for dipyridamole, with dobutamine having a sharper increase in sensitivity for a given increase in the false-positive rate. Figure 1 shows the
weighted average (true-positive, false-positive) pairs for the combinations of stressor and imaging method. The SROC regression analysis suggests that the observed differences in sensitivity and specificity are consistent with the classic tradeoff between sensitivity and specificity that is observed for a continuous diagnostic measurement when operating with different thresholds for declaring a positive test.

In the analysis of covariates, the variables of age, percent with MI history, study quality score, thallium and sestamibi, and percent men all have significance levels ≥.20, suggesting that these factors do not influence accuracy. Variables included in a backward elimination model were publication year, verification bias, 50% stenosis definition, and proportion of patients with CAD. Of these, only 50% stenosis for disease definition (β -0.648, P = .003) and percent of subjects with CAD (β 1.691, P = .014) showed significance.

Because of concerns regarding verification bias, we compared sensitivities and specificities between studies that did not have verification bias and studies that did have verification bias. There was no significant difference in the results between studies that controlled for verification bias and studies that did not, for any test (results not shown).

**Discussion**

In this meta-analysis, SPECT studies offered greater sensitivity and echocardiographic studies offered greater specificity. Clinicians can choose the appropriate test for their patients depending on the suspicion of coronary disease.

As expected, tests that use adenosine and dipyridamole yield interchangeable results. Both agents are coronary vasodilators; dipyridamole works by increasing intra-arterial levels of adenosine. Both agents cause nondiseased coronary arteries to undergo greater vasodilatation, leading to a “steal” of blood flow away from myocardium perfused by diseased coronary arteries. The selection of either agent is generally based on considerations such as side effects, cost, availability, and familiarity with the agent.

When echocardiographic and SPECT vasodilator studies are compared, we see that adenosine and dipyridamole SPECT studies offer the highest sensitivities,
90% and 89%, respectively. Adenosine and dipyridamole echocardiographic studies offer the highest specificities, 91% and 93%, respectively. Thus the choice of the imaging modality for vasodilator studies seems to involve a trade-off of higher sensitivity or specificity.

The results with dobutamine are significantly different than the results with the vasodilators. Dobutamine is an inotropic agent; instead of primarily affecting myocardial blood supply, dobutamine primarily increases myocardial demand by increasing heart rate and contractility. Thus, as a stressor, it is more comparable to exercise. Not surprisingly, dobutamine echocardiography and nuclear imaging results are comparable with the results from a previous study of exercise echocardiography and SPECT imaging.¹ The sensitivities of dobutamine echocardiography (80%) and dobutamine SPECT imaging (82%) are similar, as are the specificities of exercise echocardiography (85%) and exercise SPECT imaging (87%). Similarly, the specificity of dobutamine echocardiography (84%) is higher than that of dobutamine SPECT imaging (75%), and the specificity of exercise echocardiography (77%) is higher than that exercise SPECT imaging (64%). This suggests that echocardiography may be a better myocardial imaging modality for both dobutamine and exercise studies. Exercise studies offer a slightly higher sensitivity and a lower specificity than dobutamine studies. Of course, only exercise studies offer functional data.

The results in the sex subanalysis are limited by the fact that few studies presented sex-specific data. However, dobutamine echocardiography had a similar sensitivity and specificity in women and men, a finding that had also been confirmed in exercise echocardiographic studies.²

The finding of imaging abnormalities in >2 vascular territories appears to be a specific although not sensitive finding for the detection of multivessel disease. In clinical practice, the degree of abnormality on the imaging study is often as important as whether a study is “positive” or “negative.” Our data show that extensive imaging abnormality, if present, is a reliable sign of more extensive underlying disease, especially in nuclear imaging studies.

Several limitations of our study must be considered. First, like most meta-analyses, our study is subject to publication bias. Only published studies were examined, mostly from academic centers expert in the techniques. A previous study has demonstrated that high-volume echocardiographers had higher accuracy than “beginners” did.¹¹³ Also, studies with poor results are less likely to be accepted for publication. Therefore our results may be better than those achieved in actual practice. Second, as in all meta-analyses of diagnostic testing, verification bias is an important limitation of our study, given that about half the studies included in our meta-analysis did not control for this. Verification bias occurs when the result of the test influences which patients will receive the verification test. This can have dramatic results on the sensitivity and specificity of a test.¹¹⁹ We are unable to correct for this bias because the original studies do not provide information on the entire population tested. However, in our covariate analyses, we failed to find a significant difference between studies that did and did not control for verification bias. Also, this bias is likely to similarly occur in studies of adenosine, dipyridamole, and dobutamine, allowing comparisons between types of testing. Studies varied widely in publication year, so it is possible that the accuracy of older tests might falsely appear to be lower, but interestingly, there was no increase in accuracy in more recent studies in the SROC analysis.

The findings of our study can be used to guide the selection of the optimal pharmacologic stress test for each patient. Maximum specificity can be attained by use of a vasodilator with echocardiography. Maximum sensitivity can be attained by use of a vasodilator combined with nuclear imaging. Dobutamine echocardiography offers a good compromise between sensitivity and specificity. The clinician can customize the test selection to the clinical situation.

We thank Grace Chen for assistance with obtaining references.

References


