Chapter 2. Systematic Review of the Accuracy of Exercise Myocardial Perfusion Imaging and Echocardiography for Diagnosis of Coronary Heart Disease in Women

Introduction

Multiple studies suggest that the accuracy of diagnostic testing for coronary heart disease (CHD) may be different in women compared to men.[1-6] Many factors may account for a differential accuracy, including differences in the pre-test probability of disease, chest wall anatomy, left ventricular chamber size, ability to exercise maximally, catecholamine response to exercise or hormone levels.

One systematic review of the studies of the diagnostic accuracy of exercise electrocardiogram (ECG), exercise thallium and exercise echocardiogram in women included literature published up to 1995. The review examined five myocardial perfusion imaging (MPI) studies that included 842 women and three echocardiography studies that included 296 women.[7] MPI studies all used thallium as the radionuclide; two studies used planar imaging and three used single photon emission computed tomography (SPECT). Weighted mean sensitivity and specificity for exercise ECG in women were 61 and 70 percent; for exercise MPI 78 and 64 percent; and for exercise echocardiography 86 and 79 percent. The findings suggested that exercise stress testing without imaging has limited accuracy in women and that planar MPI is more specific than SPECT. Exercise echocardiography appeared to be the most accurate test, but data were available from only three studies. This systematic review is now outdated and provides little information on the accuracy of currently used MPI techniques that almost universally employ SPECT with technetium or technetium plus thallium imaging.

Another systematic review examined the accuracy of exercise echocardiography and exercise SPECT imaging in men and women based on literature published up to 1997.[8] Weighted mean sensitivity and specificity for exercise MPI were 87 and 64 percent and for exercise echocardiography 85 and 77 percent. The authors concluded that exercise echocardiography and exercise SPECT have similar sensitivities for the detection of coronary artery disease, but exercise echocardiography has slightly higher specificity. The total number of subjects in this study was 5,436; 70 percent were men and separate estimates for accuracy in women were not provided.

The purpose of this systematic review is to evaluate the accuracy of exercise echocardiography and MPI in women, to determine if there are differences in accuracy of these tests in men and women, and to assess test characteristics of exercise MPI with thallium compared to technitium sestamibi imaging.

Methodology

Data sources

We searched PubMed®, the Cochrane Database, and DARE for articles in English and other languages published from 1990 through January 2002. We also reviewed bibliographies and asked peer reviewers (Appendix A) to identify additional articles. The date limits of the search were chosen because both exercise echocardiography and exercise MPI using SPECT with thallium and sestamibi were in widespread use during this period.

Search Terms

We used the following search terms to identify cross-sectional studies in which the accuracy of the diagnostic tests of interest were compared to angiographic findings:

Limits	publication dates 1990 to January 2002, human *Not:* practice guideline, letter, editorial, review, meta-analysis Infant newborn, infant, preschool child, child
Predictor 1:	thallium radioisotopes, radiopharmaceuticals, tomography emission-computed single-photon, technetium TC 99M sestamibi, organotechnetium compounds, Spect, Cardiolite, Mibi *AND* exercise, exercise test, exercise tolerance, exercise*, exercising, "stress test" *AND* diagnosis, diagnoses, diagnostic, diagnosing, predictive values of test
Predictor 2:	echocardio*, ultrasound, ultrasonography *AND* exercise, exercise test, exercise tolerance, exercise*, exercising, "stress test" *AND* diagnosis, diagnoses, diagnostic, diagnosing, predictive values of test Note -- all of the commas represent "OR" statements.
Outcomes	cardiovascular diseases, heart diseases, myocardial ischemia, coronary disease

Inclusion Criteria

To be included, articles were required to fit the following criteria:

1) Contained primary data on at least 10 women who underwent exercise ECG with radionuclide injection and SPECT imaging or exercise echocardiography.
2) Estimated accuracy of noninvasive tests using angiographic evidence of CHD as the gold standard.
3) Provided data to calculate true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) for the noninvasive tests.
4) Clear definition of positive noninvasive test and positive angiogram provided.
5) Published between 1990 and January 2002. Articles published outside this date range that were recommended by peer reviewers were included.

We excluded studies that met the following criteria:

1) Noninvasive tests performed exclusively in patients after myocardial infarction (MI), percutaneous angioplasty, coronary artery bypass surgery or hospitalization for an unstable coronary syndrome. In these patients, noninvasive tests are done for the purpose of risk assessment rather than diagnosis.
2) Tests in which pharmacologic agents rather than exercise were used as the stressor. Use of pharmacologic stressors may significantly affect the accuracy of noninvasive testing; many different agents are used and protocols for their use vary substantially.

Article Identification

An initial search using the terms listed above identified articles that potentially provided evidence. Two University of California, San Francisco (UCSF)-Stanford Evidence-based Practice Center (EPC) investigators reviewed the titles and excluded those that clearly did not provide data on humans or clearly did not address the question.

The abstracts of the remaining articles were reviewed independently by two UCSF-Stanford EPC physician investigators and coded using the categories listed below. Disagreements were discussed and consensus codes were entered into a database (Access, Microsoft Corporation).

T	Test – the study clearly does not include data on exercise ECG with imaging or exercise echocardiography.
A	Angiogram - the study clearly does not compare the results of the noninvasive test with the results of angiography.
ND	Not diagnostic - The study assesses noninvasive tests performed exclusively in patients after myocardial infarction, percutaneous angioplasty, coronary artery bypass surgery or hospitalization for an unstable coronary syndrome.
R	Review – the study is a review that does not contain primary data.

NH	No humans - the study clearly does not include data on humans.
E_1	Eligible – the study may contain primary evidence regarding the research questions in women and will be reviewed in full-text.

Articles coded E_1 were retrieved and the full text was reviewed independently by two UCSF-Stanford EPC physician investigators. Names of authors and titles of journals were obscured before articles were reviewed.

Obtaining Unpublished Results in Women

Some eligible studies included women in the study population, but did not report findings separately by gender. In these instances we attempted to contact authors of these studies to obtain estimates in women. If we did not receive a response after the first contact, a second attempt was made. We contacted 34 authors[2, 9-41] and received data from five.[11, 15, 17, 18, 24]

Quality Assessment

The full text of each eligible study was reviewed independently by two UCSF-Stanford EPC physician investigators who completed a quality evaluation form (Appendix B). The studies included in this systematic review are cross-sectional. The three major quality issues affecting these studies are verification bias, biased outcome measurement and spectrum effect. Verification bias occurs when the decision to proceed to the gold standard is in part dependent on the results of the noninvasive test. Since positive noninvasive test results are more likely to be followed by an invasive test, this tends to increase the chance of detecting a true positive (TP) relative to a false negative (FN) and tends to increase the chance of detecting a false positive (FP) relative to a true negative (TN). Therefore, sensitivity may appear to be higher and specificity lower in the verified sample. Biased outcome measurement occurs when personnel performing or reading the results of the noninvasive test already know the results of angiography. Spectrum effect refers to the variation in test performance depending on the severity of disease in the population studied. Sensitivity and specificity appear higher when the persons studied either have severe disease or are healthy. For instance, in participants with significant coronary disease and healthy volunteers, the spectrum of disease is clear-cut, and both sensitivity and specificity will be higher compared to a population with intermediate prior probability of coronary disease, such as those with angina. Our quality assessment addresses verification bias and biased outcome measurement, and we recorded spectrum of disease to allow subgroup analyses.

To be categorized as *good* quality, articles were required to meet the following parameters:

- All participants who had the noninvasive test also had angiography.
- The diagnosis of coronary artery disease on angiography was made by investigators blinded to the results of the noninvasive test

Studies that did not meet these criteria were considered *fair* quality.

Data Abstraction

Two UCSF-Stanford EPC physician investigators independently reviewed the full text of each eligible study and completed a data abstraction form (Appendix C). Data abstracted included characteristics of the study (design, inclusion and exclusion criteria, noninvasive tests performed, and setting), participant characteristics (number of women and men, mean age of participants, number with prior MI, number with revascularization, cardiac risk factors in the population, and indications for cardiac testing), and test characteristics (type of exercise, average duration of exercise, percent with adequate exercise, radionuclide and imaging protocols used, criteria for positive noninvasive test, and criteria for positive coronary angiogram). For each eligible study, the numbers of true positive, true negative, false positive and false negative tests were recorded or calculated as necessary. We also abstracted accuracy measures for all subgroups evaluated. Disagreements between abstractors were discussed and decided by consensus. For studies with multiple publications, only data from the most comprehensive or recent publication were used.

Data Management and Archive

We entered all identified titles and abstracts in an EndNote® file (Niles Software, Inc.) that includes searchable key words as codes for eligibility. Information on all articles that were reviewed in full text was transferred from EndNote® to a database (Access, Microsoft® Corporation) that allows us to categorize each article by reason for exclusion. Quality assessment data for each eligible study were also entered in the database, allowing us to categorize eligible articles by quality.

Abstracted data were entered into a database (EXCEL, Microsoft® Corporation) for preparation of evidence tables and calculation of summary estimates, confidence intervals and tests of heterogeneity.

The full-text articles that were retrieved, and the abstraction forms for each article are filed in Dr. Grady's offices at the UCSF Mt. Zion Women's Health Clinical Research Center.

Data Analyses

The primary outcomes of each study were expressed as sensitivity, specificity, positive likelihood ratio and negative likelihood ratio comparing the results of the noninvasive test to angiographic findings. Summary results were calculated as the mean of the appropriate proportion (sensitivity, specificity, likelihood ratios) weighted by the sample size of each individual study. The significance level for all p-values for the weighted means was set at 0.05. All findings were assessed for heterogeneity using Z-tests. The significance level for tests of heterogeneity was 0.10. To avoid calculation problems associated with zero cells, 0.5 was added to all cells to calculate variances and standard deviations.[42] Results for women vs. men were compared using the Q* statistic, the point on the summary ROC curve where sensitivity equals specificity.[43]

Publication bias usually occurs if small studies with unremarkable findings (poor accuracy) are not published while small studies with markedly positive findings (high accuracy) are published. We calculated the correlation between individual study sample size and sensitivity using Kendall's Tau to assess potential publication bias.

Results

Study Identification

Our searches identified 3,136 titles. After eliminating ineligible studies by review of titles and abstracts, we reviewed the full text of 326 articles and found 14 eligible for inclusion in the systematic review.[2, 6, 17, 18, 23, 24, 41, 44-50] Of the 10 studies included that examined exercise MPI as the noninvasive test, five used thallium,[17, 41, 44-46] four used sestamibi[18, 23, 47, 48] and one used both[6] as radionuclide agents.

Nine MPI studies that provided accuracy estimates in women were excluded from the systematic review: four used pharmacologic agents in addition to exercise;[51-54] four reported some measures of accuracy, but did not report adequate data to allow calculation of all required estimates,[5, 55-57] and one did not provide definitions of a positive noninvasive test or positive coronary angiogram.[58]

Four eligible studies examined the accuracy of exercise echocardiography as the noninvasive test.[2, 24, 49, 50] One study of exercise echocardiography that provided accuracy estimates for women was excluded because it was published outside our date range.[59]

Description of Eligible Studies

The characteristics of the 10 studies assessing the accuracy of MPI that were included in the systematic review are shown in Evidence Table 1. The number of participants in each

study ranged from 41 to 371 and 14 to 100 percent were women. The total number of participants was 1,249, including 549 women and 700 men. Three of the studies included only women. One study included both men and women, but provided data that allowed calculation of accuracy estimates only in women.[48] The mean age of participants ranged from 51 to 62 years. Eight of the studies included participants with prior MI. All 10 studies used SPECT imaging; five used thallium only, four sestamibi only and one used both radionuclides. The definition of an abnormal test was very similar in nine of the studies (fixed or reversible perfusion defects, perfusion defects at rest or after exercise or decreased uptake at rest or after exercise). One study defined a positive test as reversible uptake defects in more than one of 22 coronary segments. All but one of the 10 studies defined 50 percent stenosis of one or more major coronary artery at angiography as the gold standard for the presence of coronary artery disease. Eight studies used treadmill exercise and two used bicycle exercise. Six of the MPI studies were judged fair quality and four were judged good quality.

The characteristics of the four studies assessing the accuracy of echocardiography that were included in the systematic review are shown in Evidence Table 2. The number of participants in each study ranged from 70 to 340 and 15 to 100 percent were women. The total number of participants was 689, including 344 women and 345 men. Two of the studies included both men and women and 2 included only women.[2, 49] The mean age of participants ranged from 55 to 66 years. One of the studies included participants with prior MI. Two of the studies defined an abnormal test as new or worse regional wall motion abnormalities after exercise and two defined a positive as regional wall motion abnormalities at rest or after exercise. All studies defined 50 percent stenosis of at least one major coronary artery at angiography as the gold standard for the presence of coronary artery disease. Two of the echocardiography studies were judged fair quality and two were judged good quality.

Findings

Exercise Myocardial Perfusion Imaging

In women, sensitivity of exercise MPI using either thallium or sestamibi ranged from 0.61 to 1.0 with a mean weighted sensitivity of 0.77 (95% CI 0.72-0.83) (Evidence Table 3). Specificity of exercise MPI in women ranged from 0.40-1.0, with a mean weighted specificity of 0.69 (95% CI 0.62-0.75). The mean weighted positive likelihood ratio for exercise MPI in women was 2.46 (95% CI 2.00-3.04) and the mean weighted negative likelihood ratio was 0.33 (95% CI 0.26-0.41).

Based on the findings of the six studies that included men, the mean weighted sensitivity of exercise MPI in men was 0.93 (95% CI 0.90-0.95) (Evidence Table 3) and the mean weighted specificity was 0.57 (95% CI 0.47-0.67). The mean weighted positive likelihood ratio for exercise MPI in men was 2.17 (95% CI 1.73-2.73) and the mean weighted negative likelihood ratio was 0.13 (95% CI 0.09-0.19).

We performed a subgroup analysis limited to the findings of the four good quality studies (Evidence Table 1). Mean weighted accuracy estimates from these analyses were not materially different from the overall mean summary estimates (Evidence Table 3).

We performed two sensitivity analyses. One eligible study evaluated the accuracy of both thallium and sestamibi.[6] The overall mean weighted results included only the accuracy estimates for sestamibi. A sensitivity analysis substituting the results for thallium produced similar overall results. We also repeated the analysis for women including the findings of one study that was excluded because no definitions of an abnormal test or abnormal angiogram were provided.[58] Including the results of this study did not materially change the accuracy estimates.

We calculated mean weighted accuracy estimates for men and women from studies that used sestamibi separately from those that used thallium (Evidence Table 4). Accuracy estimates in women for studies using sestamibi and those using thallium were very similar (p-value for the comparison of Q* statistics = 0.84)

Exercise Echocardiography

In women, sensitivity for exercise echocardiography ranged from 0.77 to 0.88 with a mean weighted sensitivity of 0.81 (95% CI 0.74-0.87) (Evidence Table 5). Specificity for exercise echocardiography in women ranged from 0.37 to 0.84, with a mean weighted specificity of 0.73 (95% CI 0.66-0.79). The mean weighted positive likelihood ratio for exercise echocardiography in women was 2.95 (95% CI 2.28-3.79) and the mean weighted negative likelihood ratio was 0.26 (95% CI 0.19-0.36).

Based on the findings of two studies of the accuracy of echocardiography that included men, the mean weighted sensitivity was 0.84 (95% CI 0.79-0.88) and the mean weighted specificity was 0.45 (95% CI 0.32-0.59) (Evidence Table 5). The mean weighted positive likelihood ratio for exercise echocardiography in men was 1.54 (95% CI 1.20-1.98) and the mean weighted negative likelihood ratio was 0.36 (95% CI 0.24-0.53).

We performed subgroup analyses limited to the findings of the two good quality studies of the accuracy of exercise echocardiography in women (Evidence Table 1). The mean weighted sensitivity estimate from the good quality studies was similar to the overall mean sensitivity (0.82 vs. 0.81), but specificity was lower (0.60 vs. 0.73) (Evidence Table 5). These differences resulted in a lower mean weighted positive likelihood ratio based on the good quality studies compared to the estimate based on all eligible studies (2.06 vs. 2.95). Mean estimates of negative likelihood ratios did not differ when results were restricted to the good quality studies (Evidence Table 5).

We performed a sensitivity analysis by adding the results of one study of the accuracy of exercise echocardiography that was published before our date range.[59] Including the results of this study did not materially change the overall accuracy estimates.

Accuracy of Exercise MPI vs. Echocardiography in Women

Based on 10 studies of the accuracy of exercise MPI and four of exercise echocardiography, the accuracy of each test for the diagnosis of CHD in women is similar (Evidence Tables 3 and 5). The mean weighted sensitivity, specificity and positive likelihood ratio for MPI are slightly lower than for echocardiography (sensitivity 0.77 vs. 0.81; specificity 0.69 vs. 0.73, positive likelihood ratio 2.46 vs. 2.95), but the differences are small and not statistically different (p-value for the Q* statistic = 0.10). The accuracy of the two tests is also similar when analyses are restricted to good quality studies (Evidence Tables 3 and 5). The use of sestamibi instead of thallium also did not change the accuracy of MPI studies in women (Evidence Table 4).

Accuracy of Noninvasive Testing in Women Compared to Men

The mean weighted sensitivity of MPI in women is somewhat lower than in men (0.77 vs. 0.93), but the specificity is higher (0.69 vs. 0.57) (Evidence Table 3). The positive likelihood ratio is slightly higher in women compared to men (2.46 vs. 2.17) as is the negative likelihood ratio (0.33 vs. 0.13). These differences in the accuracy of MPI between men and women were statistically significant (p-value for the comparison of Q* statistics 0.028), but it is not clear whether higher sensitivity or higher specificity is preferable.

Comparison of mean weighted accuracy estimates between women and men may be biased if these data are derived from different studies that may have used somewhat different methods and definitions of positive tests. To avoid this problem, we calculated mean weighted accuracy estimates for men and women restricted to the findings of studies that included both genders (Evidence Table 3). Based on the findings of these studies, sensitivity of MPI is lower in women than in men (0.86 vs. 0.93), but specificity is the same (0.57 for both genders; p-value for the comparison of Q* statistics 0.012). This analysis suggests that exercise MPI is more accurate in men than in women, but the differences are small and not clinically meaningful.

The mean weighted sensitivity of echocardiography in women is similar to that in men (0.81 vs. 0.84), but the specificity is higher (0.73 vs. 0.45) (Evidence Table 5). The positive likelihood ratio was substantially higher in women than in men (2.95 vs. 1.54), and the negative likelihood ratio was slightly lower (0.26 vs. 0.36). However, given the small numbers of men included in the analyses, we could not calculate a Q* statistic or determine any statistically significant differences between men and women with regard to exercise echocardiography.

Assessments for Heterogeneity and Publication Bias

There was no heterogeneity in any of the mean weighted estimates of accuracy. Publication bias usually occurs if small studies with unremarkable findings (poor accuracy) are not published while small studies with markedly positive findings (high

accuracy) are published. We calculated the correlation between individual study sample size and sensitivity using Kendall's Tau to assess potential publication bias. There was no evidence of publication bias in any of the summary estimates of accuracy.

Conclusions

In the last decade, both exercise echocardiography and exercise MPI have become widely available and commonly used for noninvasive diagnosis of coronary disease. It is important for both patients and providers to understand the accuracy of these tests and their limitations. We obtained results from 14 studies published between 1990 and 2002 on the accuracy of these tests in women. Based on these data, the overall accuracy of both tests in women is low with positive likelihood ratios of 2.5 to 3 and negative likelihood ratios of about 0.3.

There are several advantages of estimating accuracy of a diagnostic test using likelihood ratios rather than sensitivity and specificity. First, it is possible to achieve a high sensitivity for most diagnostic tests by accepting a low specificity; similarly, high specificity can be achieved by accepting low sensitivity. In contrast, both sensitivity and specificity must be high to achieve good likelihood ratios (positive LR = sensitivity/(1-specificity) and negative LR = (1-sensitivity)/specificity). Secondly, likelihood ratios are a powerful tool to apply clinically using Bayes' theorem; the post-test odds that a patient has the disease are estimated by multiplying the pre-test odds by the positive likelihood ratio. For instance, in a 55 year old woman with probable angina, the prior probability of CHD is about 30 percent.[60] If her exercise MPI is positive, her posterior probability of CHD would be about 50 percent (prior odds 1:2.3 multiplied by positive LR of 2.5 equal posterior odds of 2.5:2.3 which is equivalent to posterior probability of about 50 percent). Similarly, if her exercise echocardiogram is positive, her posterior probability of having CHD would be about 55 percent. If either of these studies were negative, her posterior probability would be about 10 percent. Small differences in the posterior probabilities based on exercise MPI or echocardiogram do not have different clinical implications and suggest that the value of these tests is equivalent.

The common conception that exercise testing in women should always be combined with imaging may not be true. A prior meta-analysis that evaluated the accuracy of exercise EKG in women found a mean weighted positive likelihood ratio of 2.25 and a negative likelihood ratio of 0.55.[7] These accuracy estimates are very similar to those that we calculated for exercise MPI and echocardiogram and would result in very similar estimates of posterior probability of CHD. However, women who receive exercise EKG testing without imaging are more likely to have a normal EKG at baseline and thus may be less likely to have significant CHD. Thus, comparison of the accuracy of exercise EKG with exercise imaging studies or exercise echocardgiography may be biased unless patients are randomized to receive the different tests.

The value of a diagnostic test result depends on the accuracy of the test, the prior probability of disease and the threshold for treatment. In women with low to intermediate prior probability of CHD, a positive exercise MPI or echocardiogram result in similar posterior probabilities that may warrant further testing. Our 55 year-old woman with angina, for example, has about a 50 percent probability of having CHD if she has a positive noninvasive test. Before labeling her as having CHD and beginning treatment, many clinicians may want to pursue angiography. An older woman in her mid sixties with angina has about a 50 percent prior probability of having CHD. If she has a negative noninvasive test, her posterior probability of having CHD is about 25 percent. Many clinicians may prefer a more accurate test (such as angiography) before declaring that this woman does not have CHD and forgoing treatment.

Our analysis found that the sensitivity of exercise MPI in women was lower than in men. While this difference was statistically significant, it was small and not clinically meaningful. Most of the studies included in our review reported a higher prevalence of CHD in men than in women. The prevalence of prior MI also was higher in men compared to women. This spectrum effect could account for the apparent lower sensitivity of MPI in women compared to men. Alternatively, the lower sensitivity in women may be due to differences in chest wall anatomy, left ventricular chamber size, ability to exercise maximally, catecholamine response to exercise or hormone levels. Comparison of the accuracy of exercise echocardiography in women and men was limited by the small number of studies that reported separate data for men.

Evaluation of the accuracy of noninvasive tests requires a dichotomous outcome. The conventional “gold standard” for the presence of CHD is 50 percent or more stenosis of one or more of the major coronary arteries at angiography, and this is the definition that we used. However, coronary heart disease represents a continuum of disease that may not be best measured by angiography.

In addition to diagnosis of CHD, exercise MPI and exercise echocardiography are also performed to localize disease and to determine the extent of disease. Our analysis provides no data on the accuracy of noninvasive tests for these purposes.

The value of our review is limited by the quality of the studies included and the number of persons included. Many of the studies included in this systematic review were rated only fair quality, often because it was not clear that personnel who interpreted the noninvasive test were blinded to the results of the angiogram. Prior knowledge of angiographic results could falsely increase the accuracy of the noninvasive test. In many of the studies included, only persons with a positive noninvasive test went on to have angiography. This verification bias may result in higher sensitivity and lower specificity than in studies in which all subjects undergo both tests.

Unfortunately, most studies did not report the percent of maximum predicted heart rate achieved by subjects. Failure to achieve 85 percent of maximum predicted heart rate would likely result in a higher number of false negatives and lower sensitivity. Sensitivity might be lower in women if they are less likely than men to exercise adequately.

Most of the studies of MPI and one of the studies of echocardiography included a substantial proportion of persons with prior myocardial infarction. Including persons with diagnosed disease likely increases sensitivity at the cost of specificity. Finally, we identified only 14 studies that met our inclusion criteria and the number of women included was limited. We identified many additional studies that included some women, but were unable to obtain data stratified by gender from the authors.

Future Research

Studies of the accuracy of noninvasive tests should publish all estimates of accuracy stratified by sex, or make these estimates available to public access. Stratification by sex would allow more precise estimates of the accuracy of noninvasive tests for CHD in women, but it is unlikely that this would result in a substantial improvement in the estimated accuracy of the tests. The major finding of this systematic review is that the accuracy of noninvasive tests for CHD in both men and women is low and that future research should address ways to improve accuracy. The quality of future studies of the accuracy of noninvasive tests for the diagnosis of CHD would be improved by excluding persons with known CHD, performing both the noninvasive test and angiography in all participants and assuring that the outcome of the noninvasive test is assessed by personnel blinded to the results of angiography. Finally, assessment of the value of diagnostic tests for estimating risk of future CHD events would have important long-term clinical implications.

References

1. Barolsky SM, Gilbert CA, Faruqui A, et al. Differences in electrocardiographic response to exercise of women and men: a non-Bayesian factor. Circulation 1979;60(5):1021-7.
2. Marwick TH, Torelli J, Harjai K, et al. Influence of left ventricular hypertrophy on detection of coronary artery disease using exercise echocardiography. J Am Coll Cardiol 1995;26(5):1180-6.
3. Sketch MH, Mohiuddin SM, Lynch JD, et al. Significant sex differences in the correlation of electrocardiographic exercise testing and coronary arteriograms. Am J Cardiol 1975;36(2):169-73.
4. Weiner DA, Ryan TJ, McCabe CH, et al. Exercise stress testing. Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the Coronary Artery Surgery Study (CASS). N Engl J Med 1979;301(5):230-5.
5. Hansen CL, Crabbe D, Rubin S. Lower diagnostic accuracy of thallium-201 SPECT myocardial perfusion imaging in women: an effect of smaller chamber size. J Am Coll Cardiol 1996;28(5):1214-9.
6. Taillefer R, DePuey EG, Udelson JE, et al. Comparative diagnostic accuracy of Tl-201 and Tc-99m sestamibi SPECT imaging (perfusion and ECG-gated SPECT) in detecting coronary artery disease in women. J Am Coll Cardiol 1997;29(1):69-77.
7. Kwok Y, Kim C, Grady D, et al. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999;83(5):660-6.
8. Fleischmann KE, Hunink MG, Kuntz KM, et al. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA 1998;280(10):913-20.
9. Beleslin BD, Ostojic M, Stepanovic J, et al. Stress echocardiography in the detection of myocardial ischemia. Head-to-head comparison of exercise, dobutamine, and dipyridamole tests. Circulation 1994;90(3):1168-76.
10. Berman DS, Kiat H, Friedman JD, et al. Separate acquisition rest thallium-201/stress technetium-99m sestamibi dual-isotope myocardial perfusion single-photon emission computed tomography: a clinical validation study. J Am Coll Cardiol 1993;22(5):1455-64.
11. Bjornstad K, Aakhus S, Hatle L. Comparison of digital dipyridamole stress echocardiography and upright bicycle stress echocardiography for identification of coronary artery stenosis. Cardiology 1995;86(6):514-20.
12. Christian TF, Miller TD, Bailey KR, et al. Noninvasive identification of severe coronary artery disease using exercise tomographic thallium-201 imaging. Am J Cardiol 1992;70(1):14-20.
13. Crouse LJ, Harbrecht JJ, Vacek JL, et al. Exercise echocardiography as a screening test for coronary artery disease and correlation with coronary arteriography. Am J Cardiol 1991;67(15):1213-8.

14. Dagianti A, Penco M, Agati L, et al. Stress echocardiography: comparison of exercise, dipyridamole and dobutamine in detecting and predicting the extent of coronary artery disease. J Am Coll Cardiol 1995;26(1):18-25.
15. Fleming RM, Kirkeeide RL, Taegtmeyer H, et al. Comparison of technetium-99m teboroxime tomography with automated quantitative coronary arteriography and thallium-201 tomographic imaging. J Am Coll Cardiol 1991;17(6):1297-302.
16. Galanti G, Sciagra R, Comeglio M, et al. Diagnostic accuracy of peak exercise echocardiography in coronary artery disease: comparison with thallium-201 myocardial scintigraphy. Am Heart J 1991;122(6):1609-16.
17. Gupta NC, Esterbrooks DJ, Hilleman DE, et al. Comparison of adenosine and exercise thallium-201 single-photon emission computed tomography (SPECT) myocardial perfusion imaging. The GE SPECT Multicenter Adenosine Study Group. J Am Coll Cardiol 1992;19(2):248-57.
18. Hambye AS, Vervaet A, Lieber S, et al. Diagnostic value and incremental contribution of bicycle exercise, first-pass radionuclide angiography, and 99mTc-labeled sestamibi single- photon emission computed tomography in the identification of coronary artery disease in patients without infarction. J Nucl Cardiol 1996;3(6 Pt 1):464-74.
19. Hecht HS, DeBord L, Sotomayor N, et al. Supine bicycle stress echocardiography: peak exercise imaging is superior to postexercise imaging. J Am Soc Echocardiogr 1993;6(3 Pt 1):265-71.
20. Heiba SI, Hayat NJ, Salman HS, et al. Technetium-99m-MIBI myocardial SPECT: supine versus right lateral imaging and comparison with coronary arteriography. J Nucl Med 1997;38(10):1510-4.
21. Ho YL, Wu CC, Huang PJ, et al. Dobutamine stress echocardiography compared with exercise thallium-201 single-photon emission computed tomography in detecting coronary artery disease-effect of exercise level on accuracy. Cardiology 1997;88(4):379-85.
22. Hoffmann R, Lethen H, Kleinhans E, et al. Comparative evaluation of bicycle and dobutamine stress echocardiography with perfusion scintigraphy and bicycle electrocardiogram for identification of coronary artery disease. Am J Cardiol 1993;72(7):555-9.
23. Kiat H, Van Train KF, Maddahi J, et al. Development and prospective application of quantitative 2-day stress-rest Tc-99m methoxy isobutyl isonitrile SPECT for the diagnosis of coronary artery disease. Am Heart J 1990;120(6 Pt 1):1255-66.
24. Luotolahti M, Saraste M, Hartiala J. Exercise echocardiography in the diagnosis of coronary artery disease. Ann Med 1996;28(1):73-7.
25. Mahmarian JJ, Boyce TM, Goldberg RK, et al. Quantitative exercise thallium-201 single photon emission computed tomography for the enhanced diagnosis of ischemic heart disease. J Am Coll Cardiol 1990;15(2):318-29.
26. Marangelli V, Iliceto S, Piccinni G, et al. Detection of coronary artery disease by digital stress echocardiography: comparison of exercise, transesophageal atrial pacing and dipyridamole echocardiography. J Am Coll Cardiol 1994;24(1):117-24.

27. Marwick TH, D'Hondt AM, Mairesse GH, et al. Comparative ability of dobutamine and exercise stress in inducing myocardial ischaemia in active patients. Br Heart J 1994;72(1):31-8.
28. Minoves M, Garcia A, Magrina J, et al. Evaluation of myocardial perfusion defects by means of "bull's eye" images. Clin Cardiol 1993;16(1):16-22.
29. Nguyen T, Heo J, Ogilby JD, et al. Single photon emission computed tomography with thallium-201 during adenosine-induced coronary hyperemia: correlation with coronary arteriography, exercise thallium imaging and two-dimensional echocardiography. J Am Coll Cardiol 1990;16(6):1375-83.
30. Ozdemir K, Kisacik HL, Oguzhan A, et al. Comparison of exercise stress testing with dobutamine stress echocardiography and radionuclide ventriculography for diagnosis of coronary artery disease. Jpn Heart J 1999;40(6):715-27.
31. Palmas W, Friedman JD, Diamond GA, et al. Incremental value of simultaneous assessment of myocardial function and perfusion with technetium-99m sestamibi for prediction of extent of coronary artery disease. J Am Coll Cardiol 1995;25(5):1024-31.
32. Quinones MA, Verani MS, Haichin RM, et al. Exercise echocardiography versus 201Tl single-photon emission computed tomography in evaluation of coronary artery disease. Analysis of 292 patients. Circulation 1992;85(3):1026-31.
33. Roger VL, Pellikka PA, Oh JK, et al. Identification of multivessel coronary artery disease by exercise echocardiography. J Am Coll Cardiol 1994;24(1):109-14.
34. Rubello D, Zanco P, Candelpergher G, et al. Usefulness of 99mTc-MIBI stress myocardial SPECT bull's-eye quantification in coronary artery disease. Q J Nucl Med 1995;39(2):111-5.
35. Salustri A, Pozzoli MM, Hermans W, et al. Relationship between exercise echocardiography and perfusion single-photon emission computed tomography in patients with single-vessel coronary artery disease. Am Heart J 1992;124(1):75-83.
36. Solot G, Hermans J, Merlo P, et al. Correlation of 99Tcm-sestamibi SPECT with coronary angiography in general hospital practice. Nucl Med Commun 1993;14(1):23-9.
37. Sylven C, Hagerman I, Ylen M, et al. Variance ECG detection of coronary artery disease--a comparison with exercise stress test and myocardial scintigraphy. Clin Cardiol 1994;17(3):132-40.
38. Tawa CB, Baker WB, Kleiman NS, et al. Comparison of adenosine echocardiography, with and without isometric handgrip, to exercise echocardiography in the detection of ischemia in patients with coronary artery disease. J Am Soc Echocardiogr 1996;9(1):33-43.
39. Van Train KF, Areeda J, Garcia EV, et al. Quantitative same-day rest-stress technetium-99m-sestamibi SPECT: definition and validation of stress normal limits and criteria for abnormality. J Nucl Med 1993;34(9):1494-502.
40. Van Train KF, Garcia EV, Maddahi J, et al. Multicenter trial validation for quantitative analysis of same-day rest-stress technetium-99m-sestamibi myocardial tomograms. J Nucl Med 1994;35(4):609-18.
41. Van Train KF, Maddahi J, Berman DS, et al. Quantitative analysis of tomographic stress thallium-201 myocardial scintigrams: a multicenter trial. J Nucl Med 1990;31(7):1168-79.

42. Haldane J. The estimation and significance of the logarithm of a ratio of frequencies. Ann Hum Genet 1955;20:309-314.
43. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12(14):1293-316.
44. Chae SC, Heo J, Iskandrian AS, et al. Identification of extensive coronary artery disease in women by exercise single-photon emission computed tomographic (SPECT) thallium imaging. J Am Coll Cardiol 1993;21(6):1305-11.
45. Laurienzo JM, Cannon RO, 3rd, Quyyumi AA, et al. Improved specificity of transesophageal dobutamine stress echocardiography compared to standard tests for evaluation of coronary artery disease in women presenting with chest pain. Am J Cardiol 1997;80(11):1402-7.
46. Astarita C, Nicolai E, Liguori E, et al. [Dipyridamole-echocardiography and thallium exercise myocardial scintigraphy in the diagnosis of obstructive coronary or microvascular disease in hypertensive patients with left ventricular hypertrophy and angina]. G Ital Cardiol 1998;28(9):996-1004.
47. Sciammarella MG, Fragasso G, Gerundini P, et al. 99Tcm-MIBI single photon emission tomography (SPET) for detecting myocardial ischaemia and necrosis in patients with significant coronary artery disease. Nucl Med Commun 1992;13(12):871-8.
48. Mak KH, Ang ES, Goh AS, et al. Myocardial perfusion imaging with technetium-99m sestamibi SPECT in the evaluation of coronary artery disease. Australas Radiol 1995;39(2):112-7.
49. Williams MJ, Marwick TH, O'Gorman D, et al. Comparison of exercise echocardiography with an exercise score to diagnose coronary artery disease in women. Am J Cardiol 1994;74(5):435-8.
50. Roger VL, Pellikka PA, Bell MR, et al. Sex and test verification bias. Impact on the diagnostic value of exercise echocardiography. Circulation 1997;95(2):405-10.
51. Cecil MP, Kosinski AS, Jones MT, et al. The importance of work-up (verification) bias correction in assessing the accuracy of SPECT thallium-201 testing for the diagnosis of coronary artery disease. J Clin Epidemiol 1996;49(7):735-42.
52. Kang X, Berman DS, Lewin H, et al. Comparative ability of myocardial perfusion single-photon emission computed tomography to detect coronary artery disease in patients with and without diabetes mellitus. Am Heart J 1999;137(5):949-57.
53. Santana-Boado C, Candell-Riera J, Castell-Conesa J, et al. Diagnostic accuracy of technetium-99m-MIBI myocardial SPECT in women and men. J Nucl Med 1998;39(5):751-5.
54. Niemeyer MG, Van der Wall EE, Kuyper AF, et al. Discordance of visual and quantitative analysis regarding false negative and false positive test results in thallium-201 myocardial perfusion scintigraphy. Am J Physiol Imaging 1991;6(1):34-43.
55. Takeuchi M, Sonoda S, Miura Y, et al. Comparative diagnostic value of dobutamine stress echocardiography and stress thallium-201 single-photon-emission computed tomography for detecting coronary artery disease in women. Coron Artery Dis 1996;7(11):831-5.

56. Morise AP, Diamond GA, Detrano R, et al. Incremental value of exercise electrocardiography and thallium-201 testing in men and women for the presence and extent of coronary artery disease. Am Heart J 1995;130(2):267-76.
57. Sharir T, Germano G, Waechter PB, et al. A new algorithm for the quantitation of myocardial perfusion SPECT. II: validation and diagnostic yield. J Nucl Med 2000;41(4):720-7.
58. Iskandrian AE, Heo J, Nallamothu N. Detection of coronary artery disease in women with use of stress single-photon emission computed tomography myocardial perfusion imaging. J Nucl Cardiol 1997;4(4):329-35.
59. Sawada SG, Ryan T, Fineberg NS, et al. Exercise echocardiographic detection of coronary artery disease in women. J Am Coll Cardiol 1989;14(6):1440-7.
60. Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). Circulation 1981;64(2):360-7.