Meta-analysis of exercise testing to detect coronary artery disease in women
Kwok Y, Kim C, Grady D, Segal M, Redberg R

Authors' objectives
To determine the accuracy of the exercise electrocardiogram (ECG), exercise thallium radionuclide scan and exercise echocardiogram (echo) for the diagnosis of coronary artery disease (CAD) in women.

Searching
MEDLINE was searched 1966 to December 1995. The search terms were listed. The reference lists of the included studies and review articles were checked and experts in the field were contacted. Only English language studies were eligible for inclusion.

Study selection
Study designs of evaluations included in the review
Studies of 50 or more women were eligible for inclusion.

Specific interventions included in the review
Studies of exercise ECG, thalium radionuclide exercise scan and exercise echo, alone or in combination, were eligible for inclusion. The exercise protocols included Bruce, bicycle and treadmill. Studies of noninvasive testing using pharmacologic stress were excluded. All of the included studies defined a positive exercise ECG as at least 0.1 mV horizontal or downsloping ST-segment depression measured 80 ms after the J point.

Reference standard test against which the new test was compared
The reference test was coronary angiography.

Participants included in the review
Women undergoing at least one of the exercise tests and coronary angiography. The mean age of women in included studies ranged from 49 to 63 years. The proportion of women with previous myocardial infarction (MI) ranged from 0 to 62%. CAD was defined as either 50% or 70% luminal diameter narrowing. The proportion of women with CAD varied from 18 to 67%. Studies were excluded if they did not present separate results for women. Exercise studies on post-MI risk stratification or post-angioplasty evaluation were excluded.

Outcomes assessed in the review
To be included, the studies had to present data in a manner that allowed the sensitivity and specificity of the tests to be calculated separately for women. Sensitivity, specificity and likelihood ratios (LRs) were presented for each test.

How were decisions on the relevance of primary studies made?
The authors do not state how the papers were selected for the review, or how many of the reviewers performed the selection.

Assessment of study quality
Validity was assessed according to three criteria: adequate description of the study group, the potential for verification bias (i.e. whether the test results influenced the decision to perform angiography) and the potential for diagnostic and test review bias (i.e. whether the exercise test and angiography were read blindly). Studies were classed as high quality if they met all three criteria, medium if they met two and low if they met only one. The authors do not state how the papers were assessed for validity, or how many of the reviewers performed the validity assessment.

Data extraction
Two reviewers extracted data from the included studies independently. One of the reviewers was blinded to the
authors, institutions and title. Any disagreements were resolved by consensus. Data were extracted on: the tests, proportion of women, age, exercise protocol, the proportion with prior MI, the definition of CAD and the proportion with CAD on angiography. The results of each study were recorded in 2x2 tables and the sensitivity, specificity and LRs were calculated for each study.

**Methods of synthesis**

How were the studies combined?

A summary receiver operating characteristic (ROC) curve was calculated for studies of exercise ECG, weighted by the study sample size. A summary ROC curve could not be calculated for exercise thallium or exercise echo studies because there were too few studies in each group.

Weighted (by sample size) mean sensitivities, specificities, LRs and 95% confidence intervals (CIs) were calculated for all of the groups.

How were differences between studies investigated?

For the exercise ECG studies, the summary ROC model was extended to include the following variables: validity; definition of CAD (50 or 70% stenosis); exclusion of men; upsloping ST segment counted as positive test; coronary disease prevalence (less than 40% or greater than 40%); exclusion of previous MI; baseline ECG abnormalities; taking digoxin.

**Results of the review**

Twenty-one studies were included in the meta-analysis. Some studies provided data on more than one test. Nineteen studies provided data on exercise ECGs (n=3,721), five on exercise thallium (n=842) and three on exercise echo (n=296).

The included studies had a broad range of methodological validity.

Exercise ECG: the weighted mean sensitivity was 0.61 (95% CI: 0.54, 0.68), the specificity 0.70 (95% CI: 0.64, 0.75), the positive LR 2.25 (95% CI: 1.84, 2.66) and the negative LR 0.55 (95% CI: 0.47, 0.62).

Exercise thallium: the weighted mean sensitivity was 0.78 (95% CI: 0.72, 0.83), the specificity 0.64 (95% CI: 0.51, 0.77), the positive LR 2.87 (95% CI: 1.0, 4.96) and the negative LR 0.36 (95% CI: 0.27, 0.45).

Exercise echo: the weighted mean sensitivity was 0.86 (95% CI: 0.75, 0.96), the specificity 0.79 (95% CI: 0.72, 0.86), the positive LR 4.29 (95% CI: 2.93, 5.65) and the negative LR 0.18 (95% CI: 0.05, 0.31).

Results of subset analyses: factors that produced a statistically- significant effect on the accuracy of the exercise ECG were whether the ECG was being compared with another test, the exclusion of men from the study and the prevalence of CAD. The differences were reported to be small and not clinically significant. The accuracy of the exercise thallium test was not substantially improved by any factors, with the exception that two studies using planar images had a higher specificity than tomographic thallium. No subset analyses of exercise echo were conducted because of the small number of studies.

**Authors' conclusions**

Currently available exercise tests are only moderately sensitive and specific for the diagnosis of CAD in women.

**CRD commentary**

The inclusion criteria were clearly stated, details on the included studies were presented and a validity assessment was performed. The search was restricted to one electronic database (MEDLINE) and to English language studies only, which may have resulted in relevant studies being missed. The designs of the included studies were not entirely clear and it is unclear whether pooling in a meta-analysis was appropriate given the heterogeneity, although many factors with the potential to influence the results (including methodological validity) were investigated in subset analyses.
authors’ conclusions seem to follow from the results presented, but should be treated with caution due to possible clinical heterogeneity in the pooled studies and the small number of studies in the thallium and echocardiography groups.

**Implications of the review for practice and research**
Practice: The authors did not state any implications for practice.

Research: The authors state that further studies in women using technetium-99m sestamibi are needed. Additional research is required to confirm and improve the accuracy of exercise echo.

**Bibliographic details**

**PubMedID**
10080415

**Other publications of related interest**
This additional published commentary may also be of interest. Burggraf GW. Review: exercise tests to detect CAD in women have moderate sensitivities and specificities. ACP J Club 1999;131:21.

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Coronary Angiography; Coronary Disease /diagnosis /radiography /radionuclide imaging /ultrasonography; Echocardiography; Electrocardiography; Exercise Test; Female; Humans; Likelihood Functions; Middle Aged; Physical Exertion /physiology; Predictive Value of Tests; ROC Curve; Radiopharmaceuticals; Sensitivity and Specificity; Thallium Radioisotopes

**AccessionNumber**
11999000583

**Date bibliographic record published**
30/11/2003

**Date abstract record published**
30/11/2003

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.