Diagnostic accuracy of myocardial perfusion imaging and stress echocardiography for the diagnosis of left main and triple vessel coronary artery disease: a comparative meta-analysis

N Mahajan, L Polavaram, H Vankayala, et al.

Heart 2010 96: 956-966
doi: 10.1136/hrt.2009.182295

Updated information and services can be found at:
http://heart.bmj.com/content/96/12/956.full.html

These include:

References
This article cites 95 articles, 49 of which can be accessed free at:
http://heart.bmj.com/content/96/12/956.full.html#ref-list-1

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To order reprints of this article go to:
http://heart.bmj.com/cgi/reprintform

To subscribe to Heart go to:
http://heart.bmj.com/subscriptions
Diagnostic accuracy of myocardial perfusion imaging and stress echocardiography for the diagnosis of left main and triple vessel coronary artery disease: a comparative meta-analysis

N Mahajan,1 L Polavaram,1 H Vankayala,1 B Ference,2 Y Wang,3 J Ager,3 J Kovach,2 L Afonso2

ABSTRACT

Objectives Compare the diagnostic performance of stress echocardiography (SE) and myocardial perfusion imaging (MPI) for the diagnosis of left main disease (LM) and triple vessel disease (TVD).

Background Limited comparative data on MPI and SE for the detection of LM and TVD (high-risk coronary artery disease) exist in the literature.

Methods Quantitative meta-analysis was performed using studies identified by systematic electronic literature search. Articles were included if they reported data on exercise, dobutamine SE or exercise, adenosine, dipyridamole, thallium201, technetium 99m sestamibi MPI and used coronary angiography as the reference test. Summary receiver-operating characteristic (SROC) curves were constructed for each imaging modality. Additionally, pooled sensitivity, specificity and likelihood ratios were calculated per modality. Meta-regression was performed to adjust for patient and study characteristics.

Results Thirty-two studies met inclusion criteria; 23 (MPI-15; SE-14; Common studies-6) provided sufficient data for analysis. In a SROC model comparing the two imaging modalities, SE was associated with higher area under curve (0.82 (0.03) vs 0.73 (0.02), p=0.01) and index Q* value (0.75 (0.02) vs 0.67 (0.02), p=0.01). Using pooled summary point estimates, SE had higher pooled sensitivity (94% vs 75%, p<0.001) and lower negative likelihood ratio (0.21 vs 0.47, p<0.001) compared to MPI. No evidence of a difference in the pooled specificity (40% vs 48%, p=0.16) and positive likelihood ratio (1.52 vs 1.58, p=0.36) was seen between the two stress modalities. Pooled diagnostic OR on meta-regression (9.78 vs 4.06, p=0.02) remained significantly higher for SE compared to MPI for identification of LM and TVD even after adjustment for study characteristics.

Conclusions Since LM alone or in combination with TVD are categorised as representing potentially life-threatening variants of CAD, a screening test with high sensitivity, low negative likelihood ratio or higher discriminatory capacity would be desirable for risk stratification. In the absence of a direct head-to-head comparison of the diagnostic accuracies of SE and MPI, our findings indicate that SE appears to be the preferred screening modality for high-risk coronary artery disease.

Non-invasive testing is routinely employed for the detection of clinically suspected obstructive coronary artery disease (CAD). Stress echocardiography (SE) and myocardial perfusion imaging (MPI) facilitate non-invasive risk stratification and identification of patients with significant CAD warranting coronary angiography or revascularisation.1–6

Traditionally, invasive coronary angiography has served as the gold standard for characterising the diagnostic accuracy of non-invasive testing, generally expressed in terms of sensitivity (percentage of true-positive tests in those with angiographic CAD) and specificity (percentage of true negative tests in those without angiographic CAD).

Published values for the sensitivity and specificity of existing stress modalities for the overall detection of CAD suggest comparable accuracy between MPI and SE.4–6 A pooled analysis of 18 studies involving patients who underwent exercise or pharmacologic stress in conjunction with thallium- or technetium-labeled radioisotope imaging for the diagnosis of CAD, reported sensitivity and specificity of 80% and 86% for SE and 84% and 77% for MPI, respectively.8

Left main (LM) involvement alone or in combination with additional coronary involvement, LM equivalent (proximal left anterior descending and proximal circumflex disease) and triple vessel disease (TVD) represent high-risk subsets of obstructive CAD (high-risk CAD). LM disease defined as a >50% stenosis of this vessel is found in approximately 7% of patients undergoing coronary angiography10 and is associated with significant cardiac mortality.11,12 Similarly, 5-year survival with medical therapy in patients with TVD ranges between 79% to as low as 59%, when the proximal left anterior descending artery is involved.13 Revascularisation in LM14–16 and TVD17 has been shown to positively impact survival.

While the available literature on MPI and SE suggest similar performance between these modalities for the overall detection of CAD,3–8,18–19 data on the comparative performance of SE versus MPI for the detection of “high-risk CAD” are very limited.4,6,7 Specifically, no conclusive evidence as to whether MPI or SE is more vulnerable to false negative results (low sensitivity) among persons with LM and TVD CAD exists in the literature. This meta-analysis was undertaken to compare the diagnostic performance of SE and MPI against the conventional gold standard of coronary angiography, using existing data on patients with LM and TVD.
METHODS
Data extraction
A systematic electronic literature search of OVID, EMBASE, the Cochrane database and MEDLINE using PubMed as the search engine was used to identify studies reporting comparative data on SE and MPI for LM and TVD (1980–2007 in the English language). The following keywords were used for the search: “stress echocardiography, dobutamine, dipyridamole, adenosine, technetium-99m sesambl, technetium-99m tetrofosmin and thallium-201 myocardial perfusion imaging”, single photon emission computerised tomography—myocardial perfusion imaging (SPECT-MPI) “myocardial perfusion scintigraphy”, “left main disease”, “triple vessel disease”, “coronary artery disease”, “diagnostic accuracy” and “comparative study”. Articles were also identified using the “related articles” function in PubMed. Additional references were obtained from bibliographies of original papers and review articles.

Eligibility criteria and data extraction
All studies comparing SE or MPI to conventional angiography and reporting data on LM or TVD were considered. Studies where patients did not undergo angiography and studies that did not report on the numbers of patients diagnosed with LM or TVD were excluded. Only studies that provided angiographic evidence of LM and TVD were included. Non-invasive imaging tests included exercise or pharmacologic SE and exercise or pharmacologic MPI. Authors of these studies were not contacted directly. Data were extracted on author, journal, date of publication, institution, study design, patient demographics (mean age, percentage of male patients) and technical aspects of the studies.

All data were extracted independently by three reviewers (NM, YW and JA), and any discrepancy was resolved by consensus. Sensitivity and specificity were required for analysis, and accordingly, studies providing insufficient information for calculations were excluded. Data were extracted using any “reversible perfusion deficit (PD)”, “electrocardiographic changes in ≥2 contiguous leads”, “transient ischemic dilatation” or “hypotension” or “other ancillary data” in studies with MPI and any “reversible regional wall motion abnormality” for SE as reported by the authors. Fixed perfusion abnormalities and wall motion abnormalities at rest were excluded during extraction of data for positive results to avoid any impact of prior myocardial infarction. We also extracted data for high-risk features including “LM” or “multivessel (≥2 coronary vessel)” scintigraphic pattern in studies with MPI and “reversible regional wall motion in multivessel territory (≥2 coronary vessel)” abnormalities for SE (stringent criterion) in order to define their diagnostic accuracy for precise identification of multivessel CAD.

Outcome measures and definitions
The primary outcome measure was the construction of summary receiver-operating characteristic (SROC) curves for both the imaging modalities. Summary receiver operating characteristic (AUC), diagnostic OR (DOR) and index Q* value as summary estimates. The SROC curve was described for use in the assessment of diagnostic tests in 1971 and reflects the discriminating ability of a diagnostic test. It can overcome some of the limitations associated with pooling sensitivities and specificities of published studies; accordingly, it has become the preferred method to assess performance of a diagnostic test based on data from a meta-analysis. Area under the SROC curve reflects overall performance of the test and is robust to heterogeneity. It is expressed as a value between 0 and 1, with higher values indicating better test performance. An area of 1.0 is representative of a perfect test, while an area of 0.5 represents a test whose performance is no different from a random guess. False positive rate (FPR = 1−specificity) is the x axis, while true positive rate (TPR = sensitivity) is the y axis of SROC curve. Index Q* represents the value of sensitivity and specificity implied by the SROC curve at the point where the two are equal. The closer it is to the top left corner, the better the accuracy (higher sensitivity and specificity). It is invariant to heterogeneity.

The individual study sensitivities and specificities, diagnostic odds ratio (DOR) and likelihood (positive and negative) ratios of the two imaging modalities (for high-risk CAD) were extracted using two-by-two contingency tables for each end-point of correct patient diagnosis. This was either quoted directly in the studies or extracted from the analysis of the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) on a per patient basis. Data per lesion were published in the included studies for MPI using scintigraphic perfusion abnormality along with ancillary data and visual assessment of regional left ventricular wall thickening/motion abnormalities in SE studies. The study defined all persons without either LM or TVD CAD as “not diseased” for the purposes of analysis. The accuracy based on index vessel stenosis severity was not estimated, as these data were not consistently available in the literature for SE and MPI.

Since patients with one- and two-vessel coronary artery disease were classified as “not diseased”, specificity for both modalities was artificially lowered.

We also computed likelihood ratios (LR) as they provide a measure of a test’s ability to “rule in” or “rule out” disease independent of disease probability. The positive LR represents the OR of a positive test result being obtained in a diseased patient population compared to non-diseased patients. A test with higher positive LR is useful in confirming the presence of disease. The negative LR represents the OR that a negative test result will be observed in a diseased population compared to the odds that the same result will be observed among a non-diseased population. A test with negative LR closer to 0 helps “rule out” disease. They are computed as:

\[
\text{Positive LR} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}
\]

\[
\text{Negative LR} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}
\]

An ideal screening test has high sensitivity and low negative LR.

Since sensitivity and specificity are only partly representative of a test’s discriminatory performance (ignoring threshold differences), DOR is widely utilised as a single indicator of test performance and is independent of the prevalence of disease. It was computed for the included studies as follows:

\[
DOR = \frac{\text{TP} \times \text{TN}}{\text{FP} \times \text{FN}} = \frac{(1 - \text{False positive rate})/(\text{False positive rate})}{(1 - \text{True positive rate})/(\text{True positive rate})}
\]

False positive rate is the x axis, while true positive rate is the y axis of the SROC curve. The value of a DOR can range from 0 to
Cardiac imaging and non-invasive testing

infection. Higher DOR values are reflective of better discriminatory performance of a diagnostic test. Pooled DOR can be derived from SROC curves.34

A weighted meta-regression (weighted by the inverse of variance of DOR in each study) was also performed for both imaging modalities to evaluate the importance of potential effect modifiers to explain variation between studies.25 We assessed the influence of variables including publication year, mean age, proportion of men in the study population, clinical indication, type of stressor (exercise or pharmacologic) and year of publication.

Heterogeneity among the included studies was assessed using the Cochrane Q test. This test generates a probability (p value), based on a \( \chi^2 \) distribution, that between study differences in results equal to or greater than those observed are likely to occur simply by chance (p<0.10 representing significant statistical heterogeneity). The I² statistic (measure of proportion of inconsistency not explained by chance) is calculated from Cochrane Q and is used to quantify heterogeneity. It can vary from 0% to 100%, with a value of 0% representative of absence of any observed heterogeneity (I² >50% represents substantial heterogeneity); \( \tau^2 \) gives an estimate of between study variance.26

Statistical analysis

SROC curves were constructed for the two imaging modalities and analysed to compare area under the curves, index Q* values and DOR to assess their discriminatory capacity for detection of high-risk CAD. A symmetric SROC model was used to address the lack of numerical convergence of the full SROC model. Pooled sensitivity, specificity, positive and negative likelihood ratios (weighted by the inverse of variance of sensitivity, specificity and likelihood ratios in each study, respectively), with 95% CI, were calculated using a random effects model.27 Heterogeneity was explored using sub-group analyses and meta-regression. Study-specific DORs were meta-regressed to test the effect of potential effect modifiers.25 We also calculated the effect of varying the positivity criteria using “LM” or “multivessel” scintigraphic pattern in studies with MPI and based on “reversible regional wall motion in multivessel territories” abnormalities for SE (stringent criterion). Sub-group analyses were performed for both the imaging modalities. Analysis was conducted using SPSS V15.0 (SPSS Inc.), Windows STATA V10.0 (StataCorp LP), StataDirect (StataDirect Ltd) and Meta-Test Software V0.9 (Tufts, New England Medical Centre). The Egger’s precision-weighted linear regression method was used to statistically assess publication bias (p<0.05 was considered representative of significant statistical publication bias).28 This study was undertaken in accordance with previously reported guidelines for meta-analyses evaluating diagnostic tests.25 27 28 29 30

RESULTS

Study eligibility

Thirty-two studies matched the selection and search criteria and were obtained for evaluation in full.31–62

High-risk CAD (LM+TVD)

Twenty articles utilised MPI for LM and TVD detection.33 35–37 39 40 43–47 49 50 56–61 Five of these studies were excluded due to lack of extractable data.42 59 Inability to abstract specificity as the studies included only patients with LM or TVD.33 50 61 Eighteen studies with comparative data on SE for LM and TVD were identified.31 32 34 36–41 47–49 51–55 62 Three studies using dipyridamole32 52 54 stress echocardiography (as this stress modality is not commonly employed in the US) and trans-esophageal echocardiography55 were excluded. Thus, we were left with 15 studies for MPI (exercise-9, adenosine-1, dipyridamole-1, dobutamine-3, both exercise and pharmacologic-1) and 14 studies for SE (exercise-5, dobutamine-9). Six studies reported use of both MPI and SE to diagnose CAD in the same cohort.36 37 39 40 47 49 The study characteristics of the included studies for high-risk CAD analysis are shown in table 1. Data on high-risk CAD patients with reduced left ventricular ejection fraction were not uniformly abstractable for the included studies.

Statistical analysis for MPI studies

A total of 2510 patients underwent MPI, ranging from 70 to 466 participants in the 15 studies, were included for analysis. The prevalence of high-risk CAD was 28% (649/2310) from the included studies. The SROC analysis for all the included studies yielded overall weighted AUC for high-risk CAD of 0.73 (0.02) with index Q* value of 0.67 (0.02) (figure 1). Using pooled summary point estimates, sensitivity, specificity, positive and negative likelihood ratios for MPI were 75% (72–78%), 48% (55–68%), 1.58 (1.35–1.74) and 0.47 (0.39–0.56) respectively. The forest plot for sensitivity (figure 2) showed heterogeneity (Cochrane Q p<0.1). Six small-sample-size studies were primarily responsible for this heterogeneity.37 39 43 45 47 49 However, there was no significant heterogeneity in the forest plot for negative likelihood ratio (1–Sensitivity/Specificity; figure 3), implying that it is reasonable to combine the included studies to obtain pooled estimates. The pooled DOR for high-risk CAD was 4.06 (2.84–5.80) and supported absence of any substantial heterogeneity with Cochrane Q of 19.93 (p=0.15) with \( \tau^2 = 29.7\% \) (figure 4).

Statistical analysis for SE studies

A total of 1403 patients underwent SE, with a range of 37–309 participants in the 14 studies included for analysis. The prevalence of high-risk CAD was 18% (254/1403) from the included studies. The SROC analysis for all the included studies yielded overall weighted AUC for high-risk CAD of 0.82 (0.05) with index Q* value of 0.75 (0.02) (figure 1). Using summary point estimates, sensitivity, specificity, positive and negative likelihood ratios for SE were 94% (91–97%), 40% (37–43%), 1.52 (1.38–1.68) and 0.21 (0.13–0.35), respectively. Lack of significant heterogeneity seen in the forest plots for sensitivity (figure 2) and negative likelihood ratios (figure 3), as evidenced by Cochrane Q p>0.1, permitted the combination of included studies to obtain pooled estimates. The pooled DOR for high-risk CAD was 9.78 (5.83–16.40; figure 4). The included studies showed no evidence of heterogeneity with Cochrane Q of 8.22 (p=0.83) and absence of any inconsistent studies (I²=0 attributing all the variability in estimates to sampling error within included studies) or between-study variance (\( \tau^2 = 0 \)).

Results of indirect comparison between MPI and SE

Our results suggest significantly higher area under SROC curve (0.73 (0.02) vs 0.82 (0.03), p=0.01) and index Q* value (0.75 (0.02) vs 0.67 (0.02), p=0.01) for SE supporting its overall superiority in detection of high-risk CAD. Additionally, SE had higher pooled sensitivity (94% vs 75%, p<0.001) and DOR (9.78 vs 4.06, p=0.02), and lower negative likelihood ratio (0.21 vs 0.47, p<0.001). No evidence of a difference in the pooled specificity and positive likelihood ratios was seen for the two imaging modalities in detection of high-risk CAD (table 2). Exclusion of heterogeneous studies yielded higher sensitivity for SE (94% vs 75%, p<0.01) without any evidence of a difference in
<table>
<thead>
<tr>
<th>ID</th>
<th>Year</th>
<th>Author</th>
<th>Method</th>
<th>Age (years)</th>
<th>Male (%)</th>
<th>Disease†</th>
<th>Total Sample</th>
<th>Positivity Criterion</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1981</td>
<td>Rehn</td>
<td>MPI</td>
<td>51</td>
<td>88</td>
<td>45</td>
<td>104</td>
<td>Reversible PD</td>
<td>37</td>
<td>8</td>
<td>21</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>1984</td>
<td>Nygaard</td>
<td>MPI</td>
<td>52</td>
<td>81</td>
<td>96</td>
<td>295</td>
<td>Reversible PD or hypotension or ECG changes</td>
<td>72</td>
<td>24</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1986</td>
<td>Maddahi</td>
<td>MPI</td>
<td>56</td>
<td>81</td>
<td>56</td>
<td>105</td>
<td>Reversible PD or hypotension or ECG changes</td>
<td>34</td>
<td>22</td>
<td>41</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1986</td>
<td>Wakasugi</td>
<td>MPI</td>
<td>NA</td>
<td>NA</td>
<td>56</td>
<td>79</td>
<td>Reversible PD</td>
<td>35</td>
<td>21</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1990</td>
<td>Mehmadian</td>
<td>MPI</td>
<td>56</td>
<td>74</td>
<td>13</td>
<td>236</td>
<td>Reversible PD</td>
<td>13</td>
<td>0</td>
<td>86</td>
<td>197</td>
</tr>
<tr>
<td>6</td>
<td>1991</td>
<td>Chikamori</td>
<td>MPI</td>
<td>61</td>
<td>78</td>
<td>133</td>
<td>466</td>
<td>Reversible PD or hypotension or ECG changes</td>
<td>76</td>
<td>59</td>
<td>251</td>
<td>82</td>
</tr>
<tr>
<td>7</td>
<td>1992</td>
<td>Quinones</td>
<td>MPI</td>
<td>57</td>
<td>67</td>
<td>16</td>
<td>72</td>
<td>Reversible PD</td>
<td>15</td>
<td>1</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>1993</td>
<td>Solot</td>
<td>MPI</td>
<td>61</td>
<td>62</td>
<td>20</td>
<td>127</td>
<td>Reversible PD</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>78</td>
</tr>
<tr>
<td>9</td>
<td>1993</td>
<td>Hecht</td>
<td>MPI</td>
<td>58</td>
<td>86</td>
<td>15</td>
<td>71</td>
<td>Reversible PD</td>
<td>10</td>
<td>5</td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>1994</td>
<td>Takeishi</td>
<td>MPI</td>
<td>63</td>
<td>63</td>
<td>11</td>
<td>75</td>
<td>Reversible PD or abnormal L/H ratio or LV dilation</td>
<td>9</td>
<td>2</td>
<td>42</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>1997</td>
<td>Oguzhan</td>
<td>MPI</td>
<td>51</td>
<td>84</td>
<td>14</td>
<td>70</td>
<td>Reversible PD</td>
<td>14</td>
<td>0</td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>12</td>
<td>1998</td>
<td>Ehendy</td>
<td>MPI</td>
<td>60</td>
<td>63</td>
<td>15</td>
<td>84</td>
<td>Reversible PD</td>
<td>11</td>
<td>4</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>13</td>
<td>1998</td>
<td>Ehendy</td>
<td>MPI</td>
<td>58</td>
<td>0</td>
<td>5</td>
<td>70</td>
<td>Reversible PD</td>
<td>5</td>
<td>0</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>14</td>
<td>1999</td>
<td>Fragaesso</td>
<td>MPI</td>
<td>61</td>
<td>54</td>
<td>11</td>
<td>101</td>
<td>Reversible PD</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>73</td>
</tr>
<tr>
<td>15</td>
<td>2003</td>
<td>Lima</td>
<td>MPI</td>
<td>61</td>
<td>65</td>
<td>143</td>
<td>112</td>
<td>Reversible PD or wall motion or ECG or TID or LV enlargement or reduced LVEF</td>
<td>126</td>
<td>17</td>
<td>38</td>
<td>74</td>
</tr>
<tr>
<td>16</td>
<td>1991</td>
<td>Sawada</td>
<td>SE</td>
<td>59</td>
<td>62</td>
<td>7</td>
<td>55</td>
<td>New or worse RWMA</td>
<td>7</td>
<td>0</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td>17</td>
<td>1992</td>
<td>Epstein</td>
<td>SE</td>
<td>59</td>
<td>70</td>
<td>7</td>
<td>61</td>
<td>New or worse RWMA</td>
<td>7</td>
<td>0</td>
<td>11</td>
<td>43</td>
</tr>
<tr>
<td>18</td>
<td>1992</td>
<td>Quinones</td>
<td>SE</td>
<td>57</td>
<td>67</td>
<td>16</td>
<td>112</td>
<td>New or worse RWMA</td>
<td>15</td>
<td>1</td>
<td>45</td>
<td>52</td>
</tr>
<tr>
<td>19</td>
<td>1993</td>
<td>Hecht</td>
<td>SE</td>
<td>58</td>
<td>86</td>
<td>15</td>
<td>71</td>
<td>New or worse RWMA</td>
<td>14</td>
<td>1</td>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>20</td>
<td>1993</td>
<td>Ryan</td>
<td>SE</td>
<td>57</td>
<td>75</td>
<td>55</td>
<td>309</td>
<td>New or worse RWMA</td>
<td>53</td>
<td>2</td>
<td>92</td>
<td>162</td>
</tr>
<tr>
<td>21</td>
<td>1994</td>
<td>Prince</td>
<td>SE</td>
<td>62</td>
<td>86</td>
<td>9</td>
<td>73</td>
<td>New or worse RWMA</td>
<td>8</td>
<td>1</td>
<td>50</td>
<td>14</td>
</tr>
<tr>
<td>22</td>
<td>1994</td>
<td>Sharp</td>
<td>SE</td>
<td>58</td>
<td>NA</td>
<td>9</td>
<td>54</td>
<td>New or worse RWMA</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>23</td>
<td>1995</td>
<td>Bjornstad</td>
<td>SE</td>
<td>58</td>
<td>81</td>
<td>6</td>
<td>37</td>
<td>New or worse RWMA</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>1997</td>
<td>Oguzhan</td>
<td>SE</td>
<td>51</td>
<td>84</td>
<td>14</td>
<td>70</td>
<td>New or worse RWMA</td>
<td>14</td>
<td>0</td>
<td>24</td>
<td>32</td>
</tr>
<tr>
<td>25</td>
<td>1997</td>
<td>Hennessy</td>
<td>SE</td>
<td>59</td>
<td>64</td>
<td>67</td>
<td>221</td>
<td>New or worse RWMA</td>
<td>65</td>
<td>2</td>
<td>55</td>
<td>99</td>
</tr>
<tr>
<td>26</td>
<td>1997</td>
<td>Rallihas</td>
<td>SE</td>
<td>57</td>
<td>NA</td>
<td>18</td>
<td>85</td>
<td>New or worse RWMA</td>
<td>16</td>
<td>2</td>
<td>21</td>
<td>46</td>
</tr>
<tr>
<td>27</td>
<td>1998</td>
<td>Ehendy</td>
<td>SE</td>
<td>60</td>
<td>63</td>
<td>15</td>
<td>84</td>
<td>New or worse RWMA</td>
<td>12</td>
<td>3</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>28</td>
<td>1998</td>
<td>Ehendy</td>
<td>SE</td>
<td>58</td>
<td>0</td>
<td>5</td>
<td>70</td>
<td>New or worse RWMA</td>
<td>5</td>
<td>0</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>29</td>
<td>1999</td>
<td>Fragaesso</td>
<td>SE</td>
<td>61</td>
<td>54</td>
<td>11</td>
<td>101</td>
<td>New or worse RWMA</td>
<td>10</td>
<td>1</td>
<td>41</td>
<td>49</td>
</tr>
</tbody>
</table>

The TPR or TNR have to be modified when reported TPR or TNR=1 since 0 cannot be in the 2×2 table. We add 0.5 to each cell of the 2×2 table if any 0 exists as recommended by Walter. The TPR or TNR have to be modified when reported TPR or TNR=1 since 0 cannot be in the 2×2 table. We add 0.5 to each cell of the 2×2 table if any 0 exists as recommended by Walter. TP, true positive; FN, false negative; TN, true negative; FP, false positive; MPI, myocardial perfusion imaging; SE, stress echocardiography. Disease-cumulative number of patients with left main and triple vessel disease.
Cardiac imaging and non-invasive testing

Figure 1  SROC comparison for SE and MPI for detection of high-risk CAD (LM or TVD) The points are shrunk estimates taken from the fitted random effects model, and may not correspond to the sensitivities and specificities apparent from table 1. (SROC, summary receiver operating characteristic curves; AUC, area under curves; SE, stress echocardiography; Q*, Index Q* value).

Sensitivity (95% CI)  
- Oguzhan 1.00 (0.77 - 1.00)  
- Elhendy 0.80 (0.52 - 0.96)  
- Hecht 0.93 (0.68 - 1.00)  
- Epstein 1.00 (0.59 - 1.00)  
- Prince 0.89 (0.52 - 1.00)  
- Sawada 1.00 (0.59 - 1.00)  
- Frangasso 0.91 (0.59 - 1.00)  
- Bjornstad 0.83 (0.36 - 1.00)  
- Elhendy 1.00 (0.48 - 1.00)  
- Ryan 0.96 (0.87 - 1.00)  
- Quinones 0.94 (0.70 - 1.00)  
- Hennessy 0.97 (0.90 - 1.00)  
- Sharp 1.00 (0.66 - 1.00)  
- Rallidis 0.89 (0.65 - 0.99)  

Random Effects Model  
- Pooled Sensitivity = 0.94 (0.91 to 0.97)  
- Chi-square = 12.41; df = 13 (p = 0.4944)  
- Inconsistency (I-square) = 0.0 %

SE (AUC) = 0.8195  
SE (Q*) = 0.7531  
SE (Q') = 0.0240

SROC Curve

Sensitivity

1

1

0.2

0.4

0.6

0.8

1

1-specificity

Figure 2  Comparison of sensitivity of SE and MPI for detection of high-risk CAD.
3. Comparison of the performance of SE and MPI for detection of high-risk CAD using data from studies with reported outcome data for both stress modalities by the same author.\textsuperscript{26,37,39,40,47,49}

The sample size was significantly lower for this analysis and added up to roughly one-sixth of patients with high-risk CAD. SE had a significantly higher area under SROC curve (0.80 vs 0.70, p $< 0.01$) without any evidence of a difference in specificity (46% vs 39%), p = 0.14) and DOR (7.00 vs 4.47, p = 0.04) in comparison to MPI.

4. Comparison of the performance of SE and MPI for detection of high-risk CAD excluding data from studies with reported outcome data for both stress modalities by the same author.\textsuperscript{31,32,34,39,40,41,43-46,48} 52-56 58 57 60 62

This analysis included 2757 patients. Even after excluding the six common studies that contributed to both stress modalities, the conclusion (better discriminatory ability of SE) remained unchanged with a significantly higher area under SROC curves (0.85 vs 0.75, p = 0.01), sensitivity (96% vs 74%, p < 0.001) and DOR (9.05 vs 4.30, p = 0.01).

5. Comparison of the performance of SE and MPI for detection of high-risk CAD using studies categorised by stressor modality (exercise or pharmacological).

Regardless of type of stressor used, comparison of studies involving SE and MPI using exercise or pharmacological stressor yielded results consistent with the overall trend observed (study where data on exercise and pharmacologic stressor was not abstractable (60) was excluded from this analysis). The sensitivity of SE remained significantly higher compared to MPI. (Exercise — 76% vs 95%, p < 0.01; Pharmacologic — 62% vs 93%, p < 0.01).

6. Comparison of the performance of SE and MPI for detection of Left Main disease

The absence of adequate number of studies for SE (n = 3) and MPI (n = 6) reporting abstractable data on left main disease detection precluded a meaningful analysis and therefore was not pursued further. These sub-group analyses appear to support the superior ability of SE for detection of high-risk CAD.


discussion

The results of our meta-analysis suggest that SE has a better discriminatory capacity (on SROC analysis) than MPI for the detection of high-risk CAD. This difference remains significant even after adjustment of sex, mean age of sample population and year of publication. Both tests showed high AUC values (table 2) with high DORS (figure 4). However, SE appears to have a higher sensitivity and lower negative likelihood ratio than MPI when performed in persons with LM or TVD. These findings suggest that SE is less vulnerable than MPI to false negative findings may have important clinical implications. Importantly, this is the first comparative analysis of two routinely employed stress-imaging modalities in the assessment harmonic and strain imaging) precluded any meaningful contemporary comparisons.

Figure 3 Comparison of negative likelihood ratios of SE and MPI for detection of high-risk CAD.

Figure 4 Comparison of DORs for studies using SE or MPI for detection of high-risk CAD.
of high-risk CAD. The finding that SE has a higher discriminatory capacity compared to MPI for the detection of high-risk CAD was consistently observed both in pooled analyses of studies that employed both SE and MPI in each study participant and in pooled analyses of studies that employed only one or the other modality to image each study participant.

Overall comparative data between imaging modalities

Both SE and MPI have proven to be comparable and extremely useful in the overall evaluation of CAD. Several studies comparing diagnostic accuracy of SE with MPI in the same population have shown that although both tests have comparable sensitivity for the diagnosis of CAD, SE has higher specificity. In one study that compared dobutamine–atropine SE to dipyridamole–sestamibi perfusion imaging in the same population, both modalities had similar sensitivities; however, dobutamine-atropine SE demonstrated a higher specificity (95% vs 76%). Similarly, a meta-analysis of 17 studies reported a marginal decrease in sensitivity for SE compared to MPI (80% vs 84%, p < 0.05) but a higher specificity (77% vs 86%, p < 0.001) for the overall detection of CAD. This may be ascribed to the fact that MPI assesses relative and regional differences in myocardial perfusion, the initial event in the ischemic cascade, which precedes the development of contractile abnormalities (the basis of SE).

Underestimation of high-risk CAD by MPI

It is important to recognise that the favourable overall MPI diagnostic accuracy data may not be extrapolated to high-risk CAD sub-groups. Several factors may contribute to the underestimation of high-risk CAD and include inadvertent or unrecognised ingestion of caffeine containing products or aminophylline (antidotes to adenosine or dipyridamole), anti-angina therapy, methodological artefacts and attenuation from motion, soft tissue or diaphragmatic attenuation and or aminophylline (antidotes to adenosine or dipyridamole), underestimation of high-risk CAD by MPI

<table>
<thead>
<tr>
<th>Coronary vessel</th>
<th>MPI</th>
<th>SE</th>
<th>p Value</th>
<th>MPI</th>
<th>SE</th>
<th>p Value</th>
<th>MPI</th>
<th>SE</th>
<th>p Value</th>
<th>Meta Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk CAD (LM and TVD)</td>
<td>0.75</td>
<td>0.94</td>
<td>&lt;0.0001</td>
<td>0.48</td>
<td>0.40</td>
<td>0.16</td>
<td>0.73</td>
<td>0.82</td>
<td>0.01</td>
<td>4.06</td>
</tr>
<tr>
<td>(0.72–0.78)</td>
<td>(0.91–0.97)</td>
<td>(0.35–0.68)</td>
<td>(0.37–0.43)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(2.84–5.80)</td>
<td>(5.83–16.40)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SROC, summary receiver operating characteristic curve; AUC, area under curve; SE, standard error; DOR, diagnostic odds ratio.
Cardiac imaging and non-invasive testing

CAD. Additionally, appropriate spatial normalisation, gating and the use of three-dimensional SPECT MPI have been recommended to improve detection of severe CAD. However, regrettably, these strategies remain widely under-utilised in the “real world” of contemporary nuclear cardiology practice, worldwide. Use of hybrid imaging (such as 82 Rubidium PET/CT imaging) technology has been found to provide incremental diagnostic information compared to MPI alone (permitting clinicians to delineate coronary anatomy, quantitate absolute blood flow, ejection fraction reserve and assess physiologic significance of flow limiting stenosis). It is hoped that hybrid imaging will accurately characterise patients with high-risk CAD.

Limitations of stress echocardiography

Although the current study suggests that stress echocardiography has better discriminatory capacity for the detection of high-risk CAD, this technique has its own inherent limitations. False negativity has been observed in patients who do not achieve target workload. The lack of interpretable images during exercise (obese, chronic obstructive pulmonary disease), delay in recording images post treadmill (secondary to intense respiratory effort and discomfort following exercise) and inability to image the entire left ventricle in some patients can adversely affect interpretation. Lack of quantitation and visual assessment essentially contribute to the concerning inter-observer and inter-institutional variability. Small LV size can make interpretation of RWMA more challenging. The hallmark of ischemia (per cent LV wall thickening) may be limited by increased diastolic thickness (compensatory response to decreased LV preload from dobutamine) as compared to LV systolic thickness. Conversely, localised basal inferior wall abnormalities, cardiomyopathies and abnormal septal wall motion (post bypass, left bundle branch block) can lead to false positive results. Dobutamine infusion may also be associated with ventricular arrhythmias in patients with low ejection fraction and high-risk CAD.

Referral (verification) bias

It is plausible that referral bias contributes to an overestimation of the sensitivity of MPI in the detection of high-risk CAD as the extent of ischemia is generally the primary determinant of the need for subsequent catheterisation and/or revascularisation. Notably, a significant proportion of patients with significant LM disease or multivessel disease may not undergo cardiac catheterisation because of reportedly normal perfusion scans. This issue remains poorly defined and the precise magnitude of underestimation of high-risk CAD remains obscure due to the paucity of reported data on patients who have concurrently undergone both stress testing and angiography to enable a direct comparison.

Although it is unclear what threshold of high-risk anatomy is predictive of the temporal propensity for a cardiac event, in large cohorts, a normal nuclear perfusion scan, even in patients with a high pretest likelihood of CAD, is associated with a low annual event rate (cardiac death and hard event rate of 0.6% and 1.3%, respectively) and when detected, the extent of ischemia has been shown to correlate with short-term survival benefit from revascularisation. Similarly, a normal exercise echocardiogram result is associated with a <1% annual event rate of cardiac death and nonfatal myocardial infarction, comparable to that of an age and sex-matched population.

Clinical implications

For the overall detection of CAD, two meta-analyses comparing exercise SE and exercise MPI showed similar sensitivities for the detection of CAD for both modalities but higher specificities and discriminatory capabilities for SE. Both stress modalities have been validated extensively for assessing prognosis, with a negative test result associated with a low “hard event” rate at follow-up. Based on recent meta-analytic data, the negative predictive value over approximately 5-year follow-up for both techniques is high, reported to be 98.8% and 98.4% for exercise MPI and exercise echocardiography, respectively, with equivalent prognostic utility in both women and men. Another analysis reported an annual cardiac event rate of <1.2% and an annual cardiac death rate of 0.08% for both stress modalities. Similar data have been reported for perioperative risk stratification and outcomes, with one meta analysis of patients undergoing major vascular surgery suggesting a trend for better diagnostic accuracy with SE.

The current study presents a meta-analysis of the available literature on SE and MPI for the detection of LM and TVD. In these high-risk subsets of CAD, our analysis suggests that SE may be superior to MPI for screening of patients for LM and TVD. Although SE has been shown to be sensitive for diagnosing CAD, it is less useful for the accurate localisation of the diseased vessels, particularly in the setting of LM and TVD compared to single-vessel disease. This may be due to the fact that this modality relies on the identification of a combination of symptoms, hemodynamic parameters and more importantly wall motion abnormalities. Additionally, disparate ischemic thresholds in areas subtended by functionally significant stenoses in patients with TVD may result in the rapid recognition of contractile abnormalities with or without ECG symptoms or chest pain in the territory subtended by the most significant stenosis, typically prompting the termination of the test and precluding the detection of ischemia in segments with higher ischemic thresholds. The lack of radiation, relative accuracy, availability and practicality of SE versus MPI are also important considerations that have direct impact and implications for screening CAD programs and chest pain centres.

Limitations

First, there are inherent drawbacks to comparing non-invasive tests that address the physiologic significance of a stenosis to coronary anatomy defined by angiography with its well-described shortcomings. Nevertheless, invasive coronary angiography remains the gold standard for the diagnosis and assessment of extent and severity of CAD. Second, heterogeneity between study design and results is a common problem in any systematic review or meta-analysis for diagnostic tests. Identifying its presence, investigating its cause and correctly accounting for it are important aspects to be addressed by every researcher. Such a comparison has to ideally be adjusted for study and patient characteristics to reduce the variation between studies. There is published evidence that the overall validity of the adjusted indirect comparisons depends on the internal validity and similarity of the included studies. Hence, we use weighted analysis (pooled sensitivity, specificity, likelihood ratios weighted by the inverse of variance of sensitivity, specificity and likelihood ratios, meta-regression analysis weighted by the inverse of variance of DOR in each study) using a random effects model. The presence of heterogeneity in the included studies (resulting from threshold effect, different study population, differences in technology, operators and study designs) may affect the overall validity of the pooled estimates. However, the SROC analysis is robust to heterogeneity.
Cardiac imaging and non-invasive testing

Additionally, we explored data using meta-regression analysis with the following covariates (mean age, publication year, sex).

Meta-analysis is also subject to publication bias—that is, studies with a significant result are more likely to be submitted and published. In order to minimise this bias, we did a comprehensive screening of both full-text articles and abstracts (with unpublished manuscripts). However, only those articles that used coronary angiography as the reference standard and where data was abstractable for patients with left main and triple vessel disease were included in this analysis. It was ruled out using precision-weighted linear regression.28 Additionally, we tested the effect of potential modifiers such as year of publication, mean age, sex (proportion of men), clinical indication and type of stressor on test performance in multivariate analysis (significance level of p<0.05). The protocols for conducting stress tests (including holding certain medications), tracers and methods of interpretation and reporting (with inherent inter observer variations) vary from one institution to another and may affect the results. The information on the assessment of wall motion with low dose stressor and the classic biphasic response with low- and high-dose stressor was not consistently available in studies with SE. None of the studies with SE used recovery phase wall motion assessments. The low specificities for both the stress modalities in our study are secondary to the way disease was defined. Patients without LM or TVD were defined as not diseased. However, this procedure was uniformly followed for both SE and MPI to avoid bias. Additionally, the absence of quantitative coronary angiography may have introduced observer bias in the identification of high-risk CAD. The specificity of MPI and SE was similar in our study in contrast to published literature that shows higher specificity of SE for overall detection of CAD.59

In patients with reduced left ventricular function, the issue of identification of patients with left main and triple vessel disease (vs. global hypokinesis in non-ischaemic, primary dilated cardiomyopathy) is clinically relevant. However, with the exception of one study,61 the absence of uniformly abstractable data on high-risk CAD patients with reduced ejection fraction from studies included in this meta-analysis thwarted attempts to examine this important question further. Use of computed tomographic angiography or hybrid imaging (to quantify absolute blood flow and assess the physiologic significance of stenotic lesions) may help circumvent the pitfalls of MPI and SE in this subset of patients.

Finally, the limitations of this study also include the limitations of indirect meta-regressive techniques. Results of indirect comparisons do not always agree with the results of direct comparative studies, even randomised control trials.23 A contemporary prospective head to head comparison of the two imaging modalities can overcome this limitation. In the absence of direct comparative data published or available at the present time, these findings represent the only systematic analysis of the literature on this clinically relevant issue.

CONCLUSIONS

Our meta-analysis suggests that SE has superior overall discriminatory capacity (SROC analysis) with greater sensitivity and lower negative LR compared to myocardial perfusion imaging for the detection of LM or TVD. Our findings appear to indicate that SE may be less vulnerable than myocardial perfusion imaging to false negative study results when performed in persons with high-risk CAD.

Competing interests None.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

Cardiac imaging and non-invasive testing