ORIGINAL RESEARCH

American Heart Association | American Stroke Association®

Clinical Implications of Referral Bias in the Diagnostic Performance of Exercise Testing for Coronary Artery Disease

Joseph A. Ladapo, MD, PhD; Saul Blecker, MD, MHS; Michael R. Elashoff, PhD; Jerome J. Federspiel, PhD; Dorice L. Vieira, MLS, MA, MPH; Gaurav Sharma, MD; Mark Monane, MD; Steven Rosenberg, PhD; Charles E. Phelps, PhD; Pamela S. Douglas, MD

Background—Exercise testing with echocardiography or myocardial perfusion imaging is widely used to risk-stratify patients with suspected coronary artery disease. However, reports of diagnostic performance rarely adjust for referral bias, and this practice may adversely influence patient care. Therefore, we evaluated the potential impact of referral bias on diagnostic effectiveness and clinical decision-making.

Methods and Results—Searching PubMed and EMBASE (1990–2012), 2 investigators independently evaluated eligibility and abstracted data on study characteristics and referral patterns. Diagnostic performance reported in 4 previously published meta-analyses of exercise echocardiography and myocardial perfusion imaging was adjusted using pooled referral rates and Bayesian methods. Twenty-one studies reported referral patterns in 49 006 patients (mean age 60.7 years, 39.6% women, and 0.8% prior history of myocardial infarction). Catheterization referral rates after normal and abnormal exercise tests were 4.0% (95% CI, 2.9% to 5.0%) and 42.5% (36.2% to 48.9%), respectively, with odds ratio for referral after an abnormal test of 14.6 (10.7 to 19.9). After adjustment for referral, exercise echocardiography sensitivity fell from 84% (80% to 89%) to 34% (27% to 41%), and specificity rose from 77% (69% to 86%) to 99% (99% to 100%). Similarly, exercise myocardial perfusion imaging sensitivity fell from 85% (81% to 88%) to 38% (31% to 44%), and specificity rose from 69% (61% to 78%) to 99% (99% to 100%). Summary receiver operating curve analysis demonstrated only modest changes in overall discriminatory power but adjusting for referral increased positive-predictive value and reduced negative-predictive value.

Conclusions—Exercise echocardiography and myocardial perfusion imaging are considerably less sensitive and more specific for coronary artery disease after adjustment for referral. Given these findings, future work should assess the comparative ability of these and other tests to rule-in versus rule-out coronary artery disease. (***J Am Heart Assoc***. 2013;2:e000505 doi: 10.1161/JAHA.113.000505)

Key Words: coronary artery disease • diagnostic performance • echocardiography • exercise testing • myocardial perfusion imaging

From the Department of Population Health, New York University School of Medicine, New York, NY (J.A.L., S.B.); CardioDx, Inc, Palo Alto, CA (M.R.E., M.M., S.R.); Department of Medicine, Duke University School of Medicine, Durham, NC (J.J.F., P.S.D.); Department of Health Policy and Management, UNC Gillings School of Global Public Health, Chapel Hill, NC (J.J.F.); University of North Carolina School of Medicine, Chapel Hill, NC (J.J.F.); New York University Health Sciences Libraries, New York, NY (D.L.V.); Department of Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA (G.S.); Departments of Economics and Public Health Sciences, University of Rochester, Rochester, NY (C.E.P.).

Exercise testing with radionuclide imaging or echocardiography is widely used to evaluate patients with suspected cardiovascular disease, and >12 million studies are performed annually in the United States.[1,2] The American College of Cardiology/American Heart Association practice guidelines for exercise testing recommend that physicians interpret stress test results in the context of a patient's pretest risk, using posttest disease probability as well as test performance characteristics to guide clinical decision-making.[3] However, an important limitation of this approach is that reports of stress test diagnostic performance may be influenced by referral bias[4–6]—sometimes called "verification bias" or "workup bias"—and studies do not routinely adjust for this phenomenon.[7–10] Referral bias occurs when patients with an abnormal stress test result are referred to cardiac catheterization at a higher rate than are patients with normal stress test results. While clinically appropriate, failing to adjust for this difference in referral rates when measuring test performance can significantly distort the observed diagnostic characteristics of exercise testing. This is because patients

referred for cardiac catheterization, the gold standard, may have a higher likelihood of disease than those who are not.[11,12] An exception is the work by Gibbons and colleagues, in which a posttest referral bias correction was applied, with adjustment of sensitivity and specificity.[3,11]

The clinical implications of referral bias for patient management and decision making have received little prior attention but may be substantial. However, for reasons that are unclear, referral bias is almost universally unaccounted for in studies of exercise testing.[7–10] We hypothesize that this may be because prior studies of referral bias enrolled populations from single centers,[5,11–14] thus limiting their impact due to concerns about generalizability. To address this issue, we systematically reviewed the literature on referral rates after exercise echocardiogram (ECHO) or myocardial perfusion imaging (MPI) and used these rates to adjust measures of exercise test performance. We also examined the potential impact of referral bias on posttest disease risk and clinical decision making.

Methods

Search Strategy

We searched PubMed and EMBASE from January 1990 to November 2012 for English-language articles reporting cardiac catheterization referral rates after normal or abnormal exercise MPI and ECHO. Our search terms were developed with a clinical and graduate medical education librarian (D.V.) and included the Medical Subject Headings (MeSH) *coronary disease, exercise test, myocardial perfusion imaging, single photon emission computed tomography, echocardiography,* and *humans*; keywords identifying *exercise tests, including stress test, thallium, sestamibi,* and *technetium*; and keywords identifying referral, including *refer* and *referr** (for "referral" and variants), *verif** (for "verification" and variants), and *select, selected,* and *selecti** (for "selection" and variants). We also searched the reference list of meta-analyses of exercise test performance and identified additional publications through discussion between collaborators. Our report adheres to guidelines for systematic reviews published by the Meta-analysis of Observational Studies in Epidemiology (MOOSE) group (see Appendix S1 for details).

Study Selection

Two investigators (J.L. and S.B.), working independently, in duplicate, reviewed all abstracts and identified studies that indicated or suggested that the authors reported referral rates after exercise testing. Studies then underwent full text data extraction if (1) exercise ECHO or exercise MPI was performed to detect or evaluate coronary artery disease (CAD) and (2) referral rates to cardiac catheterization were reported and stratified by stress test result (eg, normal, abnormal). Studies were excluded if they enrolled only patients with a history of myocardial infarction (MI), percutaneous transluminal coronary angioplasty (PTCA), or coronary artery bypass grafting (CABG); they enrolled patients with unstable coronary syndromes; or the majority of patients underwent pharmacological stress testing. Studies enrolling patients with a history of MI or revascularization were included if these patients comprised <15% of the study population. This threshold was considered reasonable because it was comparable to or lower than the prevalence of MI in our reference meta-analysis studies[9,15] and another systematic review of stress MPI.[7]

Data Extraction

The same 2 investigators performed data extraction independently, in duplicate, using a standardized protocol and reporting form. Study characteristics recorded included (1) identifying information (first author, journal, country, institution, publication year), (2) patient characteristics (mean age, percentage of male patients, percentage of patients with previous MI, PTCA, or CABG), (3) stress test characteristics (test used, type of exercise, positivity criterion, how authors defined significant coronary artery disease, eg, ≥50% stenosis or ≥70% stenosis), (4) referral patterns (number of patients with normal or abnormal stress test results and number of patients subsequently undergoing cardiac catheterization), and (5) diagnostic yield (number of true positives, false positives, true negatives, and false negatives). Disagreements between reviewers during the abstract screening and data extraction process were resolved through discussion.

Data Analysis

Both the proportion of patients referred for cardiac catheterization and the odds ratio for cardiac catheterization referral after a normal or abnormal stress test result were derived for each study, and 95% CIs were calculated. These estimates were pooled using the Mantel–Haenszel fixed-effects model, weighted with inverse variance, and the DerSimonian–Laird random-effects model.[16] We assessed the between-study heterogeneity using the Cochran Q statistic and study consistency using the I^2 statistic, which quantifies the proportion of heterogeneity that is not due to chance. If the *P*-value for the Q statistic was <0.10 or the I^2 statistic exceeded 50%, a random-effects model was reported instead of a fixed-effects model. Potential sources of heterogeneity also were explored. A 2-tailed *P*-value of <0.05 was judged as statistically significant. Because referral rates were generally not the primary outcome in any of the included studies, the possibility of publication bias was not explored. We used the METAN command of Stata (version 12, StataCorp) to perform all meta-analyses.

Adjusting Diagnostic Test Performance for Referral

As a reference standard for exercise test performance, we compiled studies of exercise ECHO and MPI from 4 sources: the most widely cited meta-analysis of stress test performance (Fleischmann et al),[15,17] a recent meta-analysis (Heijenbrok-Kal et al),[9] and 2 peer-reviewed meta-analyses from the Ontario Health Technology Assessment Series.[18,19] Henceforth, these 4 studies will be referred to collectively as our exercise test meta-analysis studies. They contributed a total of 45 unique studies of exercise ECHO (15 studies) or MPI performance (30 studies) (see Appendix S1).

We then used Bayesian methods developed by Begg and Greenes to adjust exercise test performance reported in each of these 45 studies for referral bias.[4] Their method assumes that referral to the gold standard test (cardiac catheterization) and disease status (the presence or absence of CAD) are conditionally independent given the exercise test result. This assumption is generally considered reasonable since the decision to perform cardiac catheterization can only be influenced by "visible" factors, such as the exercise test result and other clinical characteristics.[20] Diagnostic test performance with and without adjustment for the referral process can then be derived with the equation:

$$\frac{\Pr(T|D)}{\Pr(\bar{T}|D)} = \frac{\Pr(T|D,R+)\Pr(R+|\bar{T})}{\Pr(\bar{T}|D,R+)\Pr(R+|T)}$$

where R is referral, which is synonymous here with cardiac catheterization; T is exercise test result or its complement ($\bar{T}$), equivalent to a normal or abnormal result; D is disease status; $\Pr(T|D)$ is the sensitivity (when T is abnormal and D is disease presence) of the exercise test, *accounting* for the referral process; and $\Pr(T|D, R+)$ is the sensitivity of the exercise test, as determined in the cohort undergoing cardiac catheterization (*not accounting* for the referral process). An analogous equation can be derived for specificity by reversing disease status and test result.

Estimates of exercise test sensitivity and specificity in cohorts undergoing cardiac catheterization were derived directly from our 45 exercise test meta-analysis studies, and we adjusted these values with pooled estimates of referral rates from our literature search. To ensure we applied referral corrections appropriately, we rated each relevant study in our exercise test meta-analyses by its likelihood of referral bias. Fleischmann et al[15] also categorized studies by their likelihood of referral bias, and we limited our analyses to studies for which referral bias was rated as "likely" or "certain." We then used a random-effects model to reestimate diagnostic performance for exercise ECHO and MPI after correction for referral bias. After performing a logit transformation, we used a Taylor series expansion and the delta method to construct 95% CIs.[21] The overall methodological approach is summarized in Figure 1.

Summary Performance of Each Test

To evaluate the impact of referral bias on the overall diagnostic performance of exercise MPI and ECHO, we performed summary receiver operating characteristic (ROC) curve analysis.[22] ROC curves illustrate the trade-off between sensitivity and specificity as the threshold for defining a diagnostic test result as abnormal is varied, and they adjust for the possibility that different studies may use different test thresholds.[23] To perform our analysis, we logistically transformed the true-positive rate (TPR; sensitivity) and false-positive rate (FPR, 1−specificity) and fit a linear regression model, with the log-odds ratio (OR; log-odds TPR−log-odds FPR) as the dependent variable and the test threshold (log-odds TPR+log-odds FPR) as the independent variable.[22] The regression model included an indicator variable for referral bias correction, and its β–coefficient gives a measure of the difference in diagnostic performance after adjusting for referral. Positive coefficients indicate improved discriminatory power and negative coefficients correspond to a reduction in discriminatory power. The model's dependent variable is invariant to the referral process.[20]

Clinical Implications and Predictive Value of Diagnostic Testing

To estimate the impact of referral bias correction on posttest risk stratification and clinical decision-making, we calculated positive predictive value and negative predictive value over a
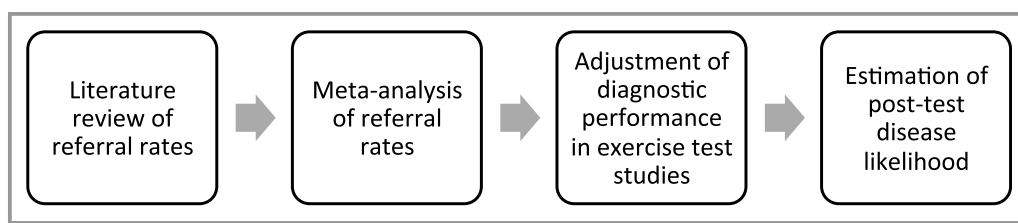
Figure 1. Flow diagram of selection process for studies included in meta-analysis.

range of pretest probabilities for CAD. We used the odds-likelihood ratio version of Bayes' theorem: posttest odds=pretest odds×likelihood ratio; where the positive likelihood ratio (LR) is sensitivity/(1−specificity) and the negative LR is (1−sensitivity)/specificity. The delta method was used to estimate 95% CIs as previously described.

Overall Methodological Approach

In summary, we searched PubMed and EMBASE for articles reporting cardiac catheterization referral rates after normal or abnormal exercise MPI and ECHO. From studies that met inclusion criteria, we extracted data on referral rates to cardiac catheterization stratified by stress test result, along with other patient and study characteristics, and pooled these referral rates. We then identified 45 studies of exercise ECHO and MPI from previously published meta-analyses, and we used Bayesian methods to adjust exercise test performance reported in each of these 45 studies for referral bias using our pooled referral rates. We also performed summary ROC curve analysis to evaluate the impact of referral bias on the overall

diagnostic performance of exercise MPI and ECHO. Finally, to estimate the impact of referral bias correction on posttest risk stratification and clinical decision making, we calculated positive predictive value and negative predictive value over a range of pretest probabilities for CAD.

Results

Literature Search

Our literature search yielded a total of 819 citations, of which 107 were selected as being potentially relevant and obtained for further screening (Figure 2). Of these 107 studies, 17 reported referral patterns after normal or abnormal exercise tests and were included in our analysis.[11,12,14,24–41] The characteristics of these studies and their 49 006 participants are shown in Table 1. The mean age was 60.7 years, 39.6% were women, 0.8% had a prior history of myocardial infarction (reported in 13 studies), and 0.1% had a prior history of revascularization (reported in 14 studies). In 13 studies, the form of exercise used was treadmill testing, and 3 studies
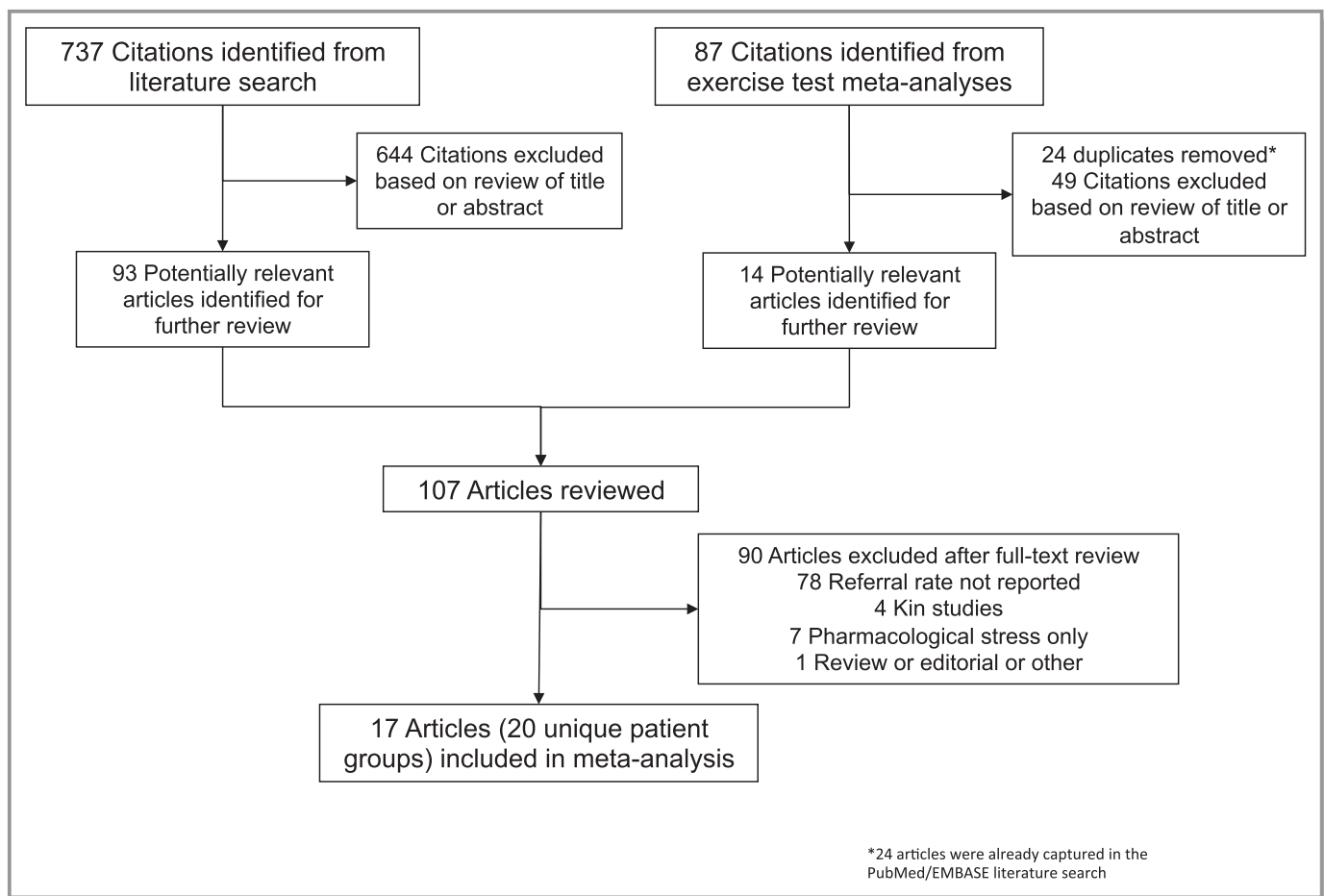

Figure 2. Overview of methodological approach.

Table 1. Characteristics of Studies Included in Meta-analysis of Referral Patterns After Normal and Abnormal Exercise Tests

Reference	Subgroup	No. of Patients	Mean Age, y	Men, %	Prior MI, %	Prior Revascularization, %*	Exercise Used	Time to Cath, d	Prevalence of CAD, %[†,‡]
ECHO									
Jang et al[35]		1287	53.2	58	0	NA	Treadmill	<30	61
Roger et al[12]	Male only	1965	60	100	0	0	Treadmill	6	80
	Female only	1714	60	0	0	0	Treadmill	6	60
Vlachopoulos et al[32]		50	59	100	0	0	Treadmill	NA	NA
Wennike et al[31]		200	62	46	NA	NA	Treadmill	NA	100
MPI									
Cecil et al[14]		2688	NA	NA	0	0	Treadmill	<90	42
Charvat et al[39]		126	59.9	60	0	NA	Bicycle	NA	56
Diamond et al[38]		9171	NA	NA	NA	NA	Treadmill	<180	29
Hachamovitch et al[37]		1021	69	55	0	0	Treadmill	<60	NA
Hannoush et al[36]		334	56	80	14	NA	Treadmill	<90	80
Hosie et al[24]		80	50	55	NA	NA	Bicycle	NA	63
Kane et al[34]		6801	61	55.4	NA	0	Treadmill	<90	60
Koistinen et al[25]		136	47.6	62	0	0	Bicycle	NA	44
Lauer et al[26]	Male only	2351	58	100	7	0	Treadmill	<90	NA
	Female only	1318	59	0	3	0	Treadmill	<90	NA
Miller et al[11]		14 273	62	59	0	0	Treadmill	<90	72
Nallamothu et al[27]		2700	59	56	0	0	NA	<180	NA
Roeters van Lennep et al[33]	Male only	322	59.4	100	0	0	Treadmill	<90	86
	Female only	294	59.1	0	0	0	Treadmill	<90	65
Schwartz et al[28]		2175	NA	100	0	0	Treadmill	NA	18

CAD indicates coronary artery disease; cath, cardiac catheterization; ECHO, echocardiography; MI, myocardial infarction; MPI, myocardial perfusion imaging; NA, not available from article.
*Jang et al,[35] Charvat et al,[39] and Hannoush et al[36] reported prior coronary artery bypass graft surgery only and did not report prior percutaneous coronary intervention.
[†]Diamond et al[38] reported prevalence only in patients with normal exercise test result.
[‡]Wennike et al,[31] Charvat et al,[39] Koistinen et al,[25] and Roeters van Lennep et al[33] reported prevalence only in patients with abnormal exercise test result.

used a bicycle; 1 study did not report the mode of exercise. Significant CAD was present in 18% to 100% of patients undergoing cardiac catheterization (reported in 13 studies). Overall, the studies were heterogeneous with respect to the population prevalence of prior coronary disease and the indications for testing and referral.

Cardiac Catheterization Referral Rates

Cardiac catheterization referral rates after normal and abnormal stress tests results from our literature review are shown in Figures 3 and 4. Because only a few studies reported referral patterns after exercise ECHO, and because we did not believe that referral patterns would differ in a clinically meaningful way between exercise ECHO and MPI, we combined all studies together. Using a random-effects model,

the pooled referral rate after a normal test result was 4.0% (95% CI, 2.9% to 5.0%), and that after an abnormal test result was 42.5% (95% CI, 36.2% to 48.9%). The pooled odds ratio for referral after an abnormal test, compared with a normal test, was 14.6 (95% CI, 10.7 to 19.9).

Diagnostic Effectiveness After Adjustment for Referral Bias

Using Bayesian methods, we adjusted the diagnostic performance reported in each of our 45 exercise test meta-analysis studies. The pooled sensitivity of exercise ECHO and MPI fell from 84% (95% CI, 80% to 89%) and 85% (95% CI, 81% to 88%) prior to adjusting for the referral process to 34% (95% CI, 27% to 41%) and 38% (95% CI, 31% to 44%) after adjustment, respectively (Table 2). The pooled specificity of exercise
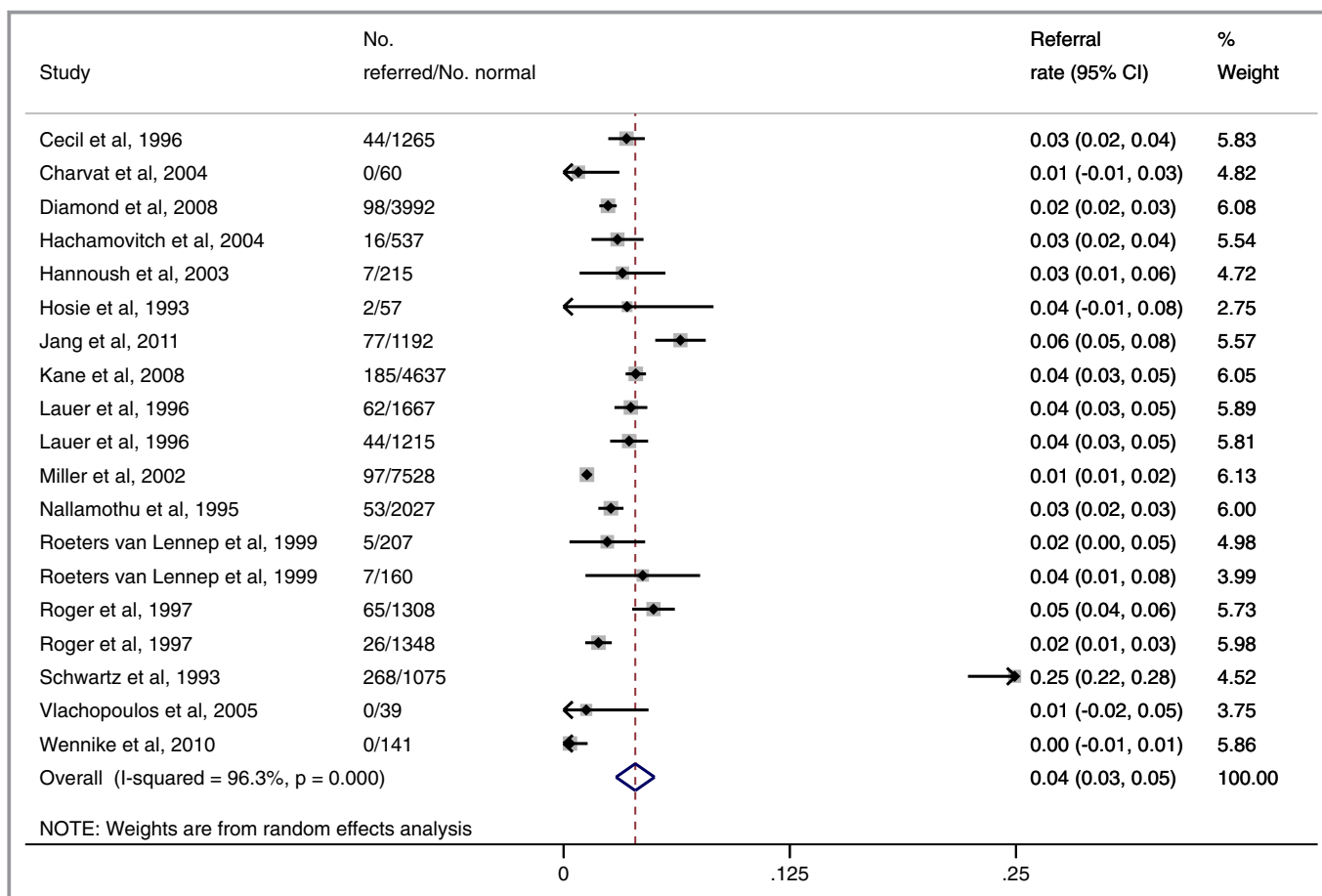
Study	No. referred/No. normal		Referral rate (95% CI)	% Weight
Cecil et al, 1996	44/1265		0.03 (0.02, 0.04)	5.83
Charvat et al, 2004	0/60		0.01 (−0.01, 0.03)	4.82
Diamond et al, 2008	98/3992		0.02 (0.02, 0.03)	6.08
Hachamovitch et al, 2004	16/537		0.03 (0.02, 0.04)	5.54
Hannoush et al, 2003	7/215		0.03 (0.01, 0.06)	4.72
Hosie et al, 1993	2/57		0.04 (−0.01, 0.08)	2.75
Jang et al, 2011	77/1192		0.06 (0.05, 0.08)	5.57
Kane et al, 2008	185/4637		0.04 (0.03, 0.05)	6.05
Lauer et al, 1996	62/1667		0.04 (0.03, 0.05)	5.89
Lauer et al, 1996	44/1215		0.04 (0.03, 0.05)	5.81
Miller et al, 2002	97/7528		0.01 (0.01, 0.02)	6.13
Nallamothu et al, 1995	53/2027		0.03 (0.02, 0.03)	6.00
Roeters van Lennep et al, 1999	5/207		0.02 (0.00, 0.05)	4.98
Roeters van Lennep et al, 1999	7/160		0.04 (0.01, 0.08)	3.99
Roger et al, 1997	65/1308		0.05 (0.04, 0.06)	5.73
Roger et al, 1997	26/1348		0.02 (0.01, 0.03)	5.98
Schwartz et al, 1993	268/1075		0.25 (0.22, 0.28)	4.52
Vlachopoulos et al, 2005	0/39		0.01 (−0.02, 0.05)	3.75
Wennike et al, 2010	0/141		0.00 (−0.01, 0.01)	5.86
Overall (I-squared = 96.3%, p = 0.000)			0.04 (0.03, 0.05)	100.00
NOTE: Weights are from random effects analysis				

	0	.125	.25

Figure 3. Cardiac catheterization referral rates after normal exercise ECHO or MPI results. Note: Lower 95% confidence intervals for some studies intersect zero. Area of each square corresponds to weight of the study in meta-analysis. CI indicates confidence interval; ECHO, echocardiography; MPI, myocardial perfusion imaging.

ECHO and MPI rose from 77% (95% CI, 69% to 86%) and 69% (95% CI, 61% to 78%) prior to adjusting for the referral process to 99% (95% CI, 99% to 100%) and 99% (95% CI, 99% to 100%) after adjustment, respectively.

Summary ROC Analysis

In a model comparing exercise ECHO without correction for referral bias to exercise ECHO with correction for referral bias, there was a trend toward a decrease in discriminatory power (parameter estimate for reduction −1.8; 95% CI, −3.6 to 0.1). No decrease in discriminatory power was found for exercise MPI (parameter estimate for reduction −0.2; 95% CI, −1.3 to 0.9).

Clinical Decision Making and Posttest Risk Stratification

Figure 5 shows how the posttest risk of CAD varies by pretest risk, with or without correction for referral bias. For both exercise ECHO and MPI, adjusting for referral resulted in an increase in the posttest risk of disease after either a normal or

abnormal test result. Posttest disease risk was comparatively higher after a normal test result primarily because referral bias adjustment significantly reduced sensitivity (thereby reducing negative predictive value). Similarly, posttest disease risk was comparatively higher after an abnormal test result primarily because referral bias adjustment significantly increased specificity (thereby increasing positive predictive value). The value of the test, in terms of guiding clinical decision making, was therefore comparatively higher after an abnormal test result than a normal test result. This is also demonstrated by the distance between the adjusted curves in Figure 5 and the 45-degree line, as the latter represents no incremental information from diagnostic testing (because pretest disease risk equals posttest disease risk along this line). Appropriately, this distance is greater after an abnormal test result than a normal test result.

Sensitivity Analyses

We evaluated the overall robustness of our results by reanalyzing our data using the upper and lower CI bounds for pooled referral rates after normal and abnormal tests,
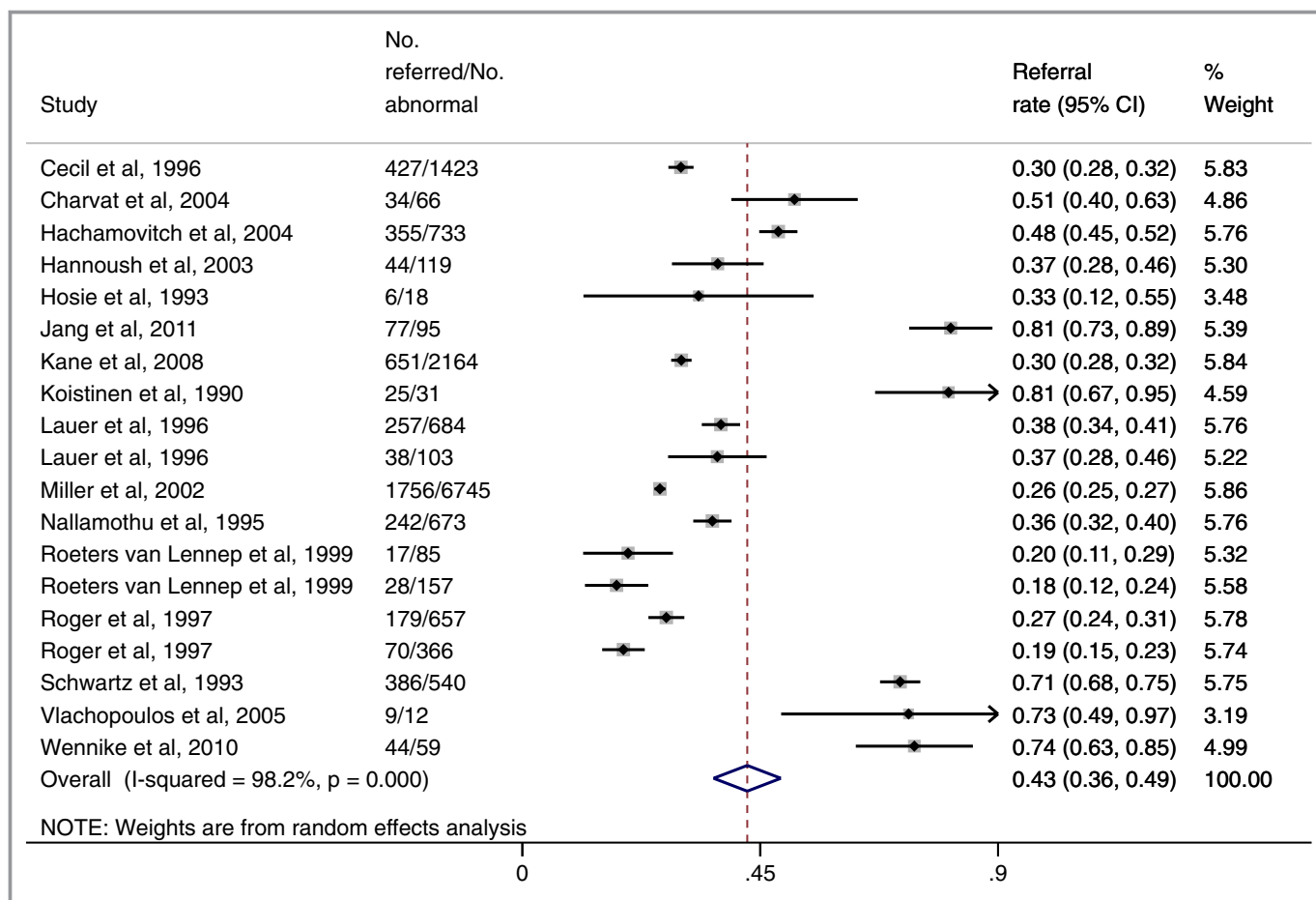
Study	No. referred/No. abnormal		Referral rate (95% CI)	% Weight
Cecil et al, 1996	427/1423		0.30 (0.28, 0.32)	5.83
Charvat et al, 2004	34/66		0.51 (0.40, 0.63)	4.86
Hachamovitch et al, 2004	355/733		0.48 (0.45, 0.52)	5.76
Hannoush et al, 2003	44/119		0.37 (0.28, 0.46)	5.30
Hosie et al, 1993	6/18		0.33 (0.12, 0.55)	3.48
Jang et al, 2011	77/95		0.81 (0.73, 0.89)	5.39
Kane et al, 2008	651/2164		0.30 (0.28, 0.32)	5.84
Koistinen et al, 1990	25/31		0.81 (0.67, 0.95)	4.59
Lauer et al, 1996	257/684		0.38 (0.34, 0.41)	5.76
Lauer et al, 1996	38/103		0.37 (0.28, 0.46)	5.22
Miller et al, 2002	1756/6745		0.26 (0.25, 0.27)	5.86
Nallamothu et al, 1995	242/673		0.36 (0.32, 0.40)	5.76
Roeters van Lennep et al, 1999	17/85		0.20 (0.11, 0.29)	5.32
Roeters van Lennep et al, 1999	28/157		0.18 (0.12, 0.24)	5.58
Roger et al, 1997	179/657		0.27 (0.24, 0.31)	5.78
Roger et al, 1997	70/366		0.19 (0.15, 0.23)	5.74
Schwartz et al, 1993	386/540		0.71 (0.68, 0.75)	5.75
Vlachopoulos et al, 2005	9/12		0.73 (0.49, 0.97)	3.19
Wennike et al, 2010	44/59		0.74 (0.63, 0.85)	4.99
Overall (I-squared = 98.2%, p = 0.000)			0.43 (0.36, 0.49)	100.00

NOTE: Weights are from random effects analysis

0   .45   .9

Figure 4. Cardiac catheterization referral rates after abnormal exercise ECHO or MPI results. CI indicates confidence interval; ECHO, echocardiography; MPI, myocardial perfusion imaging; No., number.

respectively. Selecting these values would be expected to attenuate the impact of referral bias by minimizing the relative difference in referral rates. In this analysis, the sensitivity of exercise ECHO and MPI fell to 44% (95% CI, 35% to 52%) and 48% (95% CI, 40% to 56%) and the specificity rose to 99% (95% CI, 98% to 100%) and 98% (95% CI, 97% to 98%), respectively.

We also excluded all studies of referral in which any participants were reported as having a history of MI or revascularization.[26,36] In this analysis, the referral rates after normal and abnormal exercise tests were 4.1% (95% CI, 2.8% to 5.3%) and 43.7% (36.5% to 50.9%), respectively, with odds ratio for referral after an abnormal test of 14.3 (9.8 to 20.7). After adjusting for referral, the sensitivity of exercise ECHO and MPI fell to 34% (95% CI, 26% to 41%) and 38% (95% CI, 31% to 44%) and the specificity rose to 99% (95% CI, 99% to 100%) and 99% (95% CI, 99% to 100%), respectively.

Discussion

By systematically reviewing cardiac catheterization referral rates and aggregating them to adjust pooled estimates of exercise test performance, we found that adjusting for referral

Table 2. Diagnostic Effectiveness of Exercise ECHO and MPI With and Without Adjustment for Referral

	ECHO		MPI	
	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Unadjusted*	84 (80 to 89)	77 (69 to 86)	85 (81 to 88)	69 (61 to 78)
Adjusted†	34 (27 to 41)	99 (99 to 100)	38 (31 to 44)	99 (99 to 100)

ECHO indicates echocardiography; MPI, myocardial perfusion imaging.
*Diagnostic effectiveness based on random-effects meta-analysis of sensitivity and specificity reported in 15 studies of exercise ECHO and 30 studies of exercise MPI (45 studies in total).
†Adjusted for referral rates to cardiac catheterization after abnormal or normal exercise test result.
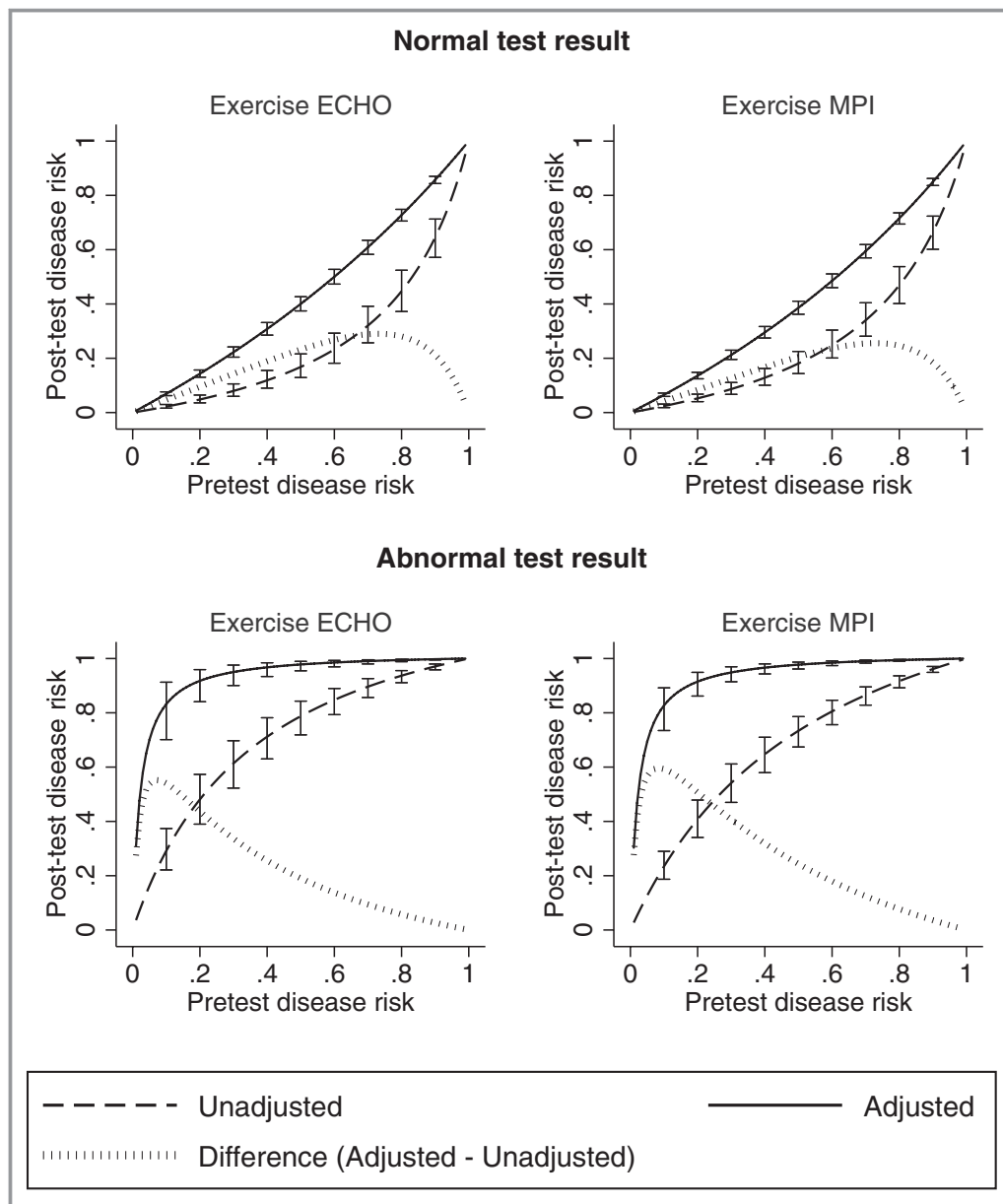
Figure 5. Posttest disease risk after normal or abnormal exercise ECHO or MPI, with and without adjustment for referral process. Note: Error bars correspond to 95% CI. ECHO indicates echocardiography; MPI, myocardial perfusion imaging.

(which was 15-fold higher after an abnormal result) significantly reduced test sensitivity and increased test specificity. While these adjustments only modestly reduced overall discriminatory power, a patient's estimated posttest disease risk shifted considerably, substantially affecting utility. To our knowledge, this analysis is the first meta-analysis to systematically apply actual cardiac catheterization referral rates after stress testing to fully understand the clinical implications of referral bias.

Our findings have important implications for clinical decision making. A recent study of Medicare patients reported that stress tests using radionuclide imaging or echocardiography account for 80% of all stress test studies performed,[42] and the volume of these procedures has grown markedly over the past 2 decades.[43] Most importantly for clinical care, we found that the sensitivity of exercise testing is much lower than previously reported. Wider recognition of this among clinicians—particularly primary care physicians and hospitalists—may influence how health care professionals use exercise ECHO and MPI to rule-in versus rule-out disease, since test specificity substantially exceeds sensitivity.

Further research is needed to understand the implications of these findings for the practicing clinician.[44] We describe 2 possible clinical scenarios that follow from our findings: A patient presenting with chest pain with a low likelihood of

CAD is considered for stress testing in an evaluation for physiologically significant CAD. If a physician preferentially aims to rule-out CAD in this patient, given the impact of referral bias on the diagnostic accuracy of stress imaging, alternative noninvasive technologies with higher sensitivity, such as coronary computed tomography angiography or other novel technologies,[45] may be more appropriate tests to use first. Another scenario is that a patient with an intermediate likelihood of CAD is similarly considered for stress testing in an evaluation for physiologically significant CAD. If a physician preferentially aims to rule-in CAD in this patient, stress imaging may be considered as the more appropriate test to use first because it is highly specific after accounting for the referral process.

We recently published the results of a multicenter, blinded trial that enrolled 537 patients with suspected CAD who were referred for stress MPI and underwent a blood-based GES.[46] To attenuate the impact of referral bias, we attempted to determine coronary anatomy in all patients using cardiac catheterization, when clinically appropriate, or coronary computed tomography angiography. Approximately 83% of eligible patients underwent at least 1 of these 2 tests, and site-read and core-lab MPI sensitivity was 27% and 36%; specificity was 92% and 90%, respectively. Though coronary computed tomography angiography is an imperfect substitute for cardiac catheterization, these results support our findings about the impact of referral bias.

Though we report significantly lower sensitivity and higher specificity than prior meta-analyses of stress testing, the prognostic value of exercise ECHO or MPI—in terms of adverse cardiovascular events—is also cited as a component of its diagnostic utility. However, while the prognostic value of a negative stress test is favorable,[47] a recent study suggests that this may be partly driven by enrollment of patients at lower risk for CAD.[48] In a study of 39 515 patients undergoing stress-rest MPI between 1991 and 2009, Rozanski and colleagues reported a significant progressive decline in the prevalence of abnormal (from 40.9% in 1991 to 8.7% in 2009) and ischemic (from 29.6% in 1991 to 5.0% in 2009) studies. These authors concluded that more cost-effective strategies for evaluating low-risk patients are needed.

Our study has several important limitations. Our adjusted values of corrected sensitivity and specificity are analytic estimates only,[49] the study populations were heterogeneous with respect to important clinical characteristics, and the validity of our results depends on how accurately referral rates from the literature review reflect those of patients comprising our exercise test meta-analysis studies. While several studies reported fairly similar referral rates, it is important to note that referral practices vary by site and provider. In addition, the diagnostic performance of exercise ECHO and MPI also varies by site, and physicians interpreting these studies may operate at different points on the ROC curve. Results obtained in nonacademic settings may also differ. Our summary ROC curve analysis partially accounted for these possibilities by evaluating overall diagnostic performance.

Another limitation is that our analysis also did not account for other important clinical characteristics that may affect diagnostic test performance and clinical decision making, including clinical factors such as patient-level risk and disease severity and nonclinical factors such as patient preferences and liability concerns. For example, patients with abnormal test results who are referred to cardiac catheterization may have more severe symptoms or a greater risk burden than patients with similar test results who are not referred. This practice would tend to attenuate the impact of referral bias on diagnostic test performance. Furthermore, the Fleischmann et al meta-analysis may not be reflective of contemporary practice, and some authors disagree with its findings. Similarly, the Ontario Health Technology Assessment series is not a widely recognized reference.

An additional limitation of our work is that cardiac catheterization is a poor gold standard for exercise testing. Recent studies have examined the importance of functional characteristics of the coronary arteries and microcirculation, rather than just coronary anatomy, and our understanding of these phenomena is growing.[50,51] However, a more accurate understanding of diagnostic performance may further improve risk factor modification in this cohort.

Conclusions

Based on pooled results from several studies and clinical sites, exercise ECHO and MPI are less sensitive and more specific for coronary artery disease after adjusting for the referral process. Accounting for this adjustment may influence how clinicians use these tests to rule-in versus rule-out disease, and more sensitive noninvasive methods for diagnosing coronary artery disease may improve patient care.

Sources of Funding

Disclosures

Drs Ladapo, Phelps, and Douglas received consulting fees from CardioDx. Drs Monane, Elashoff, and Rosenberg are employees of CardioDx. No other authors reported financial conflicts.

References

1. *The Myocardial Perfusion Imaging Market Guide (U.S.)*. Supplement to the U.S. Imaging market guide. Malvern, PA: Arlington Medical Resources, Inc; 2007.

2. *The Echocardiography Market Guide (U.S.)*. Supplement to the U.S. Imaging market guide. Malvern, PA: Arlington Medical Resources, Inc; 2009.

3. Gibbons RJ, Balady GJ, Bricker JT, Chaitman BR, Fletcher GF, Froelicher VF, Mark DB, McCallister BD, Mooss AN, O'Reilly MG, Winters WL, Antman EM, Alpert JS, Faxon DP, Fuster V, Gregoratos G, Hiratzka LF, Jacobs AK, Russell RO, Smith SC. ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (committee to update the 1997 exercise testing guidelines). *J Am Coll Cardiol*. 2002;40:1531–1540.

4. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*. 1983;39:207–215.

5. Diamond GA. Reverend Bayes' silent majority. An alternative factor affecting sensitivity and specificity of exercise electrocardiography. *Am J Cardiol*. 1986;57:1175–1180.

6. Ladapo JA, Douglas PS. Evaluation for coronary artery disease and medicare spending. *JAMA*. 2012;307:911; author reply 912.

7. Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, Fraser C, McKenzie L, Gemmell H, Hillis G, Metcalfe M. Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction. *Health Technol Assess*. 2004;8:iii–iv, 1–207.

8. Gargiulo P, Petretta M, Bruzzese D, Cuocolo A, Prastaro M, D'Amore C, Vassallo E, Savarese G, Marciano C, Paolillo S, Filardi PP. Myocardial perfusion scintigraphy and echocardiography for detecting coronary artery disease in hypertensive patients: a meta-analysis. *Eur J Nucl Med Mol Imaging*. 2011;38:2040–2049.

9. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. *Am Heart J*. 2007;154:415–423.

10. Jaarsma C, Leiner T, Bekkers SC, Crijns HJ, Wildberger JE, Nagel E, Nelemans PJ, Schalla S. Diagnostic performance of noninvasive myocardial perfusion imaging using single-photon emission computed tomography, cardiac magnetic resonance, and positron emission tomography imaging for the detection of obstructive coronary artery disease: a meta-analysis. *J Am Coll Cardiol*. 2012;59:1719–1728.

11. Miller TD, Hodge DO, Christian TF, Milavetz JJ, Bailey KR, Gibbons RJ. Effects of adjustment for referral bias on the sensitivity and specificity of single photon emission computed tomography for the diagnosis of coronary artery disease. *Am J Med*. 2002;112:290–297.

12. Roger VL, Pellikka PA, Bell MR, Chow CW, Bailey KR, Seward JB. Sex and test verification bias. Impact on the diagnostic value of exercise echocardiography. *Circulation*. 1997;95:405–410.

13. Morise AP, Diamond GA. Comparison of the sensitivity and specificity of exercise electrocardiography in biased and unbiased populations of men and women. *Am Heart J*. 1995;130:741–747.

14. Cecil MP, Kosinski AS, Jones MT, Taylor A, Alazraki NP, Pettigrew RI, Weintraub WS. 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;49:735–742.

15. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. *JAMA*. 1998;280:913–920.

16. Borenstein M. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons; 2009.

17. Literature search in Scopus for meta-analyses of exercise testing (("meta analysis" or meta-analysis) and (exercise or stress)), sorted by citation number.

18. Medical Advisory Secretariat. Stress echocardiography for the diagnosis of coronary artery disease: an evidence-based analysis. *Ont Health Technol Assess Ser*. 2010;10:1–61.

19. Medical Advisory Secretariat. Single photon emission computed tomography for the diagnosis of coronary artery disease: an evidence-based analysis. *Ont Health Technol Assess Ser*. 2010;10:1–64.

20. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. *Med Decis Making*. 1987;7:139–148.

21. Hosmer DW, Lemeshow S, May S. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Hoboken, NJ: Wiley-Interscience; 2008.

22. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48:119–130; discussion 131-112.

23. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol*. 1986;21:720–733.

24. Hosie CJ, Rodger JC, Railton R. The value of thallium scintigraphy in a district general hospital. *Nucl Med Commun*. 1993;14:12–14.

25. Koistinen MJ. Prevalence of asymptomatic myocardial ischaemia in diabetic subjects. *BMJ*. 1990;301:92–95.

26. Lauer MS, Pashkow FJ, Snader CE, Harvey SA, Thomas JD, Marwick TH. Gender and referral for coronary angiography after treadmill thallium testing. *Am J Cardiol*. 1996;78:278–283.

27. Nallamothu N, Pancholy SB, Lee KR, Heo J, Iskandrian AS. Impact on exercise single-photon emission computed tomographic thallium imaging on patient management and outcome. *J Nucl Cardiol*. 1995;2:334–338.

28. Schwartz RS, Jackson WG, Celio PV, Richardson LA, Hickman JR Jr. Accuracy of exercise 201TL myocardial scintigraphy in asymptomatic young men. *Circulation*. 1993;87:165–172.

29. Travin MI, Duca MD, Kline GM, Herman SD, Demus DD, Heller GV. Relation of gender to physician use of test results and to the prognostic value of stress technetium 99m sestamibi myocardial single-photon emission computed tomography scintigraphy. *Am Heart J*. 1997;134:73–82.

30. Vaduganathan P, He ZX, Raghavan C, Mahmarian JJ, Verani MS. Detection of left anterior descending coronary artery stenosis in patients with left bundle branch block: exercise, adenosine or dobutamine imaging? *J Am Coll Cardiol*. 1996;28:543–550.

31. Wennike N, Shah BN, Boger E, Senior R, Greaves K. Stress echocardiography in the district hospital setting: a cost-saving analysis. *Eur J Echocardiogr*. 2010;11:401–405.

32. Vlachopoulos C, Rokkas K, Ioakeimidis N, Aggeli C, Michaelides A, Roussakis G, Fassoulakis C, Askitis A, Stefanadis C. Prevalence of asymptomatic coronary artery disease in men with vasculogenic erectile dysfunction: a prospective angiographic study. *Eur Urol*. 2005;48:996–1002; discussion 1002–1003.

33. Roeters van Lennep JE, Borm JJ, Zwinderman AH, Pauwels EK, Bruschke AV, van der Wall EE. No gender bias in referral for coronary angiography after myocardial perfusion scintigraphy with technetium-99m tetrofosmin. *J Nucl Cardiol*. 1999;6:596–604.

34. Kane GC, Askew JW, Chareonthaitawee P, Miller TD, Gibbons RJ. Hypertensive response with exercise does not increase the prevalence of abnormal Tc-99m SPECT stress perfusion images. *Am Heart J*. 2008;155:930–937.

35. Jang JY, Sohn IS, Kim JN, Park JH, Park CB, Jin ES, Cho JM, Kim CJ, Bae JH. Treadmill exercise stress echocardiography in patients with no history of coronary artery disease: a single-center experience in Korean population. *Korean Circ J*. 2011;41:528–534.

36. Hannoush H, Shaar K, Alam S, Nasrallah A, Sawaya J, Dakik HA. Analysis of referral patterns, predictive accuracy, and impact on patient management of myocardial perfusion imaging in a new nuclear cardiology laboratory. *J Nucl Cardiol*. 2003;10:148–153.

37. Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. Stress myocardial perfusion single-photon emission computed tomography is clinically effective and cost effective in risk stratification of patients with a high likelihood of coronary artery disease (CAD) but no known CAD. *J Am Coll Cardiol*. 2004;43:200–208.

38. Diamond JA, Makaryus AN, Sandler DA, Machac J, Henzlova MJ. Normal or near normal myocardial perfusion stress imaging in patients with severe coronary artery disease. *J Cardiovasc Med (Hagerstown)*. 2008;9:820–825.

39. Charvat J, Michalova K, Taborska K, Chlumsky J, Kvapil M. Comparison of the exercise ECG and stress myocardial SPECT in detection of the significant coronary artery disease in the asymptomatic patients with diabetes mellitus type 2. *Bratisl Lek Listy*. 2004;105:56–61.

40. Bucerius J, Joe AY, Herder E, Brockmann H, Biermann K, Palmedo H, Tiemann K, Biersack HJ. Hemodynamic variables during stress testing can predict referral to early catheterization but failed to show a prognostic impact on emerging cardiac events in patients aged 70 years and older undergoing exercise (99m)Tc-sestamibi myocardial perfusion scintigraphy. *Int J Cardiovasc Imaging*. 2009;25:569–579.

41. McClellan JR, Dugan TM, Heller GV. Patterns of use and clinical utility of exercise thallium-201 single photon emission-computed tomography in a community hospital. *Cardiology*. 1996;87:134–140.

42. Lucas FL, DeLorenzo MA, Siewers AE, Wennberg DE. Temporal trends in the utilization of diagnostic testing and treatments for cardiovascular disease in the United States, 1993–2001. *Circulation*. 2006;113:374–379.

43. Andrus BW, Welch HG. Medicare services provided by cardiologists in the United States: 1999–2008. *Circ Cardiovasc Qual Outcomes*. 2012;5:31–36.

44. Lauer MS. What is the best test for a patient with classic angina? *Cleve Clin J Med*. 2007;74:123–126.

45. Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT, Sehnert AJ, Yau M, Kraus WE, Newby LK, Schwartz RS, Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A, Winn ME, Schork NJ, Topol EJ. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Ann Intern Med*. 2010;153:425–434.

46. Thomas GS, Voros S, McPherson JA, Lansky AJ, Winn ME, Bateman TM, Elashoff MR, Lieu HD, Johnson AM, Daniels SE, Ladapo JA, Phelps CE, Douglas PS, Rosenberg S. A blood based gene expression test for obstructive coronary artery disease tested in symptomatic non-diabetic patients referred for myocardial perfusion imaging: the COMPASS study. *Circ Cardiovasc Genet*. 2013;6:154–162.

47. Metz LD, Beattie M, Hom R, Redberg RF, Grady D, Fleischmann KE. The prognostic value of normal exercise myocardial perfusion imaging and

exercise echocardiography: a meta-analysis. *J Am Coll Cardiol*. 2007;49: 227–237.

48. Rozanski A, Gransar H, Hayes SW, Min J, Friedman JD, Thomson LE, Berman DS. Temporal trends in the frequency of inducible myocardial ischemia during cardiac stress testing: 1991 to 2009. *J Am Coll Cardiol*. 2013;61:1054–1065.

49. Cronin AM, Vickers AJ. Statistical methods to correct for verification bias in diagnostic studies are inadequate when there are few false negatives: a simulation study. *BMC Med Res Methodol*. 2008;8:75.

50. Camici PG, Crea F. Coronary microvascular dysfunction. *N Engl J Med*. 2007;356: 830–840.

51. De Bruyne B, Pijls NH, Kalesan B, Barbato E, Tonino PA, Piroth Z, Jagic N, Mobius-Winckler S, Rioufol G, Witt N, Kala P, MacCarthy P, Engstrom T, Oldroyd KG, Mavromatis K, Manoharan G, Verlee P, Frobert O, Curzen N, Johnson JB, Juni P, Fearon WF. Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease. *N Engl J Med*. 2012;367: 991–1001.