

EDITORIAL

Evaluation of the Incremental Value of a Diagnostic Test: A Worthwhile Exercise in This Era of Cost Consciousness?

The clinical worth of a diagnostic test can be evaluated by determining its incremental value in the sequence in which it is performed during the assessment of a patient. For instance, if a 55-yr-old man with a history of hypertension and tobacco abuse presents with typical angina, what is the incremental value of exercise testing in this patient for the detection of coronary artery disease (CAD)? Is it the same as in a similarly aged woman with no risk factors and atypical chest pain?

The way to evaluate the incremental value of a test is to first identify the significant demographic and clinical variables that predict CAD (such as age, type of chest pain, risk factors, etc.) and derive a model based on these variables. Then the first test done in the assessment of such a patient should be evaluated. Let us assume that the first test is the electrocardiogram (EKG). The significant variables from the EKG that identify CAD should then be determined. The next step is to assess whether inclusion of the significant EKG variables adds to the already available significant demographic and clinical variables for the detection of CAD. In this manner, each successively performed test can be independently evaluated for the variables that aid in the detection of CAD and the incremental value of these variables to those already known can then be evaluated. Several statistical methods are used to compare models containing different levels of information, and the one used by Oosterhuis et al. in a paper in this issue of the *Journal* is the likelihood ratio statistic (1).

Evaluating the incremental value of tests is of major importance in clinical

medicine for several reasons. The first is the worth of a test in the detection of disease in a specific clinical situation. For instance, in the above-mentioned case of the 55-yr-old man with multiple risk factors and typical angina, the probability of CAD is high. Performance of a test merely for the detection of CAD may therefore not be very valuable in this patient. In contrast, it may provide more useful diagnostic information in a patient with a moderate likelihood of CAD. Since risk factors and symptoms do not reflect the extent of CAD, and since the extent of CAD has important bearing on the patient's management and prognosis, a test that reflects the extent of CAD may be useful even in the aforementioned middle-aged man with typical angina and multiple risk factors. Hence, risk-stratification is the second reason to evaluate the incremental value of a test.

The third reason is cost. For example, in a patient with known CAD, previous infarctions and heart failure who presents with exercise-induced angina, an increased lung-to-heart ratio and multiple perfusion defects on the initial images and redistribution on the delayed images are very likely to be present on exercise ^{201}Tl imaging. In this patient, how much did the performance of an expensive test really add to the already available information that indicated a high-risk for an adverse outcome? In contrast, resting ^{201}Tl imaging for the assessment of viability in asynergic myocardial segments may have provided valuable information relevant to expected outcome after coronary artery bypass surgery (2).

Because of these issues, Diamond has championed the cause of assessing the incremental value of tests performed in clinical practice (3) and has demonstrated the additive value (or lack thereof) of various tests done for

the evaluation of patients with known or suspected CAD (4). Lee et al. have utilized this approach for the evaluation of the added prognostic value of exercise radionuclide angiography in patients with CAD (5) and investigators at St. Louis University have used it for assessment of dipyridamole- ^{201}Tl imaging in elderly patients unable to exercise (6). Our group has employed the same approach for a variety of clinical conditions, ranging from chronic mitral regurgitation to chest pain syndromes in the emergency room to ambulatory patients with known or suspected CAD (7-10).

In their paper, Oosterhuis and colleagues have attempted to determine the additive value of exercise ^{201}Tl imaging in patients for both the detection and management of CAD (1). They found that while quantitative assessment of ^{201}Tl images added significantly to known clinical and exercise EKG variables for the detection of any CAD, visual assessment of these images did not. They also found that both quantitative and visual assessment added significantly to known clinical and exercise EKG variables for the detection of multivessel CAD and that ^{201}Tl imaging was not useful in determining medical versus surgical management in their patient cohort.

Several issues in their paper need to be addressed. First, in evaluating the role of ^{201}Tl imaging for the detection of any CAD, these authors excluded those patients who exhibited no angina or ST-segment depression in association with inadequate (<85% of maximal predicted) exercise heart rates. This group constitutes a large proportion of patients undergoing exercise testing and it is in this particular group that ^{201}Tl imaging offers the highest yield, since the results of exercise EKG are inconclusive. Had these patients been included in the

Received Jul. 14, 1992; accepted Jul. 14, 1992.
For reprints contact: Sanjiv Kaul, MD, Division of Cardiology, Box 158, University of Virginia, Charlottesville, VA 22908.

analysis, it is conceivable that the additive value of ^{201}Tl imaging would have been greater for the detection of any CAD. When we included all patients, irrespective of level of exercise, we found quantitative ^{201}Tl imaging to provide additive value both for the detection of CAD and for the determination of outcome (7,8).

Second, since these authors included in their analysis only those patients who had undergone both ^{201}Tl imaging and coronary angiography, the true value of ^{201}Tl imaging in patient management decisions cannot be determined. Patients with low-risk ^{201}Tl findings most likely never underwent coronary angiography, and in these patients ^{201}Tl imaging in association with exercise EKG was probably the major determinant of patient management. Certainly, once the extent of CAD is observed on angiography, the decision to perform revascularization is strongly influenced by coronary anatomy and left ventricular function. Whether a patient undergoes angiography, however, may be determined by the results obtained on ^{201}Tl imaging. In our practice, physicians rarely refer patients with normal or low-risk exercise ^{201}Tl findings for cardiac catheterization (11).

Third, the incidence of false-negative ^{201}Tl imaging in patients with a high pretest likelihood of disease is disturbing. In most instances, quantitative ^{201}Tl imaging provides a high sensitivity (12–14) for the detection of CAD, and false-negative results are usually associated with minimal CAD. Although the exact cause of false-negative results in the present study are difficult to ascertain, both the quantitative method used and the quality of images obtained may play a role. We would disagree with Oosterhuis and colleagues when they state that the low sensitivity in those with a high pretest likelihood of CAD “illustrates the general Bayesian rule . . . that the highest diagnostic yield is to be expected in cases with an equivocal, not a high or low, pretest probability for a particular outcome.” The reason that a test does not contribute

in a major way in instances where the pretest probabilities are low or high is that in these situations, as expected, the results are either negative or positive, respectively.

The utility of a test in clinical practice largely depends on the confidence that physicians have in the accuracy of the test in their own medical community. Myocardial ^{201}Tl imaging is both difficult to perform and interpret and can produce a high false-positive rate if interpreters are not cognizant of image artifacts (e.g., breast attenuation) or variants of normal (e.g., exaggerated apical thinning). If the scintigrams are “over-read” to maximize sensitivity at the expense of specificity, a high false-positive interpretation rate will prevail, making the test virtually useless with respect to decision making.

Almost all studies examining the prognostic utility of exercise- ^{201}Tl imaging have employed planar imaging techniques. No large prognostic studies have been reported using SPECT ^{201}Tl imaging, which is an order of magnitude more complex than planar imaging, with even a greater probability of artifacts than is associated with the latter. If not recognized, these artifacts can lead to sub-optimal specificity (15). Thus, the need for superior equipment and quantitative techniques coupled with excellent quality control and experienced, well-trained interpreters, is even greater for SPECT than planar images.

Fourth, these authors did not include quantitative lung-to-heart ratios for determining the incremental value of exercise- ^{201}Tl imaging to clinical and exercise EKG results. An increased lung-to-heart ^{201}Tl ratio correlates with the presence of multivessel CAD, poor left ventricular function, and exercise-induced ischemia (16) and has been found to be a major discriminator of risk in patients with CAD (17,18). Its inclusion would have strengthened the results of this study. We found that analysis of the lung-to-heart ^{201}Tl ratio almost doubled the prognostic value of exercise ^{201}Tl imaging and that adding vari-

ables from cardiac catheterization did not provide any incremental prognostic value. In contrast, catheterization variables provided additional prognostic information when ^{201}Tl variables without inclusion of the lung-to-heart ratio were analyzed (8).

Finally, for those not fully familiar with the interpretation of odds ratio and confidence intervals it is worth pointing out that for any odds ratio to be statistically significant, both the low and high confidence levels should be either above 1 or below 1, depending on whether the variable is positively or negatively correlated with disease. For instance, when one examines the value of age in Table 1, the age group 55–65 yr has an odds ratio of 5.5 with the 95% low and high confidence levels of 2.2 and 13.4, respectively. Since both the low and high levels are >1 , it means that this variable is significantly correlated with the occurrence of CAD on multivariate analysis. In contrast, the age group of ≥ 65 yr yields an odds ratio of 1.6 with the low and high confidence levels of 0.5 and 5.4, respectively. Since the low confidence level is <1 and the high is >1 , this odds ratio is not statistically significant. There are four such variables in Table 1 alone. It is not clear from this paper whether these nonsignificant variables were included in their model. As a general rule, nonsignificant variables should not be included from any data category when considering the additive value of test results.

In conclusion, too many tests are often ordered for the assessment of clinical problems, such as chest pain, a number of which if not harmful, are at least unnecessary. One approach to determining which test is worth performing is to assess its incremental value in the evaluation of a particular disease in the hierarchical order in which it is performed. In their paper, Oosterhuis and colleagues have attempted to do so; however, the limitations of their study have to be considered prior to reaching any conclusions. A critical issue not addressed by these and previous investigators,

which requires our immediate concern, is that even if a test has incremental value for the detection of disease and for risk-stratification, what absolute amount of increment makes such a test worthwhile to perform given its particular cost. In this era of cost-containment, it is crucial that we as physicians make these determinations before government agencies and third-party payers make them for us.

Sanjiv Kaul
George A. Beller
University of Virginia
Charlottesville, Virginia

REFERENCES

1. Oosterhuis WP, Niemeyer MG, Kuijper AFM, et al. Evaluation of the incremental diagnostic value and impact on patient treatment of thallium scintigraphy. *J Nucl Med* 1992;33:1727-1734.
2. Ragosta M, Beller GA, Watson DD, Kaul S, Gimble LW. Can resting thallium-201 imaging predict improvement in function after revascularization in patients with severely reduced LV function? [Abstract]. *Circulation* 1991;84(suppl II):II-89.
3. Diamond GA. Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. *J Am Coll Cardiol* 1989;14:12A-22A.
4. Ladenheim ML, Kotler TS, Pollock BH, Berman DS, Diamond GS. Incremental prognostic value of history, exercise electrocardiography, and myocardial perfusion scintigraphy in suspected coronary artery disease. *Am J Cardiol* 1987;59:270-277.
5. Lee KL, Pryor DB, Pieper KS, et al. Prognostic value of radionuclide angiography in medically treated patients with coronary artery disease: a comparison with clinical and catheterization variables. *Circulation* 1990;82:1705-1717.
6. Shaw L, Chaitman BR, Hilton TC, et al. Prognostic value of dipyridamole thallium-201 imaging in elderly patients. *J Am Coll Cardiol* 1992;19:1390-1398.
7. Pollock SG, Abbott RD, Boucher CA, Watson DD, Kaul S. A model to predict multivessel coronary artery disease from the exercise thallium-201 stress test. *Am J Med* 1991;90:345-352.
8. Pollock SG, Abbott RD, Boucher CA, Beller GA, Kaul S. Independent and incremental prognostic value of tests performed in hierarchical order to evaluate patients with suspected coronary artery disease. Validation on models based on the tests. *Circulation* 1992;85:237-248.
9. Reed D, Abbott RD, Smucker ML, Kaul S. Prediction of outcome following mitral valve replacement in patients with symptomatic mitral regurgitation: the importance of left atrial size. *Circulation* 1991;84:23-34.
10. Sabia P, Abbott RD, Afrookteh A, Keller MW, Touchstone DA, Kaul S. The importance of two-dimensional echocardiographic assessment of left ventricular systolic function in patients presenting to the emergency room with cardiac-related symptoms. *Circulation* 1991;84:1615-1624.
11. Desmaris RL, Kaul S, Watson DD, Beller GA. 'False positive' thallium-201 scans are due to easily discernable artifacts and do not lead to unnecessary coronary angiography [Abstract]. *J Am Coll Cardiol* 1992;24:8A.
12. Berger BB, Watson DD, Taylor GL, et al. Quantitative thallium-201 exercise scintigraphy for the detection of coronary artery disease. *J Nucl Med* 1981;22:585-598.
13. Kaul S, Boucher CA, Newell JB, et al. Determination of the quantitative variables that optimize the detection of coronary artery disease. *J Am Coll Cardiol* 1986;7:527-537.
14. Maddahi J, Garcia EV, Berman DS, Waxman A, Swan HJC, Forrester J. Improved noninvasive assessment of coronary artery disease by quantitative analysis of regional stress myocardial distribution of washout of thallium-201. *Circulation* 1981;64:924-935.
15. Garcia E, DePuey EG, Brown M, et al. Quantification of rotational thallium-201 myocardial tomograms: a multicenter trial using bullseye polar maps and standard normal limits [Abstract]. *J Nucl Med* 1987;28:673.
16. Homma S, Kaul S, Boucher CA. Correlates of lung/heart ratio of thallium-201 in coronary artery disease. *J Nucl Med* 1987;28:1531-1535.
17. Kaul S, Finkelstein DM, Homma S, Leavitt M, Okada RD, Boucher CA. Superiority of quantitative exercise thallium-201 variables in determining long-term prognosis in ambulatory patients with chest pain: a comparison with cardiac catheterization. *J Am Coll Cardiol* 1988;12:25-34.
18. Miller DD, Kaul S, Strauss HW, Okada RD, Boucher CA. Increased exercise thallium-201 lung uptake: a noninvasive prognostic index in two-vessel coronary artery disease. *Can J Cardiol* 1988;4:270-276.