Are Human Readers Needed for Prognostication from Stress Myocardial Perfusion SPECT? Using Outcomes Research to Validate Medical Technology

Over the past decade and a half, 2 opposing forces have irrevocably altered the landscape of medicine. On the one hand, ongoing pressure has been placed on physicians to curtail health-care costs by both reducing resource use and enhancing the efficiency of the practice of medicine. Simultaneously, there is an accelerating evolution of technology that has yielded numerous new technologies as well as improvements of older technologies. Our field has obviously benefited from the latter, as evidenced by new isotopes, stress agents, hardware, and software (e.g., attenuation correction, hybrid SPECT/CT, and PET/CT) that have gained widespread acceptance, as well as from more recent developments that have similar potential. Interestingly, the economic pressures on our field have resulted in unexpected benefits. To achieve the cost savings necessitated by these pressures without compromising the quality of patient care, the field of outcomes research gained widespread attention and application. The challenge we faced was to develop a body of evidence validating the use of our technology— hence, (hopefully) ensuring reimbursement from payers. Indeed, numerous studies, focusing on pragmatic endpoints such as adverse outcomes and cost, identified when our test achieved clinical effectiveness and cost-effectiveness.

PROGNOSTIC "VALIDATION" OF STRESS MYOCARDIAL PERFUSION SCINTIGRAPHY (MPS)

Since the first published report of the prognostic value of stress planar imaging (*1*), the number, size, and quality of the published outcomes literature in nuclear cardiology have grown considerably. Importantly, the questions asked in these studies have evolved as well. For example, to justify the use of stress MPS it was necessary to show that the information provided by MPS results added significant information over and above what we already knew about the patient from prior evaluation (clinical, historical, and exercise tolerance test [ETT] data) (*2*). To ensure that this information could be freely translated into clinically useful data, this incrementalistic approach was incorporated into risk stratification schemas by demonstrating that the results of stress MPS risk stratify patients after they had already been risk stratified into low, intermediate, and high clinical risk groups (enhanced stratification) (*3*). These results have been confirmed in numerous cohorts, including women, minorities, diabetics, patients with and without prior coronary artery disease (CAD), postrevascularization, and the elderly (*4*). Further, MPS was found to be a cost-effective approach to patient evaluation when appropriate patients were selected for testing (intermediate-to-high post-ETT likelihood of CAD in patients with rest electrocardiogram in-terpretable for ETT, intermediate-to-high pre-ETT likelihood of CAD in patients with rest ECG uninterpretable for ETT) (*5,6*). Similarly, MPS was a clinically equivalent and cost-saving strategy compared with catheterization for the evaluation of stable patients presenting with chest pain (*7*).

Although no randomized trials to date have been based on MPS data, the ability of stress MPS results to assist in the selection of post-MPS patient management has been investigated. Risk stratification using stress MPS could be further refined by identifying patients with extensive abnormalities (>10% of the total myocardium abnormal) as being at significant risk for cardiac death (an endpoint that can be prevented by revascularization) as opposed to those patients with limited defects (<10% of the total myocardium abnormal), who are primarily at risk for myocardial infarction (an endpoint prevented by medical therapy, but not by revascularization) (*8*). Identification of a survival benefit with revascularization versus medical therapy as a function of MPS results was examined by a recent study of 10,627 patients without prior CAD whose post-MPS outcomes were assessed as a function of treatment given (*9*). The presence of extensive inducible ischemia (>10%– 12.5% of the total myocardium ischemic) as measured by MPS identified patients who had enhanced relative and absolute benefit with revascularization compared with medical therapy. In the absence of ischemia, irrespective of scan abnormality, no such benefit was present.

EXTENDING THE PRINCIPLE: OUTCOMES-BASED VALIDATION OF TECHNOLOGY

Although the use of outcomes research as a means to enhance the delivery of healthcare is a hypothesis rather than a proven entity, the popularity and success of an outcomes-based approach to justifying the use of MPS naturally extended these methods to investigating specific technologic aspects of this modality. These efforts have included comparison of outcomes with different isotopes (^{99m}Tc-sestamibi vs. ^{99m}Tc-tetrofosmin), protocols (^{201}Tl vs. dual isotope) and methods (viability imaging to improve prediction of improvement in LV function). In addition, a natural area of investigation for these methods would be the relative value of quantitative software in clinical practice.

INCREMENTAL PROGNOSTIC VALUE OF AUTOMATED QUANTIFICATION OF SPECT PERFUSION SOFTWARE

One of the strengths of stress perfusion imaging with nuclear techniques is the ability to apply automatic quantitative analysis software to assist in the interpretation of patient studies. It has been hypothesized that the use of these packages would generally enhance the quality and consistency of scan interpretation given the wide variability in reader expertise. The prognostic value of quantitative software was first compared with expert interpretation by Berman et al. (*10*) in a study of 984 patients who had 28 hard events (cardiac death or nonfatal myocardial infarction) over a 20-mo follow-up period. In this study, all patients had MPS interpretation by expert visual reading using a 5-point, 20-segment scoring system as well as results generated by automatic quantitative analysis software. This study reported that expert visual interpretation and quantitative scores yielded similar incremental value over clinical and historical data for prediction of adverse events. Further, risk was low with scans considered normal by either approach, and risk increased as a function of scan abnormality defined by either approach. Hence, at first glance, automatic quantitative analysis software appeared to give results similar to those of expert interpretation and, as suggested by the authors, could serve a potentially important role in "laboratories with less experienced visual interpreters." (*10*).

A closer examination of these results reveals several subtle but important findings. First, although correlations between corresponding visual and quantitative parameters identify a close association, quantitative software parameters could explain only 38%–58% of the variability in visual interpretations. Further, although both visual and quantitative parameters achieved risk stratification and normal scans (as defined by either metric) were associated with low risk, visual interpretation identified 61%–71% of all patients as normal, whereas quantitative parameters identified 50%–52% of patients as abnormal. Thus, although, as the authors stated, these metrics are similar, they are far from interchangeable.

The study by Leslie et al. (*11*), on pages 204–211 of this issue of *The Journal of Nuclear Medicine*, further extends the findings of this previous article by examining a cohort of 718 patients who underwent predominantly 2-d ^{99m}Tc exercise or pharmacologic stress protocols and were followed-up at a mean of 5.6 ± 1.1 y. The current study compared visual interpretation from 2 expert nuclear medicine physicians (who categorized results as normal, equivocal, abnormal with fixed defects, abnormal with fully reversible defects, or abnormal with partially reversible defects). Quantitative data were generated from retrospective reprocessing of studies using Cedars-Sinai software, thus generating quantitative 5-point, 20-segment scores, including summed stress, rest, and difference scores.

In this study, similar risk stratification was achieved by a normal versus abnormal categorization (the numbers of patients defined as normal and abnormal by the 2 metrics were not presented). Interestingly, the 145 patients with discordant results had an intermediate risk of adverse events. With respect to the quantitative parameters, increases in the extent and severity of stress defects or reversibility was associated with progressive increases in risk. With respect to incremental value, the addition of quantitative results expressed as a dichotomous variable (normal vs. abnormal) failed to add prognostic value over visual interpretation. Interestingly, the addition of quantitative parameters as continuous variables added significant information over dichotomous visual interpretation (normal vs. abnormal) for prediction of adverse events.

The results of Leslie et al. (*11*) both extend our understanding of the role of quantitative software in clinical practice and raise new questions. On one hand, these results suggest that if a laboratory currently scores and reports results using a categoric metric (dichotomous or otherwise, as in the current study), quantitative software will not be prognostically helpful if considered in a dichotomous fashion. On the other hand, significant additional prognostic information can be gained by the incorporation of quantitative results to report information regarding the extent and severity of perfusion defects.

This is important for several reasons. First, while the incremental prognostic value of extent and severity data over each other has long been established (*12*), the findings by Leslie et al. (*11*) extend these results in that quantitative extent and severity data are superior to visual categoric data alone. Indeed, it would have been of interest to determine whether quantitative data measuring extent or severity would have also been incremental over visual categoric data alone. Second, with respect to the clinical utility of stress MPS, it has been recognized that the extent and severity of inducible ischemia is the most powerful predictor of referral to post-MPS catheterization and revascularization by referring physicians (*3,9*) and the means by which patients who may benefit from revas-

cularization may be identified (*9*). By demonstrating the prognostic validity of quantitative software in their cohort, the authors have opened the door for other centers with similar patient profiles to also use this approach as a means to report to their referring physicians an estimate of ischemia extent and severity. Hopefully, this would permit enhanced identification by referring physicians of optimal candidates for the catheterization laboratory. Finally, this study, as was the case with the study by the Cedars-Sinai group, does not tell us whether quantitative extent and severity data yield information over visual extent and severity data. Consequently, it is unclear if laboratories willing to incorporate extent and severity into visual scoring and subsequent reporting will benefit from quantitative software. However, as this approach is undoubtedly the exception rather than the rule, many laboratories may benefit from these results and the use of quantitative software to enhance daily reporting.

METHODOLOGIC ISSUES: LIMITATIONS OF OUTCOMES-BASED VALIDATION

Although numerous methodologic issues are raised by this study, we will touch on several of the most important. Issues of generalizability must be raised with respect to whether these results are equally applicable to all quantitative software packages, to the results of all interpreters' visual analysis and to all populations examined. Variations in the methods of interpretation or reporting can negate the value of quantitative software results.

Also, it is important to point out that the statistical approach used by the current study as well as the previous study from Cedars-Sinai limits the conclusions that can be drawn with respect to the clinical value of quantitative software results. By and large, analyses of incremental value follow 1 of 2 approaches. One is a more general approach that describes the performance characteristics of a test in the overall cohort examined. This includes analyses describing an increase in χ^2

with the addition of test data to pretest information, as is the case with statistical approaches to incremental value (general tests of association)—similarly, assessing risk stratification in an overall cohort defining event rates in the setting of normal and abnormal scans as seen in various populations can be used as well. An alternative approach to these general assessments is a more patient-based approach that defines the value of a test on the basis of a change in individual patient risk. This includes, for example, studies that examine how many patients are reclassified with respect to their likelihood or risk of an outcome by the addition of the results a test (*6,13*). Similarly, assessment of test performance based on estimates of individual patient risk uses a similar approach (*14*). Although both a general approach and a patient-centered approach are important for assessing test performance characteristics, the latter would be useful to better assess the incremental value of quantitative software with respect to the impact on individual patients.

WHAT IS THE FUTURE ROLE OF AUTOMATIC QUANTITATIVE SOFTWARE PACKAGES?

The results of the study by Leslie et al. (*11*) potentially extend the use of quantitative software to daily clinical use, if the methodologic issues outlined above can be addressed. What, then, is the future for MPS software? The American Society of Nuclear Cardiology guidelines (*15*) have long advocated the incorporation of statements of risk on reporting. A major challenge, however, is how to go about fulfilling this potentially important role? It is misleading and potentially erroneous to merely use reported event rates to estimate this risk, as the patients reported and those seen at individual sites may vary significantly. Of equal concern, it has been shown that numerous patient characteristics add incremental prognostic value over MPS data. Hence, the risk associated with any particular defect may vary widely as a function of patient age, sex, presence of diabetes mellitus, type

of stress performed, and so forth. For this reason, valid reporting of statements of risk necessitates the use of validated scores that incorporate multiple patient data and weight them appropriately. This approach is best exemplified by the Duke treadmill score (*16*), a validated instrument incorporating 3 variables from ETT that permits estimation of post-ETT patient risk. This approach has been extended to estimation of risk in patients undergoing adenosine stress dual-isotope SPECT (*14*). This study presents several scores that permit estimation of 2-y cardiac mortality on the basis of clinical (age, rest electrocardiogram), stress (rest and peak stress heart rate), and MPS data (percent myocardium ischemic, percent myocardium fixed). Further, separate scores, thus prognostic estimates, are generated for medical therapy and revascularization as treatment strategies. If perfusion variables are incorporated from quantitative software, and other patient data are entered, these estimates can be made in the future entirely from automatic quantitative software.

CONCLUSION

The study by Leslie et al. (*11*) in this issue of the Journal furthers our understanding of the potential role of automatic quantitative software in the interpretation and reporting of stress MPS results. If these results can be validated in other centers, with other readers and varying patient cohorts, it would suggest that, by providing stress and rest measures of defect extent and severity, automatic quantitative software can provide interpreting physicians prognostically important and clinically useful information that can be incorporated into their daily reports to referring physicians. Many questions remain, however, including which visual methods can be augmented by this approach, whether all software packages add similar information, and whether future software will provide additional, algorithm-based estimates of CAD likelihood or risk of adverse outcome. Nonetheless,

there is little doubt that the initial steps in this area have been taken and we can look forward to hearing much more from this area of investigation.

Rory Hachamovitch, MD
Keck School of Medicine, University of Southern California
Los Angeles, California

Marcelo F. Di Carli, MD
Brigham and Women's Hospital and Harvard Medical School
Boston, Massachusetts

REFERENCES

1. Brown KA, Boucher CA, Okada RD, et al. Prognostic value of exercise thallium-201 imaging in patients presenting for evaluation of chest pain. *J Am Coll Cardiol.* 1983;1:994–1001.

2. Ladenheim ML, Kotler TS, Pollock BH, Berman DS, Diamond GA. Incremental prognostic power of clinical history, exercise electrocardiography and myocardial perfusion scintigraphy in suspected coronary artery disease. *Am J Cardiol.* 1987;59:270–277.

3. Hachamovitch R, Berman DS, Kiat H, et al. Exercise myocardial perfusion SPECT in patients without known coronary artery disease: incremental prognostic value and use in risk stratification. *Circulation.* 1996;93:905–914.

4. Berman DS, Hachamovitch R, Shaw LJ, Germano G, Hayes S. Nuclear cardiology. In: Fuster V, Alexander RW, O'Rourke RA, eds. *Hurst's The Heart.* New York, NY: McGraw-Hill; 2004:563–598.

5. Berman DS, Hachamovitch R, Kiat H, et al. Incremental value of prognostic testing in patients with known or suspected ischemic heart disease: a basis for optimal utilization of exercise technetium-99m sestamibi myocardial perfusion single-photon emission computed tomography [published erratum appears in *J Am Coll Cardiol.* 1996;27:756]. *J Am Coll Cardiol.* 1995;26:639–647.

6. Hachamovitch R, Berman DS, Kiat H, Cohen I, Friedman JD, Shaw LJ. Value of stress myocardial perfusion single photon emission computed tomography in patients with normal resting electrocardiograms: an evaluation of incremental prognostic value and cost-effectiveness. *Circulation.* 2002;105:823–829.

7. Shaw LJ, Hachamovitch R, Berman DS, et al. The economic consequences of available diagnostic and prognostic strategies for the evaluation of stable angina patients: an observational assessment of the value of precatheterization ischemia—Economics of Noninvasive Diagnosis (END) Multicenter Study Group. *J Am Coll Cardiol.* 1999;33:661–669.

8. Hachamovitch R, Berman DS, Shaw LJ, et al. Incremental prognostic value of myocardial perfusion single photon emission computed tomography for the prediction of cardiac death: differential stratification for risk of cardiac death and myocardial infarction. *Circulation.* 1998;97:535–543.

9. Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. Comparison of the short-term survival benefit associated with revascularization compared with medical therapy in patients with no prior coronary artery disease undergoing stress myocardial perfusion single photon emission computed tomography. *Circulation.* 2003;107:2900–2907.

10. Berman DS, Kang X, Van Train KF, et al. Comparative prognostic value of automatic quantitative analysis versus semiquantitative visual analysis of exercise myocardial perfusion single-photon emission computed tomography. *J Am Coll Cardiol.* 1998;32:1987–1995.

11. Leslie WB, Tully SA, Yogendran MS, Ward LM, Nour KA, Metge CJ. Prognostic value of automated quantification of ^{99m}Tc-sestamibi myocardial perfusion imaging. *J Nucl Med.* 2005;46:204–211.

12. Ladenheim ML, Pollock BH, Rozanski A, et al. Extent and severity of myocardial hypoperfusion as predictors of prognosis in patients with suspected coronary artery disease. *J Am Coll Cardiol.* 1986;7:464–471.

13. Christian TF, Miller TD, Bailey KR, Gibbons RJ. Exercise tomographic thallium-201 imaging in patients with severe coronary artery disease and normal electrocardiograms. *Ann Intern Med.* 1994;121:825–832.

14. Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. A prognostic score for prediction of cardiac mortality risk after adenosine stress myocardial perfusion scintigraphy. *J Am Coll Cardiol.* In press.

15. Imaging guidelines for nuclear cardiology procedures. Part 2. American Society of Nuclear Cardiology. *J Nucl Cardiol.* 1999;6:G47–G84.

16. Mark DB, Hlatky MA, Harrell FE Jr, Lee KL, Califf RM, Pryor DB. Exercise treadmill score for predicting prognosis in coronary artery disease. *Ann Intern Med.* 1987;106:793–800.