

## The price for probability: Comparing the costs of diagnostic testing strategies

Prem Soman, MD, PhD, MRCP(UK),<sup>a</sup> and James E. Udelson, MD<sup>b</sup>

---

**See related article, p. 174**

---

Despite the universal acceptance of the concept of evidence-based medicine, many practices in medicine are widely embraced and implemented without a thorough assessment of both efficacy (how a test or procedure performs in a controlled setting) and effectiveness (performance in more real-life settings). This is perhaps more true of diagnostic tests than of therapeutic interventions, because the latter are more closely regulated by national regulatory agencies. It has been suggested that the use of a diagnostic test in medicine should be evaluated at the following hierarchic levels<sup>1</sup>: (1) technical efficacy, (2) diagnostic accuracy efficacy, (3) diagnostic thinking efficacy, (4) therapeutic planning efficacy, (5) patient outcome efficacy, and (6) societal efficacy. Efficacy at higher levels is contingent upon efficacy at lower levels, but the converse is not a requirement.

Technical efficacy, which refers to the development and refinement of a diagnostic modality before its clinical implementation, and diagnostic accuracy efficacy, measured as sensitivity, specificity, and positive and negative predictive values, are usually well established before broader clinical implementation. Diagnostic thinking efficacy is a measure of the effect of a test result on the thinking of the physician and is often used as a surrogate measure of the relationship between test result and patient outcome, because the latter is almost always difficult to establish. Therapeutic planning efficacy refers to the initiation of appropriate therapy, or the determination that no therapy is necessary, based on the results of a diagnostic test. Patient outcome efficacy is a determination of whether the test results eventually translate into improved outcome. This is often challenging to

clearly establish, as the downstream clinical outcome is often removed by several steps of clinical decision making and other influential variables from the upstream test result. Better diagnostic accuracy efficacy may not always translate into better patient outcome efficacy. A case in point might be the study by Siebelink et al,<sup>2</sup> in which outcome efficiency was compared in patients with ischemic cardiomyopathy treated either by revascularization or medical therapy, based on the presence of residual myocardial viability. Patients were prospectively randomized to have the management decision based either on a single photon emission computed tomography (SPECT) or positron emission tomography (PET) study. Although PET imaging is generally considered to be more sensitive than SPECT for viability detection, this opinion is based on the number of viable segments detected by these modalities in comparative studies.<sup>3</sup> When patient outcome, rather than the extent of viability detected, was assessed in the Siebelink trial, there was equivalent efficacy of these imaging modalities in terms of influencing patient outcome.

The final and highest level of efficacy measurement, societal efficacy, is the true measure of cost-effectiveness—that is, whether the overall benefit to society is worth the costs associated with the test. These latter levels of efficacy, levels 4 through 6, are not always fully ascertained before a test is clinically implemented, because they are extremely challenging to evaluate, as nicely illustrated in the study by Sabharwal et al<sup>4</sup> in this issue of the *Journal of Nuclear Cardiology*.

The authors should truly be complimented for applying the rigor of a prospective, randomized clinical trial (RCT) in a single center to evaluate the diagnostic and cost implications of a widely accepted clinical practice paradigm, namely, the use of myocardial perfusion imaging (MPI) as the initial diagnostic step in patients with suspected coronary artery disease (CAD). Although a considerable body of literature does exist in regard to the cost-effectiveness of MPI for CAD diagnosis, much of the data are observational or derived from multicenter database compilations comparing different diagnostic strategies in different groups of patients studied in different eras.<sup>5</sup> Moreover, comparisons of the diagnostic performance of MPI versus exercise electrocardiography (ECG) are often based on comparisons of data from distinct populations<sup>6,7</sup> or are performed by

From the Division of Cardiology, University of Pittsburgh Medical Center, Pittsburgh, Pa,<sup>a</sup> and Division of Cardiology, Tufts-New England Medical Center, Boston, Mass.<sup>b</sup>

Reprint requests: Prem Soman, MD, PhD, MRCP(UK), Presbyterian Hospital, 200 Lothrop St, Room A 429, Scaife Hall, Pittsburgh, PA 15238; [somanp@upmc.edu](mailto:somanp@upmc.edu).

*J Nucl Cardiol* 2007;14:142-4.

1071-3581/\$32.00

Copyright © 2007 by the American Society of Nuclear Cardiology.

doi:10.1016/j.nuclcard.2007.01.038

comparing the diagnostic or prognostic value of these two modalities when performed in the same patients.<sup>8,9</sup> This latter approach is surely biased to some degree by the referral of patients in such studies to an MPI study, from which the stress ECG data are then analyzed separately.

The interesting and well-conceived study by Sabharwal et al<sup>4</sup> prospectively randomized 503 patients with stable chest pain to an initial diagnostic strategy of either exercise ECG or stress MPI. Surprisingly, even at this point in our large collective experience with MPI, this approach is quite novel and unique in the exercise ECG and SPECT literature and is a significant strength of the study. The main endpoint was “cost to diagnosis,” and subsequent testing after the initial strategy was performed in patients who remained in an intermediate-likelihood category for CAD (ie, those who were not definitively stratified by the initial test as having a low or high probability of CAD).

At first look, many of the results were expected, and they can be summarized as follows:

1. MPI produced more definitive test results as indicated by the smaller number of patients who remained in the intermediate-likelihood category after initial testing.
2. MPI resulted in a reduction in the requirement for subsequent testing (somewhat by study design).
3. MPI could be construed as resulting in more effective utilization of coronary angiography as indicated by the lower rate of normal coronary angiograms in patients who had MPI as the initial diagnostic strategy.

However, despite these findings, and somewhat counterintuitively, the analysis of cost to diagnosis did not show an advantage of the initial MPI strategy, either in the overall population or in any of the pretest likelihood subgroups. In fact, in the group with a low pretest likelihood, exercise ECG as the initial strategy was associated with a lower cost to diagnosis. This seeming paradox illustrates not only the challenges associated with efficacy or effectiveness testing of diagnostic strategies in medicine, but also that assumptions based simply on test performance efficacy or on data from observational studies do not always play out as expected when subjected to scrutiny in a randomized trial format.

There are some issues to consider in the interpretation of the results of this study. First, the assignment of cost to a test can be approached in many different ways. In this single-center study the authors used costs that were specific to their institution, and they included all costs that were incurred up to the point of diagnosis. However, applying average national costs derived from the National Institute for Health and Clinical Excellence database produced different results, in that MPI was less

expensive than exercise ECG in the groups with an intermediate or high pretest probability. This raises questions on the appropriateness of generalizing cost data generated from a single-center study even in a single-payer health care system such as the National Health Service of the United Kingdom. Cost considerations are more complex in the US health care system, where multiple vested parties including the government, payers, hospitals, and physician practices have different perspectives on health care costs. Nonetheless, the data and patient flow from this study can be used by investigators to model costs within their own systems.

Second, what was tested by Sabharwal et al<sup>4</sup> is cost to diagnosis and not cost-effectiveness. A study by Patterson et al<sup>10</sup> clearly illustrates the difference. These authors retrospectively analyzed published data on test accuracy and complication rates to compare the cost of using exercise ECG, stress MPI, PET, or invasive coronary angiography as the initial strategy for the diagnosis of CAD. In the cost analysis, indirect test costs arising from complications and death associated with each test, the cost associated with subsequent testing generated by results of the initial test, and the estimated cost and mortality rate associated with missing a diagnosis of CAD because of false-negative tests were also added to the direct cost of the test. Cost per effect (where *effect* was defined as the “ability to accurately diagnose CAD”) and cost per utility (where *utility* was defined as the increment in the number of quality-adjusted life-years over a 10-year period) were calculated in patients stratified based on pretest CAD probability. The results showed that the direct cost is only a fraction of the total costs incurred for using a particular test in a diagnostic algorithm. More importantly, using direct cost versus measures of cost per effect or cost per utility produced dramatically different conclusions regarding the clinical utility of these tests. For example, whereas direct costs increased linearly with pretest probability of CAD for all of the noninvasive tests (high-risk noninvasive test results are more often associated with further [invasive] testing and therapeutic interventions, whereas low-risk test results in high-probability patients are associated with the cost of a missed CAD diagnosis), the cost per effect and cost per utility had a “hyperbolic” relationship such that these parameters decreased as pretest probability increased. Thus, whereas absolute costs and mortality rates increased with increasing pretest probability of CAD, the cost per effect and cost per utility lessened. Therefore using cost-effectiveness parameters provides a more comprehensive indication of the clinical utility of a test compared with a cost-to-diagnosis assessment alone. The study by Sabharwal et al<sup>4</sup> is concordant with the analysis of Patterson et al in that exercise ECG is least costly in patients with a low probability of CAD.

However, as the pretest probability of CAD increases to the intermediate range, MPI was more cost-effective in the analysis of Patterson et al, despite greater absolute costs.

Third, although the RCT is considered the most rigorous level of clinical investigation, the specific inclusion and exclusion criteria generally applied in RCTs may result in study populations that are somewhat unrepresentative of real life. In this study patients with an abnormal baseline electrocardiogram were excluded, whereas in reality, these patients would usually be directly referred to MPI.<sup>7</sup> Similarly, many patients referred for stress testing are deemed unable to perform meaningful treadmill exercise, and they undergo pharmacologic stress testing with MPI as the initial strategy. In the present study patients could be randomized without regard to their potential ability to perform adequate exercise. One could argue that this is the full expression of an effectiveness trial (ie, randomizing all comers). On the other hand, inclusion of these patients in the exercise ECG arm of this study, as most had indeterminate exercise ECG results almost by definition and subsequently moved on to MPI as the second step, resulted in a bias against exercise ECG. It is of interest that even despite this bias, in no patient subset was exercise ECG more costly as the initial strategy.

Ultimately, it would be of great interest if follow-up studies established actual patient outcomes from this data set so that a true cost-effectiveness analysis could be performed.

Finally, although we live in a cost-conscious era where the curtailing of spiraling health care costs is (rightly) considered a priority, it must be remembered that cost is not the only factor that a physician has to consider when making management decisions for an individual patient. Other considerations might include the rapidity with which a diagnosis is established, the number of visits to a hospital or outpatient facility for testing, the physician's confidence in results from a particular modality, and the certainty of diagnosis that is afforded by the test.

This unique study by Sabharwal et al<sup>4</sup> illustrates that cost analysis of diagnostic strategies is complex, and applying the results to patient care even more so. Nonetheless, this study, performed in a single center, illustrates that rigorous evaluation of diagnostic testing strategies is quite feasible, and the approach taken by

these investigators can inform the development and potentially wider deployment of newer testing modalities such as computed tomographic coronary angiography or PET.

## References

1. Gazelle GS, McMahon PM, Siebert U, Beinfeld MT. Cost-effectiveness analysis in the assessment of diagnostic imaging technologies. *Radiology* 2005;235:361-70.
2. Siebelink HM, Blanksma PK, Crijns HJ, Bax JJ, van Boven AJ, Kingma T, et al. No difference in cardiac event-free survival between positron emission tomography-guided and single-photon emission computed tomography-guided patient management: a prospective, randomized comparison of patients with suspicion of jeopardized myocardium. *J Am Coll Cardiol* 2001;37:81-8.
3. Bax JJ, Wijns W, Cornel JH, Visser FC, Boersma E, Fioretti PM. Accuracy of currently available techniques for prediction of functional recovery after revascularization in patients with left ventricular dysfunction due to chronic coronary artery disease: comparison of pooled data. *J Am Coll Cardiol* 1997;30:1451-60.
4. Sabharwal NK, Stoykova B, Taneja AK, Lahiri A. A randomized trial of exercise treadmill ECG versus stress SPECT myocardial perfusion imaging as an initial diagnostic strategy in stable patients with chest pain and suspected CAD: cost analysis. *J Nucl Cardiol* 2007;14:174-86.
5. Des Prez R, Gillespie R, Jaber W, Noble G, Shaw L, Soman P, et al. Cost-effectiveness of myocardial perfusion imaging: a summary of the currently available literature. *J Nucl Cardiol* 2005;12:750-9.
6. Gianrossi R, Detrano R, Mulvihill D, Lehmann K, Dubach, Colombo A, et al. Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. *Circulation* 1989;80: 87-98.
7. Klocke FJ, Baird MG, Lorell BH, Bateman TM, Messer JV, Berman DS, et al. ACC/AHA/ASNC guidelines for the clinical use of cardiac radionuclide imaging—executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/ASNC Committee to Revise the 1995 Guidelines for the Clinical Use of Cardiac Radionuclide Imaging). *Circulation* 2003;108:1404-18.
8. Hachamovitch R, Berman DS, Kiat H, Cohen I, Cabico JA, Friedman J, et al. Exercise myocardial perfusion SPECT in patients without known coronary artery disease: incremental prognostic value and use in risk stratification. *Circulation* 1996;93:905-14.
9. Gibbons RJ, Hodge DO, Berman DS, Akinboboye OO, Heo J, Hachamovitch R, et al. Long-term outcome of patients with intermediate-risk exercise electrocardiograms who do not have myocardial perfusion defects on radionuclide imaging. *Circulation* 1999;100:2140-5.
10. Patterson RE, Eisner RL, Horowitz SF. Comparison of cost-effectiveness and utility of exercise ECG, single photon emission computed tomography, positron emission tomography, and coronary angiography for diagnosis of coronary artery disease. *Circulation* 1995;91:54-65.