EDITORIAL

Assessment of the Information Boondoggle Resulting from the Evaluation of Noninvasive Stress Tests in Cardiology

Increasingly, the practicing clinician is exposed to a veritable information boondoggle regarding the use of noninvasive stress tests in the evaluation of myocardial ischemia. As illustrated in Table 1, there are now a plethora of noninvasive stress tests, each with its own technical considerations. Investigators often differ in terms of testing goals, selection of stressor(s), criteria for test abnormality and statistical analyses. Biases associated with the selection of the study population and the referent standard for normality are additional factors which may compound the interpretation of test results. As a consequence, discerning clinicians must become increasingly sophisticated in stress-testing technology, the complex pathophysiology associated with myocardial ischemia (which these tests are designed to measure) and important principles of testing which may not have been taught during many physicians' clinical training.

When a new test or noninvasive test variable is evaluated in the literature, practicing clinicians' concerns can be reduced to two basic questions: (1) How accurate is the noninvasive test or variable being proposed, and (2) How relevant is the test for the patient population seen in my clinical practice? The study by Wu et al. (*1*) was evaluated with these two questions in mind. In particular, testing principles relevant to the assessment of noninvasive stress testing technology will be explored.

Wu et al. employed rest and post-hyperventilation radionuclide ventriculography, including the phase analysis of the resting equilibrium blood-pool scintigrams, to evaluate the pathophysiologic consequences of coronary vasoconstriction in patients with vasospastic angina (*1*). Approximately two-thirds of the 36 vasospastic angina patients had a 5% or greater fall in left ventricular ejection fraction (LVEF) during hyperventilation in comparison to the baseline resting value. In addition, nonhomogenous contraction of the left ventricle, as reflected by the standard deviation of the left ventricular phase values (SD-LV), was present at rest in 94% of the patients with vasospastic angina, and the value for SD-LV was apparently related to the magnitude of coronary vasospasm induced during ergonovine testing. (The frequency of segmental wall motion abnormalities, which would be redundant with phase abnormalities, is not reported, but is pressumably low given the normal mean resting LVEF values in the study population). Finally, the investigators found a significant correlation between SD-LV at rest and the response of LVEF to hyperventilation. While these results are interesting in terms of pathophysiologic investigation, can they now be relied upon for clinical application?

Considerations Pertinent to the Assessment of Test Accuracy

The diagnostic accuracy of a noninvasive stress test is based on measurement of its sensitivity and specificity for detecting disease. Since many factors bias these measurements, this assessment is not straightforward. For example, unlike tests for many disease processes, the values for test sensitivity and specificity are variable when noninvasive stress tests are used to diagnose coronary artery disease (CAD), since these values are influenced by disease prevalence within the population being tested. The measurement of test sensitivity and specificity may be further influenced by pretest and post-test referral biases. Pretest referral bias results from the assessment of test results in narrow (instead of broad) spectrum populations. Post-test referral bias results from the preferential selection of positive test responses to angiography and negative test responses away from angiography. These biases are fully described elsewhere (*2–4*). In particular, because the assessment of test specificity has become so troublesome, due to post-test referral bias (*2,4*), investigators have struggled to attempt to define an alternative standard referent group when the angiographically normal population cannot be relied upon for test specificity. The controversy over the appropriate referent standard for test specificity also leads to a second debate: What is the appropriate referent standard on which to base normal test limits for normality, such as the SD-LV limits used in the Wu et al. study?

The three most common referent standards for evaluating test specificity and developing normal limits for test results are listed in Table 2. Each standard is flawed. Just as the use of patients with normal coronary arteriograms will tend to underestimate the true specificity of noninvasive tests due to the aforementioned post test referral bias, the use of healthy volunteers and low-likelihood CAD patients will result in an overestimation of a test's true specificity. Debiasing techniques for the evaluation of test sensitivity and specificity have been proposed (*5*) but have not been widely evaluated yet. As a compromise to this problem, many investigators have resorted to reporting both test specificity, based on the frequency of abnormal test results in patients with normal coronary arteriograms, and a normalcy rate based on the frequency of test abnormality in low-likelihood CAD patients. Low-likelihood CAD patients have also become the most

Received Nov. 2, 1994; accepted Jan. 17, 1995.
For correspondence or reprints contact: Alan Rozanski, MD, Division of Cardiology, St. Luke's-Roosevelt Hospital Center, 114th St. and Amsterdam Ave., New York, NY 10025.

TABLE 1
Common Considerations Associated with Noninvasive Stress Testing

Factors	Illustrative Examples of differences among studies
General category of testing	Exercise and ambulatory electrocardiography; exercise RNV; exercise and dobutamine echocardiography; exercise and pharmacologic myocardial perfusion SPECT; PET
Technical approach	Exercise RNV performed by the first pass vs. multiple gated equilibrium technique
Selection of test variables	Assessment of exercise RNV based on routine parameters such as exercise LVEF and wall motion, vs. additional nonroutine parameters such as those based on phase analyses, diastolic parameters
Goal of testing	Diagnostic vs. prognostic testing
Different stressors	Exercise/mental stress/hyperventilation/pharmacologic
Criteria for test abnormality	Exercise ST-segment depression may be analyzed as a dichotomas vs. categorical vs. continuous variable; or myocardial perfusion scintigrams may be assessed by subjective or quantitative analysis
Statistical analyses	Using multivariate analyses vs. area under ROC curve analyses to assess the added incremental information of a given test variable
Referent standards for normality and disease	NCA patients vs. low-likelihood CAD patients vs. healthy volunteers as a referent standard for normal test limits or test specificity
Study population	Differences in the acuity and spectrum of the diseased population, governed, in part, by differing magnitudes of pretest and/or post-test referral biases

commonly used referent standard for defining normal limits for scintigraphic tests. Low-risk CAD patients are more advantageous than healthy volunteers due to the fact that since they are already patients, the need for patient recruitment and unnecessary radiation exposure may be overcome, and because an explicit, albeit arbitrary, definition of normality can be defined: CAD risk level, based on Bayesian analysis of patient age, sex, symptoms, risk factors and any prior noninvasive tests (*6*).

Initially, patients with <1% CAD likelihood were the first group used as a referent standard for normal limits and for reporting normalcy rates for scintigraphic tests. Diamond, however, has characterized the use of patients with <1% CAD likelihood as misleading due to the supernormal nature of this group (*7*); they are too narrow a spectrum of the healthy population, manifesting uniformly normal test responses (Fig. 1) which are analogous to those observed in healthy volunteers. More recently, investigators have utilized CAD likelihood value of <5% to define normal referent populations for scintigraphic tests. But even the 5% limit may be too stringent a criterion. So far, no study has evaluated the most effective cut off value for CAD likelihood when using low-likelihood CAD patients as a referent standard.

Given this perspective, we can now evaluate the study by Wu et al. with respect to their assessment of test accuracy. Their results are clearly based on the evaluation of a highly preselected and biased group of normals: individuals who lacked any anginal symptoms, had normal rest and exercise electrocardiograms, normal echocardiograms and normal exercise myocardial perfusion scintigraphy results. As a result, these individuals had <1% mean pretest likelihood of CAD. As mentioned, when such supernormals (*7*) are used to define normal test limits for generating test abnormality criteria, we should expect that very high sensitivity test criteria would be generated. In this study, abnormal asynchronous LV contraction at rest, assessed by phase analysis, occurred in 96% of patients with vasospastic angina. But, are these results based on a criterion which is accordingly nonspecific? To assess this possibility, the authors should also have evaluated their technology in appropriate normal populations. Since

TABLE 2
Common Referent Populations to Assess Test Specificity and Define Normal Noninvasive Stress Test Limits

Referent standard	Test specificity measurement limits	Limitations for developing normal test limits
Patients with normal coronary arteriograms	Can significantly underestimate test specificity, due to a high concentration of false positive test responders when post-test referral bias is operative	Unusable when the population is significantly skewed by post-test referral bias
Healthy volunteers	Significantly overestimates test specificity, as it represents the "healthiest" extreme of the population without disease	Results in criteria for test abnormality which are too strict
Patients with a low CAD likelihood (<5%)	Overestimates test specificity, since it also represents a normal spectrum of the healthy patients	Can also result in criteria for test abnormality which are too strict; the optimal CAD likelihood value for this standard is currently undefined

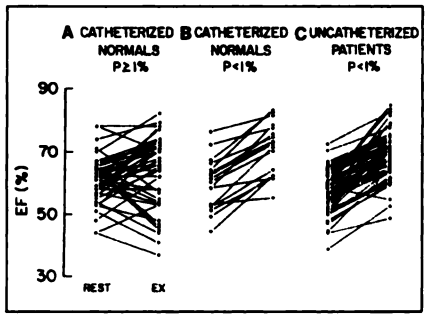

FIGURE 1. Rest and exercise (EX) LVEF responses for three groups: (A) patients with normal coronary arteriograms (catheterized normals) with concomitant pretest (i.e., pre-exercise RNV) probability (P) of CAD based on Bayesian analysis of >1; (B) a concurrent group of catheterized normals whose pretest CAD probability was also <1%; and (C) uncatheterized patients, whose pretest CAD probability was also <1%. Catherized patients with a pretest probability <1% manifested almost uniformly normal exercise LVEF responses, comparable in those who were or were not catheterized, with only one patient manifesting a fall in exercise LVEF.

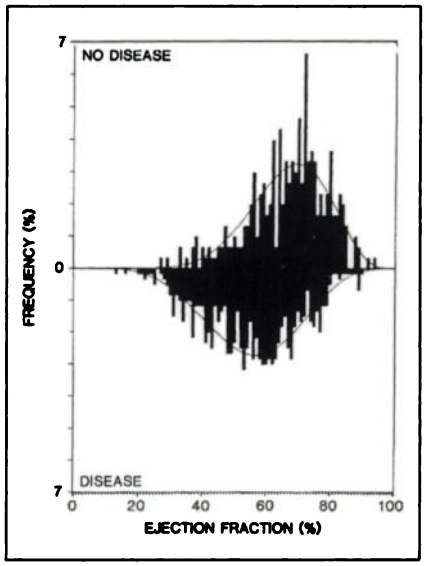

FIGURE 2. Peak exercise EF values during exercise radionuclide ventriculography plotted for 854 patients from a previous multicenter trial (*8*). The frequency of each given peak exercise EF value is shown for 297 patients without significant CAD above the horizontal line and for 557 patients with significant CAD (≥50 stenosis) below the horizontal line. The solid curves show the fitted beta distribution for exercise LVEF in angiographically normal and CAD patients. A considerable overlap in exercise EF values was noted in the normal versus diseased patients in this study.

they did not do so, their accuracy has not really been evaluated.

As an analogy, the criteria for normal versus abnormal LVEF responses to exercise, as measured by radionuclide ventriculography, were first assessed by comparing exercise LVEF responses in CAD patients to those noted in healthy volunteers and/or angiographic normals with a concomitantly low pretest likelihood of CAD (*2*). Test specificity was also based on the responses in such normals in initial validation studies. Extraordinary high specificity values were recorded (*2*), and the 5% rise in LVEF, a standard for defining a normal exercise LVEF response that was based on these initial studies, has remained the standard, despite significant flaws in this criterion (*8*). Later, following widespread clinical practice, it became clear that there was a considerable overlap in exercise LVEF responses among CAD and angiographically normal patients (*8*) (Fig. 2), which could not have been appreciated when only a very narrow (very well) spectrum of the available normal population was used for analysis. This created confusion in the clinical application of this test.

Pathophysiological Considerations

Pathophysiologic considerations associated with vasospastic angina will now be contrasted with those associated with atheromatous CAD (Table 3). In this study, the investigators evaluated the potential effects of vasospastic angina in two ways: by evaluating the frequency and magnitude of asynchronous contraction of the left ventricle (LV) at rest and by evaluating the change in LVEF following hyperventilation. The use of hyperventilation as a stimulus to induce coronary vasocon-

TABLE 3
Clinical Manifestations of Vasospastic Angina Versus Atheromatous CAD

Parameter	Vasospastic angina	Atheromatous CAD
Vasoconstriction	Present, by definition	Common
Nature	Focal, intense	More generalized, mild at site of atheromatous disease
Atheromatous disease	May be present or absent	Present, by definition
Pathophysiologic basis of vasoconstriction	Localized, segmental hypersensitivity	Coronary endothelial dysfunction
Induction of myocardial ischemia* by:		
Exercise	++	+++++
Mental stress	? (unknown)	+++
Cold stimulation	++	++
Hyperventilation	++++	+
Occurrence of silent ischemia	Yes	Yes
Greatest circadian density of ischemic episodes	2–6 am	Within first hour after morning awakening
Most common diagnostic method	Ergonivine testing in the catheterization lab	Noninvasive stress testing

*Subjective estimation of frequency (from 1+ to 5+) based on literature reports.

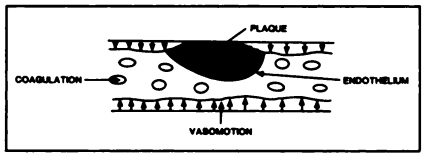

FIGURE 3. Schematic representation of pathophysiologic factors contributing to the presence of coronary vasoconstriction in patients with atherosclerotic CAD.

striction and/or ischemic findings in vasospastic angina patients has been previously studied, occurring in 54% to 80% of such patients in published reports (*9–11*). Thus, the investigators' experience with hyperventilation stress are compatible with those reported by others. The two novel findings in the current study are the high frequency of LV contraction abnormality at rest, and the relationship between rest SD-LV values and the magnitude of LVEF response to exercise. The investigators attribute the former finding to either resting silent myocardial ischemia or to myocardial stunning. The latter finding is largely unexplained.

To put these findings in context, it should be appreciated that vasospastic angina is a relatively uncommon phenomenon among patients presenting with chest pain in the United States. Furthermore, recent studies indicate that coronary vasoconstriction commonly occurs in patients with atheromatous coronary heart disease as well through different pathophysiologic mechanisms. As schematized in Figure 3, the development of atheromatous coronary plaque is associated with the concomitant presence of endothelial dysfunction: normal endogenous endothelial vasodilator relaxing factors (EDRF), such as nitrous oxide, are lost from the endothelium as it is rendered increasingly porous by the atherosclerotic process. In the presence of normal endothelium, neurohumoral substances, when activated by physiologic stimuli (e.g., exercise, mental stress), interact with nitrous oxide in the coronary endothelium to cause coronary vasodilation. In the absence of EDRF, neurohumoral substances secreted during physiologic stimuli interact with other receptor sites, to cause paradoxic vasoconstriction of the coronary artery, instead of vasodilation. In contrast to the intense focal vasospasm seen in patients with vasospastic angina, vasoconstriction in atheromatous CAD tends to be relatively mild and more generalized. Because of the curvilinear relationship between coronary luminal size and resistance to coronary blood flow, however, even small degrees of coronary vasoconstriction can be very important at sites of significant atheromatous obstruction in patients with CAD. Since physiologic stimuli, such as exercise, mental stress and smoking, may enhance coronary vasoconstriction in CAD patients, it is not surprising that these same stimuli can likewise induce myocardial ischemia in CAD patients during provocative stress in the laboratory. Hyperventilation seems to be a more important stimulus for ischemia in vasospastic angina patients, however, inducing ischemia in only 8% to 25% of patients with atheromatous CAD in prior studies (*11–13*).

These observations support the investigators' diagnostic approach in this study, but there are no pre-existing data to indicate that resting asynchronous LV contraction should be expected to occur more commonly in vasospastic angina patients as opposed to CAD patients. Rather, atheromatous CAD frequently results in resting left ventricular dysfunction in the absence of prior myocardial infarction. Thus, reversal of resting myocardial asynergy is a common phenomenon following the performance of coronary artery bypass surgery (*14*). Therefore, while the current report by Wu et al. is of interest as a pathophysiologic investigation, the relevance of their findings would be furthered by comparing this experience to patients with atheromatous CAD. Along these lines, is the relationship between rest SD-LV and the magnitude of LVEF change to hyperventilation, noted in this study, reproducible and unique for vasospastic angina?

Before such pathophysiologic investigation can be pursued, however, the accuracy of the measurements used by Wu et al. must be ascertained. A long list of test parameters which performed well in initial validation studies but has failed to remain robust when applied to a broad spectrum of diseased and normal patients now exists. To avoid the continual reoccurrence of this problem, it is time for those who assess and review the efficacy of noninvasive stress tests to adopt practices which might serve to diminish the information boondoggle which results from the publication of incomplete noninvasive stress tests evaluations. Just as strict guidelines must be fulfilled before the approval and release of therapeutic medications by the FDA, guidelines for test validation and reporting of results should be developed and disseminated by the experts in noninvasive stress testing. The following are minimal suggestions:

1. It is time to recognize that the process of evaluating noninvasive cardiac stress tests constitutes a scientific discipline. This discipline should integrate an understanding of: test technology, the pathophysiology and clinical significance of myocardial ischemia and sound principles of testing, as identified by an accumulating experience.
2. An appropriate scientific body (e.g., the American College of Cardiology, Society of Nuclear Medicine, or American Society of Nuclear Cardiology) or groups of investigators should take on the task of identifying and consolidating those principles of testing which may help to diminish reports based on a deficient evaluation of test technology. Examples of the principles that could be tackled include: definition of what constitutes a sufficiently "broad spectrum" population for test validation and reporting of results, development of referent standard(s) for cardiac normality, development/evaluation of debiasing methods to overcome the effects of pretest and post-test referral

bias, guidelines for defining and reporting equivocal test responses and delineation of the optimal method(s) for analyzing the added incremental information provided by new noninvasive tests relative to baseline information readily available.

3. Due to the short "half-life" of medical knowledge, there must be methods of keeping these principles in fresh existence. An example could be the publication of a page of guidelines, submitted to reviewers by editors of journals at the time of peer review. In addition, some of these principles could be incorporated into additional course work for postgraduate fellows in fields related to stress testing and imaging technology.

By identifying and promulgating standard testing principles, the information boondoggle which now plagues stress testing literature can be reduced. This would be a timely outcome in light of the increasing need to identify the true efficacy of each form of noninvasive stress testing within managed care environments.

ACKNOWLEDGMENTS

The author thanks Ms. Robyn Horowitz and Ms. Giti Bessinger for their technical assistance in the preparation of this manuscript.

Alan Rozanski
St. Luke's-Roosevelt Hospital Center
New York, New York

REFERENCES

1. Wu J, Takeda T, Toyama H, et al. Resting asynchronous left ventricular contraction abnormality analyzed by phase method in spastic angina pectoris. *J Nucl Med* 1995;36:1003–1008.
2. Rozanski A, Diamond GA, Berman D, et al. The declining specificity of exercise radionuclide ventriculography. *N Engl J Med* 1983;309:518–522.
3. Rozanski A, Diamond GA, Forrester JS, et al. Comparison of alternative referent standards for cardiac normality. Implications for diagnostic testing. *Ann Intern Med* 1984;101:164–171.
4. Rozanski A. Referral bias and the efficacy of radionuclide stress tests: problems and solutions. *J Nucl Med* 1992;3:2074–2079.
5. Diamond GA, Rozanski A, Forrester JS, et al. A model for assessing the sensitivity and specificity of tests subject to selection bias: application to exercise radionuclide ventriculography for diagnosis of coronary artery disease. *J Chron Dis* 1986;39:343–355.
6. Diamond GA, Forrester JS, Hirsch M, et al. Application of conditional probability analysis to the clinical diagnosis of coronary artery disease. *J Clin Invest* 1980;65:1210–21.
7. Diamond GA. An improbable criterion of normality. *Circulation* 1982;66:618.
8. Rozanski A, Diamond GA, Jones R, et al. A format for integrating the interpretation of exercise ejection fraction and wall motion with special reference to equivocal responses. *J Am Coll Cardiol* 1985;5:238–248.
9. Kaski JC, Crea F, Meran D, et al. Local coronary supersensitivity to diverse vasoconstrictive stimuli in patients with variant angina. *Circulation* 1986;74:1255–1265.
10. Previtali M, Ardissino D, Barberis P, et al. Hyperventilation and ergonovine tests in prinzmetal's variant angina pectoris in men. *Am J Cardiol* 1989;63:17–20.
11. Perez Balino NA, Liprandi AS, Masoli O, et al. Usefulness of radionuclide ventriculography in assessment of coronary artery spasm. *Am J Cardiol* 1987;59:552–558.
12. Ardissino D, DeServi S, Barberis P, et al. Significance of hyperventilation-induced ST segment depression in patients with coronary artery disease. *J Am Coll Cardiol* 1989;13:804–810.
13. Ardissino D, Barberis P, De Servi S, et al. Usefulness of the hyperventilation test in stable exertional angina pectoris in selecting medical therapy. *Am J Cardiol* 1990;65:417–421.
14. Rozanski A, Berman D, Gray R, et al. Preoperative prediction of reversible myocardial asynergy by postexercise radionuclide ventriculography. *N Engl J Med* 1982;307:212–216.