Editorial Comment

Agreement on the Accuracy of Thallium Stress Testing*

K. LANCE GOULD, MD, FACC
Houston, Texas

The diagnostic accuracy of thallium stress testing. My earlier editorial comment (1) on the accuracy of thallium stress testing for the diagnosis of coronary artery disease, which accompanied the study of Iskandrian et al. (2), made three essential points:

1. The average reported sensitivity and specificity of thallium exercise testing in 2,877 cumulative cases published since 1983 (2–7), weighted for the number of cases in each study, is 83% and 53%, respectively, associated with approximately 17% false negative and 47% false positive results (Table 1).

2. Referral bias cannot completely explain this poor diagnostic accuracy, which is therefore a real limitation, reducing clinical utility of thallium stress testing as currently used.

3. This degree of error in a noninvasive diagnostic test incurs considerable economic and personal costs related to making the correct diagnosis by coronary arteriography, as reported in studies (14–16) published since that editorial comment (1).

Diamond’s superb editorial in this issue of the Journal (8) confirms the essential points of my editorial comment by more rigorous logic. In the studies cited in his Table 5 (8), observed sensitivity ranged from 78% to 91% and observed specificity from 34% to 63%. With use of a sophisticated analysis that corrected for referral bias, Diamond calculated a “corrected” sensitivity and specificity from the observed values. After accounting for referral bias, his average “corrected” sensitivity was 67.5% and average “corrected” specificity 75%.

My editorial comment (1) carefully concluded that referral bias could not completely explain the reported poor sensitivity and specificity of thallium stress testing. Depending on whether images were aggressively or conservatively interpreted as abnormal, I estimated that the currently reported average 83% sensitivity and 53% specificity were approximately equivalent to 65% sensitivity and 65% specificity if images were correctly interpreted by approximately the same criteria as in the years before 1983, without correction for referral bias. Diamond’s estimates (8) with a correction for referral bias were an average of 67.5% “corrected” sensitivity and 75% “corrected” specificity, compared with my estimates of 65% and 65%, respectively, without correction for referral bias. Therefore, his results are remarkably consistent with mine and verify the poor diagnostic accuracy of thallium stress testing regardless of how the analysis is done.

Definitions and terminology of diagnostic accuracy. There are minor differences in our approaches that Diamond (8) criticizes. He eloquently defends referral bias as a concept. I did not claim otherwise, but only that referral bias cannot be invoked to completely explain the poor diagnostic result confirmed by his own analysis. With regard to the definition of sensitivity and specificity that Diamond criticizes, it is clear that Iskandrian and I understand the term. This criticism merely establishes Diamond’s expertise and the “rigor” of his analysis, all of which I accept.

However, there is a valid semantic difference here because we have no accepted term for a hypothetical sensitivity and specificity, for a hypothetical group affected by a hypothetical referral bias, none of which can be empirically verified. Which terms should we use: “true,” “measured,” “corrected for referral bias” or “apparent” sensitivities and specificities—all of which may or may not be invariant, according to Diamond’s claims, depending on how they are defined? Diamond’s use of positive and negative predicted accuracy may avoid the semantic issue, but his calculations also confirm the poor diagnostic accuracy of thallium stress testing. It is important that our minor semantic differences do not detract from our remarkable agreement, namely, that the diagnostic accuracy of thallium stress testing, as currently practiced, is not satisfactory in modern cardiovascular medicine.

Two other statements in Diamond’s analysis (8) are worthy of emphasis. First, he confirms the error in using patients not undergoing catheterization as normal subjects, which I also criticized. Second, he emphasizes that single photon emission computed tomographic (SPECT) thallium testing has not yet been demonstrated to have greater accuracy than planar imaging, a point of view for which I am often accused of bias.

The personal and economic costs of initial misdiagnosis by thallium stress testing. These costs may be substantial, including unnecessary coronary arteriography in patients with a false positive test, as recently analyzed (1.14–16). The price of not detecting disease by noninvasive testing may also be substantial in terms of both lives and medical

*Editorials published in Journal of the American College of Cardiology reflect the views of the authors and do not necessarily represent the views of JACC or the American College of Cardiology.

From the Division of Cardiology and the Positron Diagnostic and Research Center, The University of Texas Health Sciences Center at Houston, Houston, Texas.

Address for reprints: K. Lance Gould, MD, Box 20708, Room 4.258, Division of Cardiology, University of Texas Medical School at Houston, Houston, Texas 77225.

©1990 by the American College of Cardiology 0735-1097/90/$3.50
expense. These hidden costs need to be considered when analyzing the cost-effectiveness of positron emission tomography (PET) with its greater diagnostic accuracy.

In my personal practice, I have seen several cases of sudden death or myocardial infarction, or both, within several weeks or months after a normal thallium scan. In addition, I commonly see normal thallium studies in patients who are identified by PET to have severe coronary artery disease requiring mechanical intervention, or moderate disease requiring vigorous cholesterol lowering that would not otherwise be instituted. As a practicing cardiologist with a reasonably intense interest in my patients' well-being, I find it untenable to say to them 50% to 60% of the time, "sorry, my noninvasive test showing that you have coronary artery disease (or not) was wrong." After personally performing over 1,000 clinical PET studies and making therapeutic decisions based on them, I don't need to say that very often anymore.

It is gratifying to see this series of concurring reports by Iskandrian, Diamond and myself, which finally clearly states the questions and available data to allow cardiovascular practitioners to make their own judgments on the accuracy of thallium exercise testing.

References