Editorial Comment

How Accurate Is Thallium Exercise Testing for the Diagnosis of Coronary Artery Disease?*

K. LANCE GOULD, MD, FACC

Houston, Texas

The report by Iskandrian et al. (1) in this issue of the Journal contains startling data definitively answering the single most important current question about thallium exercise testing—how accurate is it? Although the sensitivity and specificity of exercise thallium imaging in earlier published reports (2–4) are 80% to 90% in symptomatic patients, more recent studies (5–10) report a sensitivity of 70% to 85% and a specificity of 50% to 60% in both symptomatic (5–7) and asymptomatic (8,9) subjects or those with atypical presentations (10), consistent with the results of Iskandrian et al. (1).

Reasons for low specificity of thallium stress test. One explanation proposed for low specificity in recent reports is that patients with a negative thallium stress test no longer undergo cardiac catheterization (6). The catheterized subjects in such a study would then be biased by this exclusion of normal subjects, thereby skewing the study group toward a higher prevalence of disease. According to Bayes' theorem, as the study population is skewed toward greater prevalence of disease, the post-test probability of having no disease with a normal test or observed test specificity decreases in that population (11,12). The second argument made in support of this point of view is that the "normalcy rate" for exercise thallium single photon emission computed tomography (SPECT) is about 90% in uncatheterized patients having <5% probability of coronary artery disease by virtue of young age and no risk factors or family history (6).

The *fact* of low specificity at 50% to 60% *reported* in recent studies including that of Iskandrian et al. (1) is not disputed (5–10). The explanation for this low specificity, and therefore the value of thallium stress testing for diagnostic purposes, depends on which of the alternative, opposing explanations is true: 1) thallium stress testing has sufficiently low true specificity to make it of limited value in a population characterized by moderate prevalence of disease (10% to 30%) in recent studies and is therefore not economically or medically appropriate, or 2) thallium stress testing is so good that it excludes so many normal subjects from cardiac catheterization in current study populations that the catheterized population of recent studies is now skewed toward high prevalence and low reported apparent specificity because normal subjects are no longer being catheterized.

The currently reported low specificity of 50% to 60% is unlikely to be due to referral bias, as indicated by the Bayes' theorem graph in Figure 1A (11,12). Let us hypothesize generously in favor of exercise thallium SPECT that its true sensitivity and specificity are both 90%. Then let us ask whether all the data fit this hypothesis, thereby making it true or false. If the observed post-test probability of disease for a normal noninvasive test is 50% in a study population, the apparent specificity in the study population would be 50% (with 50% having a false positive result). The referral bias hypothesis proposed to explain this low specificity would claim that referral bias has increased the prevalence of disease in the study group. Therefore, the next question is how high would the prevalence of disease in a study population have to be to have a 50% post-test probability of a normal arteriogram for a normal thallium test result, i.e., an apparent specificity of 50% in the study population. From Figure 1A, the prevalence of disease would have to be 90% *and*, at that prevalence of disease, the post-test probability of having disease for an *abnormal* thallium test, or apparent sensitivity, would approach 99%. The apparent sensitivity in the study population, i.e., the post-test probability of disease from a positive test, should be nearly perfect because the prevalence of disease is so high. Thus, if the true specificity of the test was 90% but appeared to be 50% because of referral bias affecting the study population, then the prevalence of disease in the study population would have to be 90% and the post-test probability of disease or the apparent sensitivity would have to be 99%. However, prevalence of disease in published studies ranges from 55% to 75% and sensitivity from 65% to 85%, values inconsistent with the hypothesis that true specificity is 90% appearing to be 50% because of referral bias in the study population. Therefore, the low specificity of 50% cannot be explained by a high prevalence of disease produced by referral bias. In addition, a large recent study (9) with a sensitivity of 76% and a specificity of 49% was not biased by referral selection because all 832 subjects, asymptomatic Air Force personnel, had coronary arteriograms regardless of thallium exercise test results.

*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 University of Texas Medical School, Houston, Texas. This study was supported in part by Grants RO1-HL26862 and RO1-HL26885 from the National Institutes of Health, Bethesda, Maryland and by a joint collaborative research project with the Clayton Foundation for Research, Houston, Texas.

Address for reprints: K. Lance Gould, MD, Room 4.258 MSMB, P.O. Box 20708, University of Texas Medical School, Houston, Texas 77225.

0735-1097/89/$3.50

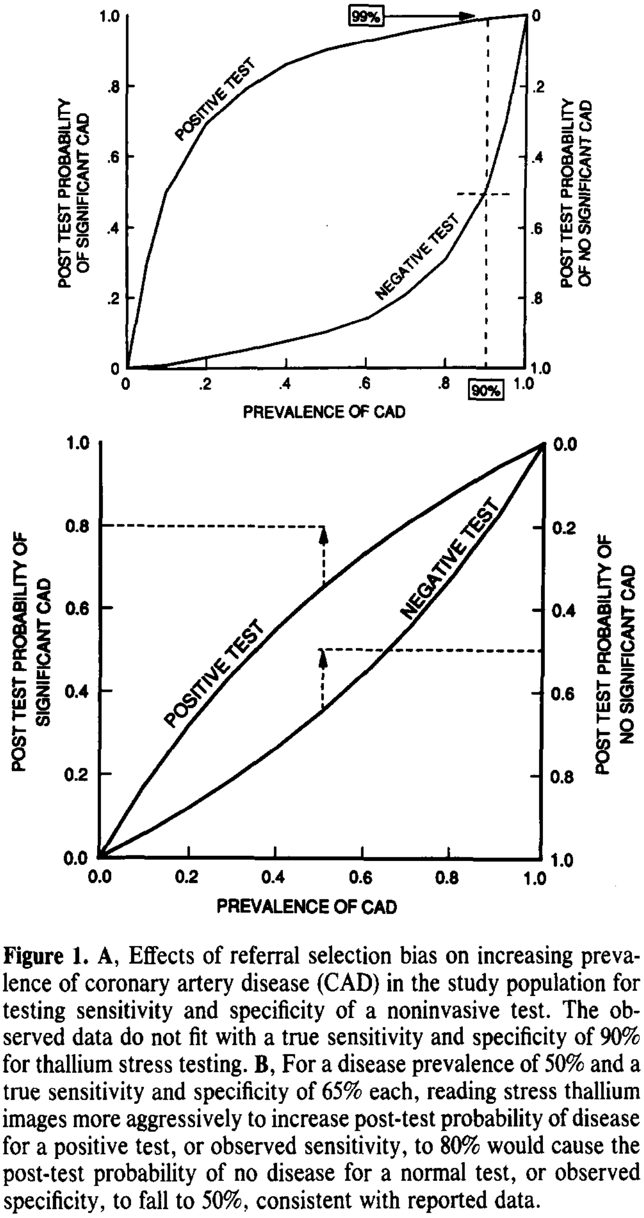

Figure 1. **A**, Effects of referral selection bias on increasing prevalence of coronary artery disease (CAD) in the study population for testing sensitivity and specificity of a noninvasive test. The observed data do not fit with a true sensitivity and specificity of 90% for thallium stress testing. **B**, For a disease prevalence of 50% and a true sensitivity and specificity of 65% each, reading stress thallium images more aggressively to increase post-test probability of disease for a positive test, or observed sensitivity, to 80% would cause the post-test probability of no disease for a normal test, or observed specificity, to fall to 50%, consistent with reported data.

The present study. On the basis of Figure 1, the prevalence of disease in the study population of Iskandrian et al. would have to be 85% to explain their specificity of 62% as due to referral bias. However, of 461 patients in their study population, 272 were classified as having coronary artery disease, a prevalence of only 59%. In their discussion of clinical implications, Iskandrian et al. unintentionally demonstrate the major point here by incorrectly excluding the low probability normal group to calculate a disease prevalence of 272 of 330 or 82% in the catheterization group, while at the same time including this group as normal as if catheterized in the overall complete study. Because the low probability group is classified as normal in the study, the true prevalence of disease in the study population is 59%. Thus, referral bias cannot explain the low specificity of Iskandrian et al. because actual disease prevalence of 59% was much lower than the prevalence of 85% necessary to explain low specificity on this basis.

The study population with <5% probability of coronary artery disease on which "normalcy rates" of thallium tests are calculated (6) is characteristically a younger, more vigorous population selected to have few risk factors causing coronary artery disease. That younger population is also likely to have fewer causes for a false positive thallium stress test than would an older group of patients, who have a greater risk of coronary artery disease and a greater likelihood of a false positive exercise test because of differences in the anatomy of the older, heavier person, including greater chest diameter, greater body mass, bone density, greater diaphragmatic and breast attenuation and other factors that cause false-positive thallium images. Table III of the report of Iskandrian et al. (1) documents the difference in body habitus between the low risk normal subjects and the group at risk for disease. Therefore, the "normalcy rate" for thallium testing from a young population with low probability of disease cannot be extrapolated to the study population at risk for coronary artery disease. In this and previous reports the uncatheterized low risk normal subjects are used as if they had been catheterized, to increase the number of patients in the study and to show a favorable "normalcy rate" to offset low observed specificity while simultaneously classifying this group as uncatheterized to avoid the fact of low disease prevalence that disproves referral bias.

Sensitivity and specificity of the thallium stress test. How accurate is thallium stress testing? The answer consistent with recently published data is suggested by Figure 1B, a corresponding graph for a true sensitivity and specificity of 65% each. Because of the inverse relation between sensitivity and specificity, at a disease prevalence of 50%, if images were read so as to increase sensitivity to 80% (indicated by the upper dashed arrow) then specificity would decrease to 50% (lower dashed arrow), consistent with published data. If images were read so as to give higher specificity, then sensitivity would be lower. The net or combined sensitivity-specificity data content of thallium stress imaging corresponds to an approximate equivalent sensitivity and specificity of 65% each, where reported sensitivity and specificity vary inversely depending on how aggressively or conservatively abnormalities are called on scans. By conservative or aggressive interpretation of images as abnormal, the scan reader chooses the preponderance or balance of false positive or false negative results but cannot substantially decrease both to clinically acceptable levels.

How does this analysis fit with higher reported sensitivity and specificity in earlier studies? The study populations in these early studies consisted of patients undergoing cardiac catheterization on clinical grounds. Therefore, these early study populations had a higher prevalence and more advanced disease than do current study populations in which thallium stress testing is applied to determine whether catheterization is indicated in the absence of a clear clinical diagnosis. Conservative image interpretation in a population of high disease prevalence will give reasonably good sensitivity and specificity as reported in early publications. However, conservative image interpretation with the same number of false negative and false positive results applied to a population of lower disease prevalence (30% to 50%) decreases sensitivity markedly whereas specificity increases. To maintain higher sensitivity, the reader then interprets images more aggressively, thereby keeping sensitivity high but at a cost of decreasing specificity. Therefore, recent work showing the same sensitivity but markedly lower specificity compared with earlier studies can be explained by decreased disease prevalence in the study populations combined with more aggressive image interpretation. With more aggressive interpretations in current lower prevalence study populations, 65% sensitivity and 65% specificity become skewed to a sensitivity of 80% and specificity of 50% to 60% reported in current studies, confirmed by the study (1) of Iskandrian et al. and explained by Figure 1B.

Clinical implications. For a sensitivity of 80% as most images are currently read, the 60% specificity of thallium testing has major medical and economic consequences. With use of sensitivity and specificity values of approximately 65% and 50%, respectively, for exercise electrocardiography and 80% and 60% for exercise thallium scintigraphy, the efficacy of conventional sequential testing for detecting coronary artery disease can be analyzed. In a population with 10% to 15% prevalence of coronary disease, as would be expected in patients with positive risk factors (13,14), if the combination of a positive treadmill electrocardiogram (ECG) followed by a confirmatory positive thallium scan were required for proceeding to catheterization, 48% of patients $[1.0-(0.65)(0.8)]$ with angiographically significant coronary artery disease would not be catheterized. If thallium testing were used in all patients instead of an initial exercise ECG exercise test, *only 18% of the catheterized patients would have disease* in a population with a prevalence of disease of 10%, calculated as $(0.8 \times 10) \div (0.8 \times 10 + 0.4 \times 90)$. Thus, the low specificity of thallium stress testing as recently reported (5–10) and confirmed by the study of Iskandrian et al. (11) markedly reduces the clinical utility of thallium stress testing in populations of moderate prevalence in which a noninvasive test is most needed. In addition, Iskandrian et al. have demonstrated that sensitivity is highly dependent on achieving maximal exercise stress.

In contrast, for a noninvasive test with 95% sensitivity and 95% specificity as reported for positron emission tomography (15–19), unnecessary cardiac catheterizations are largely avoided and severity of coronary artery disease is categorized noninvasively (16–20), thereby eliminating further catheterizations in patients having mild disease by positron emission tomography suitable for medical management. This greater accuracy of positron emission tomography has major economic benefits by preventing unnecessary catheterizations due to false positive or equivocal thallium stress tests that more than compensate for its cost in comparison to thallium stress testing, as well as providing improved medical management (21). Therefore, serious consideration should be given to positron emission tomography dipyridamole perfusion imaging, which provides sufficiently high diagnostic accuracy to substitute for diagnostic coronary arteriography in many patients.

References

1. Iskandrian AS, Heo J, Kong B, Lyons E. Effect of exercise level on the ability of thallium-201 tomographic imaging in detecting coronary artery disease: analysis of 461 patients. J Am Coll Cardiol 1989;14:1477–86.
2. Ritchie JL, Trobaugh GB, Hamilton GW, et al. Myocardial imaging with thallium-201 at rest and during exercise. Comparison with coronary arteriography and resting and stress electrocardiography. Circulation 1977:56:66–78.
3. Ritchie JL, Zaret BL, Strauss HW, et al. Myocardial imaging with thallium-201: a multicenter study in patients with angina pectoris or acute myocardial infarction. Am J Cardiol 1978;42:345–50.
4. 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 and washout of thallium-201. Circulation 1981;64:924–35.
5. DePasquale EE, Nody AC, DePuey EG, et al. Quantitative rotational thallium-201 tomography for identifying and localizing coronary artery disease. Circulation 1987;77:316–27.
6. Van Train KF, Berman DS, Garcia EV, et al. Quantitative analysis of stress thallium-201 myocardial scintigrams: a multicenter trial. J Nucl Med 1986;27:17–25.
7. Ranhosky A, Gerlag DM. Quantitative interpretation provides no advantage over qualitative interpretation in intravenous dipyridamole thallium imaging (abstr). Circulation 1988;78(II):II-432.
8. Nolewajka AJ, Kostuk WJ, Howard J, Rechnitzer PA, Cunningham DA. 201-thallium stress myocardial imaging: an evaluation of fifty-eight asymptomatic males. Clin Cardiol 1981;4:134–8.
9. Schwartz RS, Jackson WG, Celio PV, Hickman JR. Exercise thallium-201 scintigraphy for detecting coronary artery disease in asymptomatic young men (abstr). J Am Coll Cardiol 1988;11:80A.
10. Bungo MW, Leland OS. Discordance of exercise thallium testing with coronary arteriography in patients with atypical presentations. Chest 1983;83:112–6.
11. Hamilton GW, Trobaugh GB, Ritchie JL, Gould KL, DeRouen TA, Williams DL. Myocardial imaging with 201-thallium: an analysis of clinical usefulness based on Bayes' theorem. Semin Nucl Med 1978; 8:358–64.
12. Hamilton GW. Myocardial imaging with thallium-201: the controversy over its clinical usefulness in ischemic heart disease. J Nucl Med 1979;20:1201–5.
13. Uhl GS, Froelicher V. Screening for asymptomatic coronary artery disease. J Am Coll Cardiol 1983;1:946–55.
14. Houck PD. Epidemiology of total cholesterol to HDL ratio in 11,669 Air Force personnel and coronary artery anatomy in 305 healthy aviators (abstr). J Am Coll Cardiol 1988;11:222A.
15. Gould KL, Goldstein RA, Mullani NA, et al. Noninvasive assessment of coronary stenoses by myocardial perfusion imaging during pharmacologic coronary vasodilation. VIII. Clinical feasibility of positron cardiac imaging without a cyclotron using generator-produced rubidium-82. J Am Coll Cardiol 1986;7:775–89.

16. Demer LL, Gould KL, Goldstein RA, Kirkeeide RL. Diagnosis of coronary artery disease by positron emission tomography: comparison to quantitative coronary arteriography in 193 patients. Circulation 1989; 79:825–35.

17. Gould KL. Identifying and measuring severity of coronary artery stenosis: quantitative coronary arteriography and positron emission tomography. Circulation 1988;68:237–45.

18. Schelbert HR, Wisenberg G, Phelps ME, et al. Noninvasive assessment of coronary stenosis by myocardial imaging during pharmacologic coronary vasodilation. VI. Detection of coronary artery disease in man with intravenous NH3 and positron computed tomography. Am J Cardiol 1982;49:1197–207.

19. Tamaki N, Yonekura Y, Senda M, et al. Myocardial positron computed tomography with N-13 ammonia at rest and during exercise. Eur J Nucl Med 1985;11:246–51.

20. Goldstein RA, Kirkeeide R, Demer L, et al. Relations between geometric dimensions of coronary artery stenoses and myocardial perfusion reserve in man. J Clin Invest 1987;79:1473–8.

21. Gould KL, Goldstein RA, Mullani NA. Economic analysis of clinical positron emission tomography of the heart with Rubidium-82. J Nucl Med 1989;30:707–17.