The Accuracy of Quantitative Analysis of Stress/Delayed Thallium-201 Myocardial Tomograms

John C. Lasher
Ralph Blumhardt
William T. Phillips
University of Texas Health Science Center at San Antonio
San Antonio, Texas

TO THE EDITOR: We have a few questions regarding a recent paper by Garcia et al. (*1*) describing a multicenter study for validating the accuracy of a new Emory program for quantitating reversibility of stress-induced thallium-201 myocardial perfusion images.

The first question concerns the definition of accuracy. The calculation of accuracy requires determining sensitivity and specificity (*2*). This study evaluates sensitivity but not specificity when comparing the new Emory program for detecting reversibility with visual interpretation of four experts. It would be important to know if there were patients, and if so how many, who demonstrated no reversibility according to the experts but did show reversibility with the new Emory program, i.e., what is the false-positive rate? Do the authors recommend diagnosing reversible ischemia if the new Emory program is positive and the images appear normal?

We also question this method when quantitating data in patients with balanced multi-vessel disease. Wouldn't one miss abnormal reversibility with disease involving the three major coronary vessels with this normalizing technique? In a previously described method of quantifying rotational thallium-201 myocardial tomography (*3*), the relative change in counts between stress and delayed images was handled by "multiplicative scale factors provided by commercial programs." Perhaps this relative method would detect multi-vessel disease, however, we would like to know how these scale factors were derived and validated since they were not discussed.

Our last question concerns the wisdom of recommending the use or this new program to train "diagnosticians with limited experience in interpreting thallium tomograms," since even the experts did not fully agree. There was a significant difference in performance of two of the four experts in diagnosing reversibility compared to the new Emory program (Table 2). One expert "tended to relate more subtle reversibility with significant ischemia." Furthermore, when there is disagreement between the experts and the new Emory program, there is no way to tell who or which is correct.

We belive these questions need to be addressed since the use of this new program will undoubtedly become widespread. The article will be very valuable to General Electric Medical Systems, one of the institutions participating in this study. We predict that they will heavily market this technology just as they have the previous Emory program.

REFERENCES

1. Garcia EV, DePuey EG, Sonnemaker RE, et al. Quantification of the reversibility of stress-induced thallium-201 myocardial perfusion defects: a multicenter trial using bull's-eye polar maps and standard normal limits. *J Nucl Med* 1990;31:1761–1765.
2. Goris, ML. *Sensitivity and specificity of common scintigraphic procedures.* Chicago: Year Book Medical Publishers, Inc.; 1985:2.
3. Garcia EV, Van Train K, Maddahi J, et al. Quantification of rotational thallium-201 myocardial tomography. *J Nucl Med* 1985;26:17–36.

REPLY: In our multicenter validation, we chose to establish the accuracy of the method in terms of how well the results of the program agreed with those of experts in determining reversible versus fixed defects. This decision was based on the difficulty of establishing a gold standard to measure ischemia or infarction in vivo. Because of this choice and because in this analysis the absence of detecting reversibility did not mean a normal finding but rather a fixed defect, we chose to avoid terms like sensitivity and specificity. Nevertheless, it is not difficult to determine how many defects, or patients, were assessed as reversible by the program but demonstrated no reversibility by the expert's interpretations. Table 1 on our multicenter paper (*1*) gives a detailed comparison on a vascular territory basis. From the right column of this table, it can be determined that of 83 defects assessed to be fixed by the experts 15 were determined to be reversible by the new method. This comparison yields an 18% disagreement rate, which would correspond to what Lasher et al. call a false-positive rate. Since the purpose of this analysis is to determine how well the program's results agree with experts, it is clear that if there is a disagreement it is the program that is wrong. Previously, we established using five experts that there is only a 7% interobserver variability in the visual assessment of defect reversibility (*2*). The fact that experts disagree is not different from the fact that repeated measurements with any "gold standard" yield different results at least some of the time.

In addition to analyzing results in large populations, it is also important to analyze how to use the program in specific cases. One case of concern described by Lasher et al. is when the program suggests there is a reversible defect when the images appear normal to the physician. The program will show regions that change between stress and delayed imaging and the magnitude of the change. But as implemented, the program will not flag a region as reversible unless it was first determined to be associated with a stress-perfusion defect using the quantitative critera. Another case of concern is that the program will miss determining reversibility when the patient has balanced multivessel disease. If the flow reduction to all vascular beds is truly balanced (something we suspect happens rather infrequently), then no stress-perfusion defect will be detected since there is no myocardial region demonstrating a relative reduction in counts no matter what scale or normalization factor is used. This will confuse both the program and the expert into interpreting the scan as normal unless other markers of disease are used such as lung uptake of thallium-201 or slow washout of thallium-201 from the myocardium. One feature of our approach is that we continue to quantify the percent washout from the myocardium between stress and delayed imaging. Although we have not systematically analyzed how this independent parameter comple-