

Gated SPECT and the Visual Gold Standard: Gold Standard or Not?

In this issue of *The Journal of Nuclear Medicine*, Sharir et al. (1) report normal limits for quantitative regional motion and thickening measurements by gated myocardial perfusion SPECT. These investigators sought to “assess the normal heterogeneity in poststress motion and thickening by ^{99m}Tc gated myocardial perfusion SPECT and to determine and validate quantitative criteria for abnormal poststress motion and thickening for individual myocardial segments.” Specifically, they report “a substantial apex-to-base decline in thickening” and “circumferential heterogeneity in endocardial motion.” Additionally, they report that the criteria they developed for assigning semiquantitative measures of abnormality were accurate for the identification of motion and thickening abnormalities using these algorithms. The gold standard for these studies was expert visual interpretation of the same images analyzed quantitatively for the validation of the automatic computer algorithms reported in the study. It is an unusual experimental design that relies on visual interpretation of the same images rather than independent gold standards from 1 or more different imaging modalities. We might call this the “visual gold standard.”

Before discussing the interesting article by Sharir et al. (1), it should be stated that the authors of this commentary have also developed and commercialized quantitative software that automatically analyzes gated perfusion SPECT images for the same

parameters of regional ventricular function as studied in the report by Sharir et al. Commercially, we compete directly with Sharir et al. Although this may be seen as a conflict of interest, we have tried to be objective and fair both to the article by Sharir et al. and to the medical community. Any criticisms of their article can generally be regarded as equally self-critical. We have in preliminary form, and in the near future hope to publish, somewhat similar findings resulting from our own computer algorithms (2). Just as it can be surmised that Sharir et al. will further pursue the quantitative analysis and clinical application of these gated SPECT methods, we also will continue our pursuits.

During the 1990s, stimulated by the availability of ^{99m}Tc perfusion tracers and advances in imaging hardware and software, gated perfusion SPECT became the standard image acquisition technique used in most nuclear medicine laboratories (3,4). Gated perfusion SPECT has been shown to have important applications in the assessment of coronary heart disease, patient prognosis, regional viability, and the differentiation of fixed attenuation artifacts from myocardial infarction (5,6). Quantitative computer software for the analysis and visualization of these studies has played an important role in these developments. Computer applications such as these are mandatory for the accurate, reproducible, and time-efficient interpretation of the large volumes of data that these studies provide.

The validation procedure used in the report of Sharir et al. (1) raises some concerns. Validation means to establish the soundness of a method or, in this case, to corroborate the findings the authors' new method provides. In medical imaging, validation has tradi-

tionally meant confirmation of the accuracy of a new technique in comparison with an accepted independent standard (e.g., perfusion imaging has generally been compared with coronary angiography). Gated blood-pool imaging and echocardiography have been validated against contrast ventriculography (CVG). Subsequently, they also have been used as gold standards. More recently, gated and fast cine MRI and 3-dimensional (3D) echocardiography have been used as confirming standards. The study design by Sharir et al. is in most ways excellent. The authors developed reference ranges for regional wall motion and wall thickening in a healthy population. Using receiver operator characteristic (ROC) analyses, they taught their algorithms the required thresholds for grades of abnormality so that the program “saw” what the expert interpreters saw (7). Finally, they tested their algorithms and thresholds in a separate test population, comparing the algorithm-determined semiquantitative regional function scores with similar scores determined by expert visual interpretation of the same images. This study does not compare the algorithm-determined automatic quantitative results with an independent gold standard. The visual interpretations reported were not of an independent imaging technique (e.g., CVG, echocardiography, or MRI). The algorithm-determined scores were derived from the same gated SPECT images that served as the gold standard for visual interpretation. In this case, it might be better stated by the authors that they confirmed the correspondence between automatic and semiquantitative visual assessments of regional left ventricle (LV) function from gated myocardial perfusion SPECT. Unfortunately, the gold standard in this

Received Jul. 11, 2001; revision accepted Jul. 17, 2001.

For correspondence or reprints contact: James R. Corbett, MD, Division of Nuclear Medicine, University of Michigan Hospitals, B1G412, Box 0028, 1500 E. Medical Center Dr., Ann Arbor, MI 48109.

study—the visual interpretation of gated perfusion SPECT—is neither accurate nor independent, which is a problem.

Is an independent gold standard important? Perhaps more to the point, is an accurate independent gold standard important? The answer must be yes. In their study, Sharir et al. (1) report reference values for regional myocardial thickening and motion. The problem is that it is known from quantitative analysis of other highly credible imaging modalities that the reference values reported by Sharir et al. err significantly in several important respects. Imaging modalities such as CVG, echocardiography, and MRI have much higher spatial resolution than does gated SPECT perfusion imaging. In images obtained during breath holding, resolution is further enhanced compared with SPECT. As reported by Sharir et al., the high values for wall thickening and motion at the apex and the very low values for thickening at the base are incorrect, according to other studies (8–16).

At the apex, the excessive wall thickening in normal hearts reported by the authors is likely the result of a combination of factors, including the steeper relationship between wall thickness and count density at the anatomically thinned apex, as discussed in the article (17). Increased scatter into the region of the LV apex as adjacent myocardial walls approach and the papillary muscles descend into a more tightly packed distal LV chamber during ventricular systole also clearly plays a role, especially in relatively small normal hearts. These factors probably contributed to the excessive apical thickening described by Sharir et al. (1).

The steep fall-off in wall thickening at the LV base noted in the article is another variance from other imaging modalities. The explanation for this is less clear, although early reports from MRI literature may provide some insight (18). It is known from CVG, echocardiography, and MRI that the valve plane descends toward the apex approximately 1 cm in normal hearts during ventricular systole. One centi-

meter is the width of 1.5–2.1 SPECT short-axis sections, which typically vary in thickness from 4.8 to 6.5 mm. It is also known that cardiac MR images have excellent in-plane resolution (1–2 mm) but slices are 5- to 10-mm thick. Perhaps because of the severely anisotropic voxels MRI provides, early studies analyzed wall thickening and motion within individual slices without taking into account motion parallel to the LV long axis and, like Sharir et al. (1), they reported a large apex-to-base gradient in wall thickening. When MRI studies underwent 3D analysis to account for the motion of the base and apex parallel to the long axis, very different results were obtained (13–16). There was little if any gradient in thickening from apex to base, and the apex showed less motion than other ventricular segments. Wall thickening measurements from gated SPECT are based on the percentage change in wall counts from end-diastole to end-systole. The authors have not clearly described how the valve plane is handled by their algorithms (19–21). However, if the mitral valve plane is artificially fixed or its motion is underestimated, depending on the amplitude of its motion in the individual healthy volunteer or patient, the end-diastolic myocardial counts at the base of the heart will be compared with end-systolic counts not only from the LV but also from the mitral valve apparatus and left atrium (LA). Because the LA wall is much thinner than the LV wall and the valve is not a muscular structure, the basal LV counts at end-systole will be greatly underestimated and a significant underestimation of LV thickening at the LV base will result.

The exaggerated apex-to-base gradient in wall thickening and the excessive motion of the LV apical and distal anterior segments all run counter to the established physiology of the normal heart. It has been established by CVG and cine fluoroscopy in instrumented animal models and man 2 decades ago, and, more recently, confirmed by echocardiography and MRI as discussed above that, in normal hearts, the LV apex moves less than most other segments, including the mitral valve plane

(22,23). There is no argument that when one views gated SPECT studies from patients with relatively normal hearts, the apex appears to thicken and move more than any other segment, just as Sharir et al. (1) reported. Echocardiography and MRI have shown high in-section endocardial area change as the LV cavity tends to ablate near the apex during ventricular systole (10,14). The tendency toward cavity ablation near the apex, combined with the relatively low spatial resolution of gated SPECT, undoubtedly contributes to the visual impression of relatively high-amplitude thickening and motion in this region. When the gold-standard images are the same as the investigated images, errors such as these may occur and, because they are the same images as those that the experts interpreted, the errors are correlated and accepted as correct. One must remember that the automatic algorithms were trained using ROC analysis and the same expert interpreters' visual analyses of the training population. Although it is possible that such an analysis or validation could be correct, in this case, higher-resolution methods for making these same measurements have proven that the SPECT measurements are wrong. Although the gated SPECT measurements are in error, because they are closely correlated with expert visual interpretations, they will likely be accepted as correct.

Inaccurate or not, quantitative measurements obtained from gated SPECT may still be useful. As long as the quantitative and visual interpretations closely match and the reader understands and recognizes these variations from absolute truth, perhaps they are good enough. Although these gated SPECT measurements may be biased, if normal and abnormal values correlate well with normal and diseased, viable and nonviable, then it may not matter. It might be more comforting if values were accurate in an absolute sense, but perhaps that is unnecessary. Although the higher spatial resolution and more-accurate measurements of wall thickening and motion from other imaging modalities are not rivaled by

those provided by gated SPECT, the perfusion data provided by SPECT are generally not rivaled by the higher-resolution modalities. An extensive volume of work has revealed the diagnostic and prognostic importance of the perfusion assessments alone. Global LV ejection fractions repeatedly have been shown to be accurate. The availability of regional wall function assessments from gated SPECT is not only convenient but, with robust quantitative methods, will be clinically useful and widely applied.

Certainly, it is difficult to argue with the cost savings alone from having perfectly registered assessments of perfusion and function. It has been shown that the retained wall function in regions of abnormal perfusion is highly predictive of defect reversibility and viability (5,24). It is not unreasonable to predict that visual and automatic semiquantitative scoring of regional LV wall function from gated SPECT will prove useful for the assessment of several aspects of heart disease (e.g., regional viability; the effectiveness of various medical, interventional, and surgical therapies; and patient prognosis). Sharir et al. (1) have previously used similar comparisons to validate semiquantitative perfusion scores and summed stress, rest, and difference scores, but comparing computed values with visual scores from the same image is not a comparison with an independent gold standard, whatever the measured variables may be (25–27). Although such a comparison is an appropriate part of software development and a part of the validation process, it is not validation in the usual sense and is certainly not full validation.

The sensitivities and specificities reported in the study by Sharir et al. (1) are misleading. In previous reports, Sharir et al. used a similar approach to develop optimal criteria for perfusion defect identification (25,26,28). However, unlike the study under discussion (1), in previous studies they validated their criteria against a separate patient population with angiographic correlates. Even when this approach is used

to assess myocardial perfusion, a variable that SPECT imaging measures with reasonable accuracy compared with coronary angiography or PET perfusion imaging, it has not always worked very well. In consecutive articles, they reported the development of methods and criteria for quantification of same-day ^{99m}Tc-sestamibi perfusion images (25,29). In the first article, they reported 97% sensitivity and 67% specificity (25). In the second, they reported 87% sensitivity and 36% specificity (29). The sensitivities and specificities for motion (88% and 92%) and thickening (87% and 89%) reported in the current article (1) are not comparisons with independent gold standards but are comparisons with what expert interpreters saw in the same images. Because those interpretations are incorrect, it can be assumed that the automatic scoring algorithms are incorrect as well. The authors have documented closely correlated errors in the current report, which is potentially a significant problem.

Confirmation that a quantitative algorithm provides values comparable with those of expert visual interpretations is a reasonable part of the validation process. It is not surprising that visual and computer-generated values are similar. The training population and ROC analyses were used to teach the algorithm how the experts saw gated SPECT wall motion and wall thickening. However, even experts can read images incorrectly. Reliance on the same images to serve as both the test population and the gold standard is risky. When those images represent a complex interaction between multiple physical principles and physiologic phenomena, they may be misleading. When we look at gated SPECT, we see similar phenomena at the apex and want to believe the correlated errors as truth. It may be proven that, imperfect or not, wall motion and thickening estimated from gated SPECT provide important diagnostic information. It may also be proven, for example, that reliance on inaccurate measurements frequently results in incorrect assessment of regional myocardial viability

and futile attempts at revascularizing irreversibly injured myocardium because of the bias to overestimate regional function in the apical and distal anterior segments typical of the distal left anterior descending coronary artery territory. Considerable additional work remains to prove the clinical utility of regional assessments of regional wall function from gated SPECT.

In conclusion, expert visual interpretation of the same test images is not a gold standard in the usual sense. As applied in the study by Sharir et al. (1), it could turn out to be a misleading standard. Whether their approach proves useful or not, these investigators and others who have developed quantification software for gated myocardial perfusion SPECT (including ourselves) should not be satisfied. It is likely that chamber volume and geometry, reconstruction filters, systolic performance, photon scatter, and respiratory motion influence these assessments. The influence of these factors may not be easily resolved, but accurate, robust solutions should be pursued. While awaiting further progress in this important methodologic approach to regional LV function assessment, it seems clear that the “visual gold standard” can be a risky and potentially inaccurate standard.

James R. Corbett
Edward P. Ficaro

*University of Michigan Medical Center
Ann Arbor, Michigan*

REFERENCES

1. Sharir T, Berman DS, Waechter PB, et al. Quantitative analysis of regional motion and thickening by gated myocardial perfusion SPECT: normal heterogeneity and criteria for abnormality. *J Nucl Med.* 2001;42:1630–1638.
2. Ficaro EP, Kritzman JN, Corbett JR. Automatic segmental scoring of myocardial wall thickening and motion: validation of a new semi-quantitative algorithm [abstract]. *J Nucl Med.* 2001;42(suppl):171P.
3. DePuey EG, Garcia EV, eds. Updated imaging guidelines for nuclear cardiology procedures: part 1. *J Nucl Cardiol.* 2001;8:G5–G58.
4. Port SC, ed. Updated imaging guidelines for nuclear cardiology procedures: part 2. *J Nucl Cardiol.* 1999;6:G53–G83.
5. Chua T, Kiat H, Germano G, et al. Gated technetium-99m sestamibi for simultaneous assessment of stress myocardial perfusion, post exercise regional ventricular function and myocardial viability: correlation with echocardiography and rest

- thallium-201 scintigraphy. *J Am Coll Cardiol*. 1994;23:1107–1114.
6. Taillefer R, DePuey EG, Udelson JE, et al. Comparative diagnostic accuracy of Tl-201 and Tc-99m sestamibi SPECT imaging (perfusion and ECG-gated SPECT) in detecting coronary artery disease in women. *J Am Coll Cardiol*. 1997;29:69–77.
 7. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol*. 1986;21:720–733.
 8. Sheehan FH, Stewart DK, Dodge HT, et al. Variability in the measurement of regional left ventricular wall motion from contrast angiograms. *Circulation*. 1983;68:550–559.
 9. Sheehan FH, Bolson EL, Dodge HT, et al. Advantages and applications of the centerline method for characterizing regional ventricular function. *Circulation*. 1986;74:293–305.
 10. Haendchen RV, Wyatt HL, Maurer G, et al. Quantitation of regional cardiac function by two-dimensional echocardiography. I. Patterns of contraction in the normal left ventricle. *Circulation*. 1983;67:1234–1245.
 11. Assmann PE, Slager CJ, van der Borden SG, et al. Reference systems in echocardiographic quantitative wall motion analysis with registration of respiration. *J Am Soc Echocardiogr*. 1991;4:224–234.
 12. Assmann PE, Slager CJ, Roelandt JR. Systolic excursion of the mitral annulus as an index of left ventricular systolic function. *Am J Cardiol*. 1991;68:829–830.
 13. Beyar R, Shapiro EP, Graves WL, et al. Quantification and validation of left ventricular wall thickening by a three-dimensional volume element magnetic resonance imaging approach. *Circulation*. 1990;81:297–307.
 14. Beyar R, Weiss JL, Shapiro EP, et al. Small apex-to-base heterogeneity in radius-to-thickness ratio by three-dimensional magnetic resonance imaging. *Am J Physiol*. 1993;264:H133–H140.
 15. Young AA, Axel L. Three-dimensional motion and deformation of the heart wall: estimation with spatial modulation of magnetization: a model-based approach. *Radiology*. 1992;185:241–247.
 16. Dong SJ, MacGregor JH, Crawley AP, et al. Left ventricular wall thickness and regional systolic function in patients with hypertrophic cardiomyopathy: a three-dimensional tagged magnetic resonance imaging study. *Circulation*. 1994;90:1200–1209.
 17. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography. I. Effect of object size. *J Comput Assist Tomogr*. 1979;3:299–308.
 18. Buser PT, Auffermann W, Holt WW, et al. Non-invasive evaluation of global left ventricular function with use of cine nuclear magnetic resonance. *J Am Coll Cardiol*. 1989;13:1294–1300.
 19. Germano G, Kiat H, Kavanagh PB, et al. Automatic quantification of ejection fraction from gated myocardial perfusion SPECT. *J Nucl Med*. 1995;36:2138–2147.
 20. Germano G, Erel J, Lewin H, et al. Automatic quantitation of regional myocardial wall motion and thickening from gated technetium-99m sestamibi myocardial perfusion single-photon emission computed tomography. *J Am Coll Cardiol*. 1997;30:1360–1367.
 21. Germano G, Kavanagh P, Waechter P, et al. A new algorithm for the quantitation of myocardial perfusion SPECT. I. Technical principles and reproducibility. *J Nucl Med*. 2000;41:712–719.
 22. Ingels NB, Daughters GT 2nd, Stinson EB, et al. Measurement of midwall myocardial dynamics in intact man by radiography of surgically implanted markers. *Circulation*. 1975;52:859–867.
 23. Ingels NB, Daughters GT 2nd, Stinson EB, et al. Evaluation of methods for quantitating left ventricular segmental wall motion in man using myocardial markers as a standard. *Circulation*. 1980;61:966–972.
 24. Kahn JK, Henderson EB, Akers MS, et al. Prediction of reversibility of perfusion defects with a single post-exercise technetium-99m RP-30A gated tomographic image: the role of residual systolic thickening [abstract]. *J Am Coll Cardiol*. 1988;11:31A.
 25. Van Train KF, Garcia EV, Maddahi J, et al. Multicenter trial validation for quantitative analysis of same-day rest-stress technetium-99m-sestamibi myocardial tomograms. *J Nucl Med*. 1994;35:609–618.
 26. Sharir T, Germano G, Waechter P, et al. A new algorithm for the quantitation of myocardial perfusion SPECT. II. Validation and diagnostic yield. *J Nucl Med*. 2000;41:720–727.
 27. Berman DS, Kang X, Van Train KF, et al. Comparative prognostic value of automatic quantitative analysis versus semiquantitative visual analysis of exercise myocardial perfusion single-photon emission computed tomography. *J Am Coll Cardiol*. 1998;32:1987–1995.
 28. Maddahi J, Van Train K, Prigent F, et al. Quantitative single photon emission computed thallium-201 tomography for detection and localization of coronary artery disease: optimization and prospective validation of a new technique. *J Am Coll Cardiol*. 1989;14:1689–1699.
 29. Van Train KF, Areeda J, Garcia EV, et al. Quantitative same-day rest-stress technetium-99m-sestamibi SPECT: definition and validation of stress normal limits and criteria for abnormality. *J Nucl Med*. 1993;34:1494–1502.

