
Receiver Operating Characteristic Evaluation of Iterative Reconstruction with Attenuation Correction in ^{99m}Tc -Sestamibi Myocardial SPECT Images

Karen J. LaCroix, Benjamin M.W. Tsui, Eric C. Frey, and Ronald J. Jaszcak

Departments of Biomedical Engineering and Radiology, The University of North Carolina at Chapel Hill, Chapel Hill; Department of Radiology, Duke University Medical Center, Durham; and Department of Biomedical Engineering, Duke University, Durham, North Carolina

The purpose of this study was to evaluate differences in myocardial defect detection between ^{99m}Tc -sestamibi myocardial SPECT images reconstructed using conventional filtered backprojection (FBP) without attenuation correction (AC) and those reconstructed using maximum-likelihood expectation maximization with nonuniform attenuation correction (MLAC). **Methods:** An observer study and receiver operating characteristic (ROC) curve analysis were performed using simulated ^{99m}Tc -sestamibi SPECT data from a population of 24 mathematic anthropomorphic torso phantoms, which realistically modeled a wide range of anatomic variations. The phantoms modeled male patients with a flat diaphragm, male patients with a diaphragm raised to the level of the heart, and female patients with large breasts. Transmural, cold defects with a contrast of 0.25 were simulated in the left ventricular wall for 6 locations. Noisy projection data were generated from the phantoms and included the effects of nonuniform attenuation, collimator-detector response, and scatter. The data were then reconstructed using FBP and MLAC. Images were displayed in the short- and long-axis formats, as in clinical practice. Eight observers viewed blocks of FBP and MLAC images and, for each image, indicated on a continuous rating scale the probability that a defect was present. From the rating data, FBP and MLAC ROC curves were generated, and their areas (A_z) were estimated and compared. **Results:** In general, the FBP and MLAC ROC curves did not cross and the MLAC curve showed a higher A_z than did the corresponding FBP curve. For male phantoms with a flat diaphragm, the average difference in A_z was 0.04 and was not statistically significant (at the $P = 0.05$ level) for 6 of 8 observers. For male phantoms with a raised diaphragm, the average difference in A_z was 0.22 and was statistically significant for 6 of 8 observers. For female phantoms with large breasts, the average difference in A_z was 0.19 and was statistically significant for all 8 observers. **Conclusion:** This study showed an improvement in defect detection in myocardial SPECT images using MLAC in comparison with images using FBP without AC, particularly for patients with large breasts or with a diaphragm raised to the level of the heart.

Key Words: cardiac SPECT; attenuation correction; ^{99m}Tc -sestamibi radionuclide; ROC study; phantom study

J Nucl Med 2000; 41:502-513

Myocardial perfusion imaging using SPECT with ^{201}Tl or ^{99m}Tc -sestamibi has shown relatively high sensitivity but low specificity for the diagnosis of coronary artery disease (1-3). The relatively low specificity can be attributed to the artifacts resulting from nonuniform photon attenuation, which is not compensated for in the commonly used reconstruction technique of filtered backprojection (FBP) (1-3). Iterative reconstruction techniques have emerged as an alternative to FBP reconstruction, because the iterative algorithms can use the information provided by patient-specific attenuation maps to compensate for nonuniform attenuation and possibly many other physical factors (4-7). The most popular and widely studied of these iterative algorithms is the maximum-likelihood expectation maximization (MLEM) algorithm (8,9), which models the Poisson noise inherent in single-photon data.

Only 2 published studies have compared iteratively reconstructed attenuation correction (AC) images and FBP images by measuring human observer performance at a detection task, and those 2 studies had apparently conflicting results. Gilland et al. (10) compared FBP without AC and MLEM both with and without nonuniform AC using simulated ^{201}Tl myocardial SPECT. This study showed no improvement in myocardial defect detection as a result of applying AC in the reconstruction. However, the simulated images shown to the observers were not representative of typical clinical myocardial perfusion images with respect to the defect size, number of phantom anatomies, postreconstruction filtering, and image display methods used. More recently, Ficaro et al. (11) compared FBP without AC to iterative penalized weighted least squares with nonuniform AC (12) using clinical ^{99m}Tc -sestamibi myocardial SPECT images. The results of

Received Nov. 12, 1998; revision accepted Jul. 30, 1999.
For correspondence or reprints contact: Karen J. LaCroix, PhD, Department of Biomedical Engineering, CB 7575, 152 MacNider, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599.

this study showed a statistically significant improvement in defect detection for the penalized weighted least squares images with AC compared with the FBP images without AC. In the study of Ficaro et al., the patient population was relatively small and possibly not representative of populations in other clinics. Furthermore, because clinical images were used, the presence of a true perfusion defect could not be known with certainty in all cases. Thus, the accuracy of the sensitivity, specificity, and receiver operating characteristic (ROC) curve estimates was, to some degree, unknown. Because of the apparently conflicting results of the studies of Gilland et al. and Ficaro et al. and the limitations in both studies, further human observer studies were needed in this area.

The purpose of this study was to compare ^{99m}Tc -sestamibi myocardial SPECT images reconstructed using MLEM with nonuniform AC (MLAC) and those reconstructed using FBP without AC for detecting myocardial defects using human observers. This study has a general purpose similar to that of the 2 previous studies but pursues that purpose in a context different from that of either to create a broader understanding of the relative differences between these 2 types of reconstruction techniques.

Unlike the study of Ficaro et al. (11), this study used simulated rather than clinical SPECT images because of the following benefits: the presence of a defect was known with certainty; many simulated images could be obtained in far less time and using far fewer resources, allowing better statistical precision of the results; and the presence and location of perfusion defects and the anatomy of patients could be systematically varied, allowing a clearer evaluation of the types of patients for whom AC may improve defect detection. In addition, this study used an iterative reconstruction algorithm different from, but comparable with, that used in the study of Ficaro et al.

This study is similar to the study of Gilland et al. (10) in its use of simulated SPECT images and its observer study methodology but differs in several important aspects. First, the SPECT images were simulated from a large and anatomically diverse population of mathematic phantoms, resulting in a set of simulated SPECT images varying broadly in the extent and severity of attenuation artifacts. This population specifically included phantoms modeling patients with large breasts or a raised diaphragm, because breast and diaphragmatic attenuation are often singled out as the causes of especially severe attenuation artifacts in SPECT images reconstructed using FBP without AC (2,13–17). Second, this study modeled large, diffuse perfusion defects as opposed to small, focal defects. The detection of large defects is more relevant to clinical diagnosis of coronary artery disease because physicians often require that an area of decreased count density be of a certain minimum size—typically at least 3 contiguous slices—to qualify as a physiologically significant defect. These large defects are likely to be difficult to distinguish from attenuation artifacts

because attenuation artifacts also tend to be large and diffuse. Third, the reconstructed images were filtered and displayed in a manner similar to that used for clinical myocardial SPECT images.

MATERIALS AND METHODS

Three-Dimensional Mathematic Phantom Population

Realistic SPECT projection data were simulated using the 3-dimensional mathematic cardiac torso (MCAT) phantom with a dynamic, beating heart (18,19). For this study, the MCAT phantom was upgraded from previous versions to model a more realistic lung shape and variations in torso anatomy (20). A population of 24 MCAT phantoms was created to model 8 male patients with a flat diaphragm, 8 male patients with a raised diaphragm, and 8 female patients with large breasts and a flat diaphragm. The last 2 anatomic classes were formed separately because severe attenuation artifacts are generally attributed to breast or diaphragmatic attenuation (13–17). All phantom anatomies were based on PET transmission images of actual patients, taken primarily from the PET Thorax Model Database (21).

Four of the 8 phantoms modeling typical male anatomies with a flat diaphragm are shown in Figure 1. The anatomy of the phantoms varied with respect to the size and shape of the torso, the size and shape of the rib cage and lungs, and the thickness of the chest wall. The variation in torso shape can be appreciated by comparing phantoms A and C. Phantom C models a round torso shape in which the eccentricity (ratio of lateral width to anteroposterior width) of the outer boundary is nearly equal to 1, whereas phantom A models a flat torso shape in which the eccentricity of the outer boundary is much greater than 1. The variation in lung size and chest wall thickness can be appreciated by comparing phantom B, which has smaller lungs and thicker chest walls, and phantom D, which has larger lungs and thinner chest walls.

Four of the 8 phantoms modeling the typical male anatomies with a raised diaphragm are shown in Figure 2. The phantoms in Figures 1 and 2 were based on the same human patient images (top row of Fig. 1). The only anatomic difference between the phantoms in Figures 1 and 2 is the shape of the liver and diaphragm, which can best be appreciated by comparing the simulated chest radiographs in the bottom row of Figures 1 and 2.

Four of the 8 phantoms modeling typical female anatomies with large breasts are shown in Figure 3. The female phantoms vary with respect to the size, shape, and position of the breasts and the size and shape of the torso and lungs. All female phantoms were modeled with a flat diaphragm.

For this study, the assumed ^{99m}Tc -sestamibi relative organ uptakes per unit volume were 38, 38, 2, 30, 38, and 1 for the myocardium, liver, lungs, spleen, kidneys, and background, respectively. The background included all organs not specifically listed. These estimates were based on the average organ concentration ratios estimated by Gilland et al. (22), except for the liver, for which uptake was simulated at roughly twice that of the estimated average. The high liver uptake modeled in this study is not uncommon with ^{99m}Tc -sestamibi and is a cause of inferoseptal artifacts that possibly degrade defect detection in images that are reconstructed with AC (but not with scatter and collimator–detector response compensation) (11,23,24). We chose to model this high liver uptake to add difficulty to the observers' detection task.

Given the variability in coronary artery anatomy, the location of arterial blocks, and the degree of collateral circulation from patient

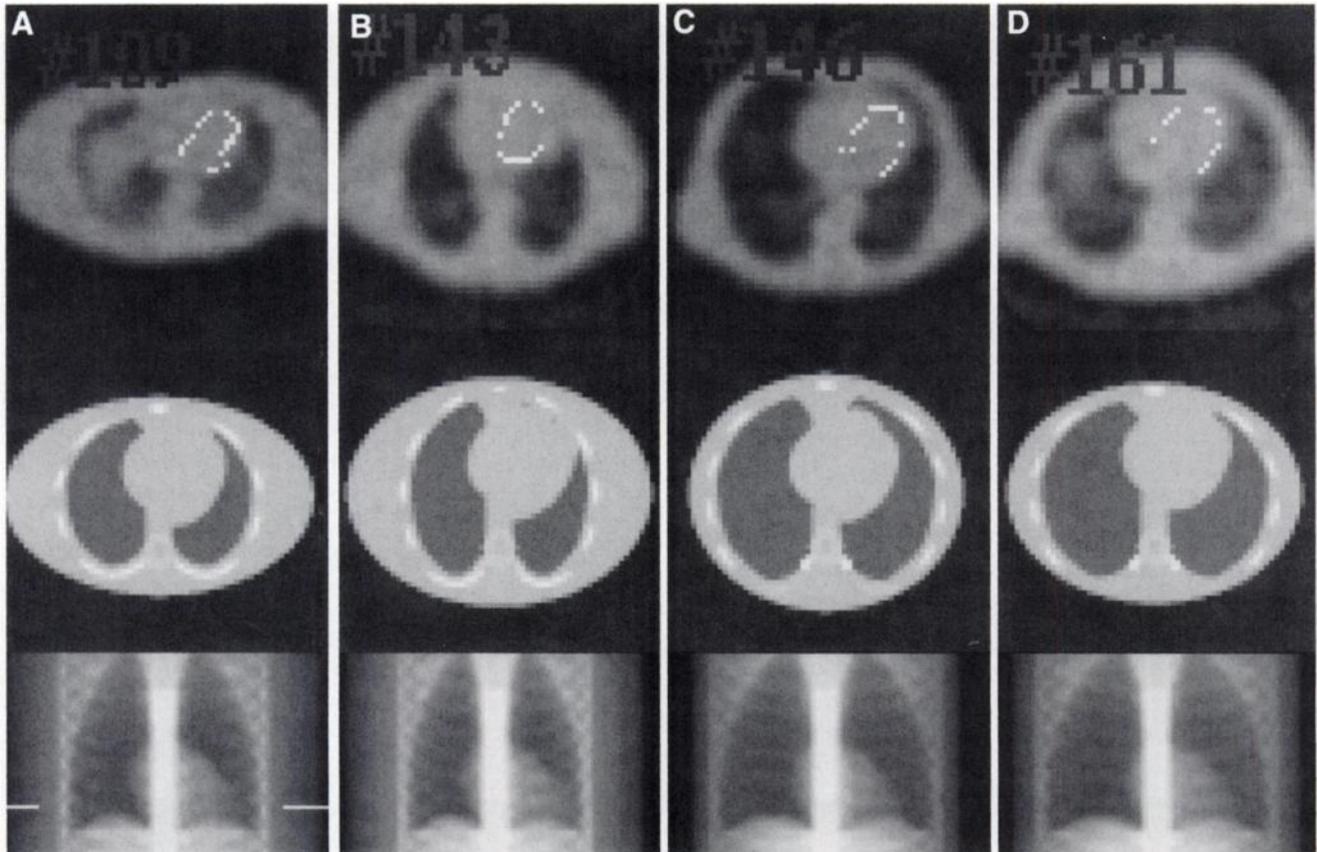


FIGURE 1. (A–D) Shown are 4 of 8 male MCAT phantoms with flat diaphragm anatomy. Top row shows single transaxial slice of attenuation map of actual patients on whom phantoms were based. Middle row shows corresponding transaxial slice of MCAT phantom attenuation map. Bottom row shows simulated chest radiograph (anteroposterior transmission projection image) for each MCAT phantom. Horizontal lines drawn on chest radiograph of phantom A show location of transaxial slices.

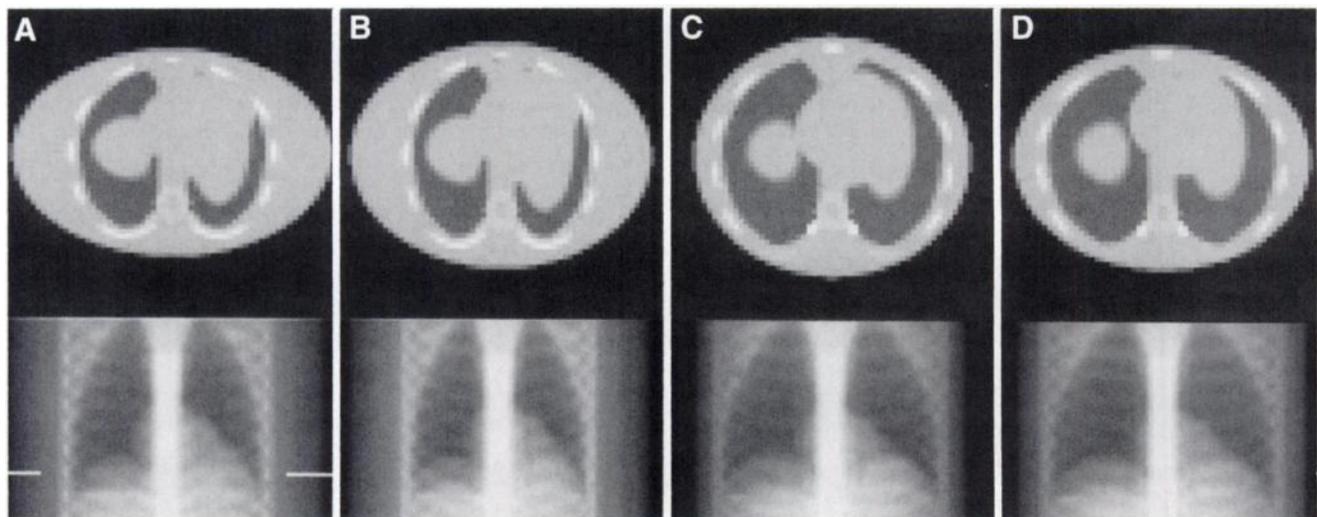


FIGURE 2. (A–D) Shown are 4 of 8 male MCAT phantoms with raised diaphragm and liver anatomy. These MCAT phantom anatomies are identical to those shown in Figure 1, except for liver and diaphragm shape. Top row shows transaxial slices of attenuation maps. Bottom row shows simulated chest radiograph. Horizontal lines drawn on chest radiographs of phantom A show location of transaxial slices.

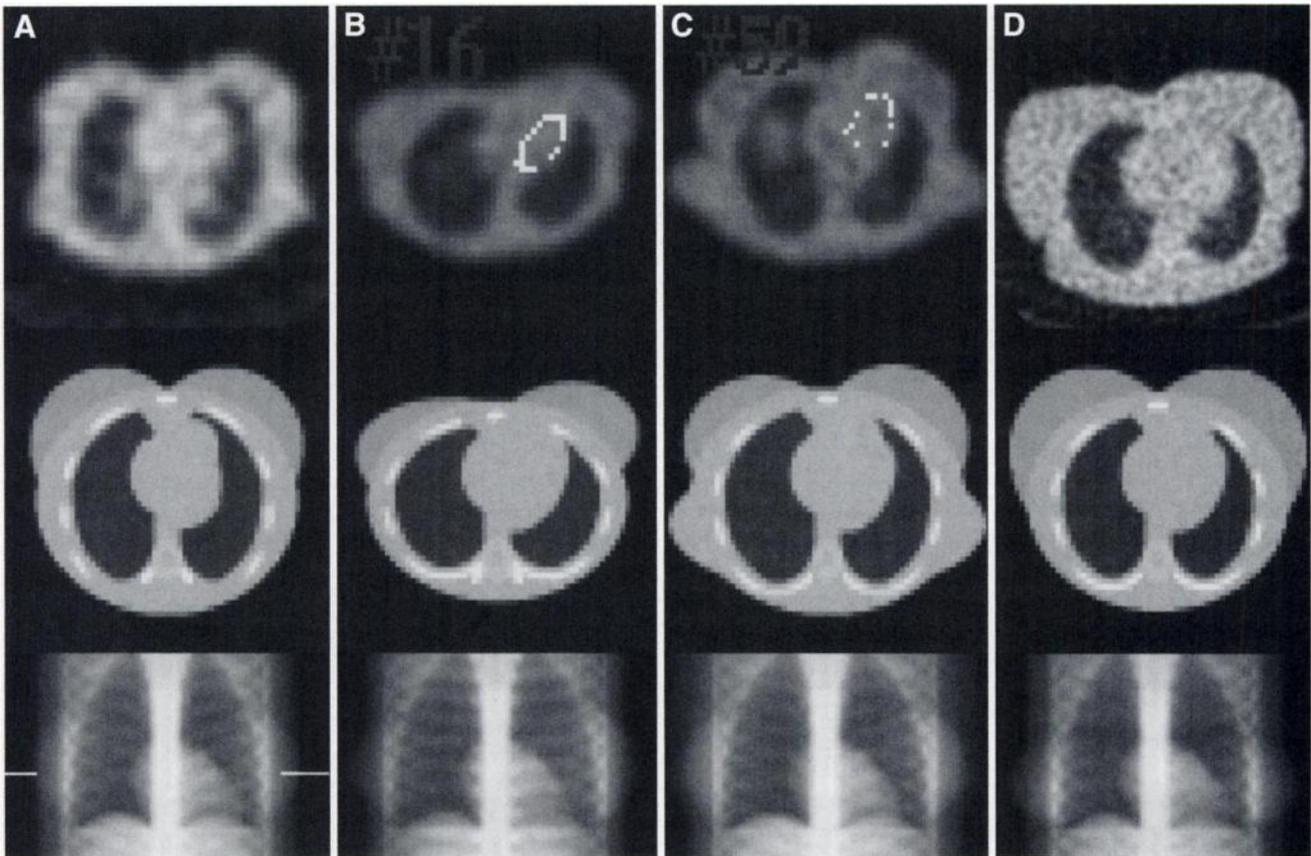


FIGURE 3. (A–D) Shown are 4 of 8 female MCAT phantoms with large breasts and flat diaphragm anatomy. Top row shows single transaxial slice of attenuation map of each patient on whom phantoms were based. Middle row shows corresponding transaxial slice of MCAT phantom attenuation map. Bottom row shows simulated chest radiograph (anteroposterior transmission projection image) of each MCAT phantom. Horizontal lines drawn on chest radiograph of phantom A show location of transaxial slices.

to patient, myocardial perfusion defects can occur in a variety of locations, shapes, and sizes (25). As a result, there were no specific guidelines for modeling myocardial defects. Four of the 6 defects used in this study are shown in Figure 4. All defects were transmural, and their approximate volume varied from 16% to 22% of the total volume of the left ventricular myocardium. They spanned at least 3 contiguous slices (1.9 cm) in both the short- and long-axis orientations. The defect location was varied throughout the left ventricular wall, except that no defects were placed at the apex or in the lateral left ventricular wall, because preliminary studies showed that defect detection was too difficult in the apex (as is often the case in clinical images) and too easy in the lateral wall

regardless of reconstruction technique. All defects were simulated at a contrast of 0.25, that is, the uptake in the defect was 75% of normal in the left ventricular wall. This contrast was chosen because it gave the desired difficulty for observer detection while still being at a level that physicians consider physiologically significant.

Simulation of Emission Projection Data

Noise-free single-photon emission projection data were simulated with the effects of nonuniform attenuation, collimator–detector blurring, and scatter. Because of the large number of projection data sets that needed to be simulated, an analytic

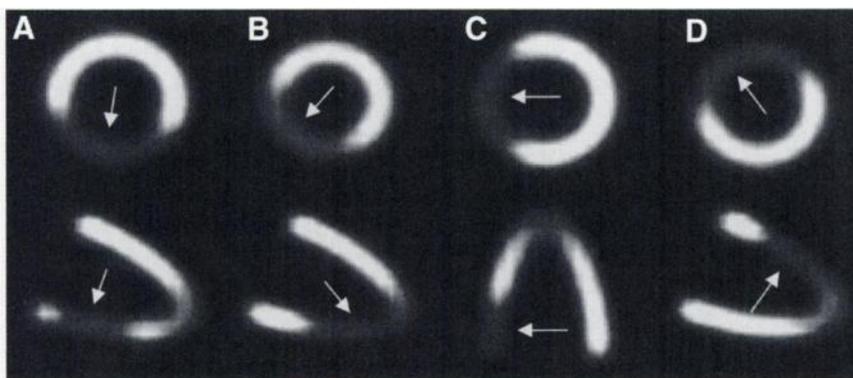


FIGURE 4. Size and location of 4 of 6 defects used in this study are shown. Top row shows short-axis slices and bottom row shows either vertical long-axis (A, B, D) or horizontal long-axis (C) slice. Defect location is indicated by arrows.

projector, as opposed to a computationally intensive Monte Carlo simulation, was used to simulate the noise-free projection data. The analytic projector simulated the effect of scattered photons using a slab-derived scatter model (26,27), assuming a γ camera energy resolution of 10% and a 20% energy window centered on the 140-keV ^{99m}Tc photopeak. Preliminary simulations showed the differences between Monte Carlo (SimSET code (28)) and analytic projection data to be negligible.

Projection data were generated for 64 angular views over a 180° arc extending from 45° right anterior oblique to 45° left posterior oblique. For each view, the simulated 2-dimensional projection data were generated for 128 bins and 82 slices with a bin width and slice thickness of 0.31 cm. The projection data were then collapsed to 64 bins and 41 slices with a bin width and slice thickness of 0.62 cm. The radius of rotation was 22.5 cm. Collimator-detector response was simulated assuming a low-energy high-resolution parallel-hole collimator with a thickness of 4.1 cm and a circular hole diameter of 0.19 cm and assuming an intrinsic resolution of 0.40 cm, full width at half maximum, for the detector.

Poisson noise was simulated in the scaled, noise-free data. Noise-free data were scaled assuming the same total acquisition time for all phantoms, resulting in a variable number of acquired counts among data sets, depending on the anatomy of the phantom. Preliminary observer studies determined the image noise level for the final study. At the count level of clinical ^{99m}Tc -sestamibi data ($\sim 200,000$ counts per myocardial slice), defect detection in the MLAC images was trivial; that is, the observer performance was 100% accurate. To perform a meaningful ROC analysis with adequate statistical power, the noise level was increased by scaling the projection data to approximately 25,000–40,000 total counts in the cardiac slices for the final observer study. The high noise level did not significantly alter the task or results.

Image Reconstruction, Processing, and Display

Reconstruction without AC was performed using the conventional FBP method with a ramp filter. The iterative reconstruction with nonuniform AC was performed using MLEM for 30 iterations with a uniform initial estimate. All reconstructed images were filtered afterward with a stationary Butterworth filter of order 5 and a cutoff of 0.25 cycles/pixel (0.4 cycles/cm), which is typical of that used in clinical ^{99m}Tc -sestamibi myocardial SPECT studies (29–31). The stopping point of 30 iterations was chosen because the filtered MLAC images did not change appreciably beyond 30 iterations. All reconstructed images had a pixel width and slice thickness of 0.62 cm. No compensations for scatter or detector response were used in either the FBP or the MLAC reconstructions.

The phantom-specific attenuation maps used in the MLAC reconstructions were obtained by blurring the narrow-beam attenuation coefficient map for each phantom with a gaussian filter of 0.7 cm (~ 1.1 pixels), full width at half maximum, which is on the order of the simulated intrinsic resolution of the camera. Because the noise in SPECT images is dominated by the emission data statistics (32), and because the scatter in properly collimated transmission data is relatively low (33), this method produces attenuation maps that are a reasonable approximation for reconstructing simulated SPECT data. The resulting attenuation maps have roughly the same spatial resolution as the emission data to avoid resolution-mismatch artifacts (34,35).

All reconstructed images were reformatted from transaxial slices to short- and long-axis slices. Any negative values in the images were set to 0, and images were individually scaled for display, as is

typically done for clinical images. For the final display, a square region of interest encompassing the heart and the top of the liver was extracted from each short- and long-axis slice and subsequently scaled to 4 times its original area by bilinear interpolation. The observers were shown 12 short-axis slices that encompassed the entire left ventricular wall, as well as 3 vertical long-axis and 3 horizontal long-axis slices. Each slice was approximately 2.5×2.5 cm when displayed on the monitor.

The images were displayed on a Diamondscan 53-cm (21-in.) color monitor (Mitsubishi Electronics America, Somerset, NJ) in gray scale using 256 levels of gray. The luminance of the monitor for each of the 256 display-driving intensities was measured with a photometer (20). The display settings were held constant during the course of the observer study, and the observers were instructed not to adjust the display. A crosshair was used to indicate the myocardial section where a defect, if present, was located. The observer could remove or redisplay the crosshairs as desired but was required, for each image, to remove the crosshairs at least once before entering a rating.

Observer Study Methodology and ROC Analysis

This observer study and ROC analysis followed the basic methodology established by Metz (36). The 8 observers of this study included 2 board-certified nuclear medicine physicians at the University of North Carolina Memorial Hospital, 2 doctoral faculty members, 1 postdoctoral research associate, and 3 graduate students in the Department of Biomedical Engineering at the University of North Carolina at Chapel Hill.

All observers were read a set of instructions that explained the context of the observer study, the reconstruction techniques, and image processing information that would be readily available to physicians. Each observer was instructed to rate, for each image, the probability that a perfusion defect was present in the section of the left ventricular wall indicated by the crosshair. The rating scale was essentially continuous (256 levels) and ranged from 1 to 5, with 1 corresponding to “defect definitely absent” and 5 corresponding to “defect definitely present.” The observers were told which reconstruction technique was used to create each displayed image but were not told anything about the patient’s anatomy, because anatomic information is not always available to physicians.

The observer study was split into 2 sessions, each of which consisted of 288 training and 288 test images and lasted 1–2 h. Sessions were separated by at least 24 h but no more than 72 h. As shown by Table 1, the FBP and MLAC images were presented in blocks. Each session consisted of 2 initial long training blocks of 72 images and then 4 training–test blocks of 36 training images followed by 72 test images. Observers were allowed to take short breaks between blocks.

To eliminate reading-order effects, the FBP and the corresponding MLAC images from the same set of projection data were shown in separate sessions, and the order of the blocks was different for each observer and for each session, as shown in Table 1. Within each block, the images were randomized differently for each observer, and an equal number of cases was selected from each of the 3 anatomic classes and from phantoms with defects present and defects absent.

From the experimental rating data, ROC curves were generated using the CLABROC program to estimate parameters of the binormal ROC curve that best fit the experimental rating data (37). The binormal ROC curve model assumes that ratings for the defect-present and defect-absent images form 2 normal distribu-

TABLE 1
Image Block Orders for Sessions 1 and 2 for the 8 Observers

Session 1						Session 2					
Train		Train/ test	Train/ test	Train/ test	Train/ test	Train		Train/ test	Train/ test	Train/ test	Train/ test
FB	ML	FB1	ML3	ML4	FB2	ML	FB	ML2	FB4	FB3	ML1
FB	ML	FB2	ML4	ML3	FB1	ML	FB	ML1	FB3	FB4	ML2
FB	ML	FB3	ML1	ML2	FB4	ML	FB	ML4	FB2	FB1	ML3
FB	ML	FB4	ML2	ML1	FB3	ML	FB	ML3	FB1	FB2	ML4
ML	FB	ML1	FB2	FB3	ML4	FB	ML	FB4	ML3	ML2	FB1
ML	FB	ML2	FB4	FB1	ML3	FB	ML	FB3	ML1	ML4	FB2
ML	FB	ML3	FB1	FB4	ML2	FB	ML	FB2	ML4	ML1	FB3
ML	FB	ML4	FB3	FB2	ML1	FB	ML	FB1	ML2	ML3	FB4

FBn = block of FBP images from group n; MLn = block of MLAC images from group n.
FBP and ML blocks with same group number, n, contain images reconstructed from same sets of projection data.

tions, each with its own mean and SD. The binormal ROC curve depends only on the relative differences between the 2 normal distributions and, as a result, is characterized by only 2 parameters. These are commonly denoted "A," which is the difference in the means of the 2 normal distributions divided by the SD of 1 normal distribution, and "B," which is the ratio of the SDs of the 2 normal distributions. For each of the 8 observers, 6 ROC curves were estimated (2 reconstruction techniques \times 3 anatomic classes). Each ROC curve was estimated from the rating data of 48 defect-present images (6 defect locations \times 8 anatomies \times 1 noise realization) and 48 defect-absent images (8 anatomies \times 6 noise realizations). The estimated parameters, A and B, were averaged over all observers to produce the average ROC curves for each reconstruction technique and anatomic class.

From the estimated A and B values, the estimated area under the ROC curve (A_z) was calculated. The A_z was used as an index for comparing the difference in defect detection between the FBP and corresponding MLAC images. For each anatomic class, 2 types of statistical tests were performed. First, for each observer, the difference in A_z between the FBP and the MLAC ROC curves was tested for statistical significance using the z score with the null hypothesis that the difference in A_z between the FBP and MLAC ROC curves was 0. This test was provided by the CLABROC program. Second, combining the results from all 8 observers, the average difference in A_z between the FBP and the MLAC ROC curves was tested for statistical significance using the paired 2-tailed Student's t test with the null hypothesis that the difference in A_z between the FBP and MLAC ROC curves was 0. The A_z values from FBP and MLAC ROC curves were paired for each of the 8 observers; thus, the t test had 7 degrees of freedom. For both the z score and the t test, the difference in A_z was considered to be significant if P was less than or equal to 0.01, marginally significant if P was greater than 0.01 but not more than 0.05, and not significant if P was greater than 0.05.

RESULTS

Reconstructed Images

For each anatomic class, representative short-axis slices of noisy FBP and MLAC reconstructed images are shown in Figures 5 and 6, respectively. The first column in both figures shows myocardial uptake in the corresponding slice

of the male phantom with a flat diaphragm. Although not shown, the phantoms for the other 2 anatomic types had the same myocardial uptake as for the phantoms in the first column.

The images in Figure 5 show the relative location and severity of attenuation artifacts in the FBP images. In male phantoms with a flat diaphragm, when perfusion was normal the reconstructed count density in the left ventricular wall was relatively uniform (B), and when perfusion defects were present they were generally apparent in the reconstructed images (F and J). In male phantoms with a raised diaphragm, the diaphragmatic attenuation caused an artifactual decrease in count density in the inferior left ventricular wall. As a result, all images (C, G, and K) exhibited an apparent defect in the inferior left ventricular wall, even though an inferior wall defect is actually present in only 1 of the images (K). In female phantoms with large breasts, the breast attenuation caused an artifactual decrease in count density of varying severity in the anterior wall of the left ventricle. Images D and H show an apparent perfusion defect in the anterior wall, even though an anterior wall defect is actually present in only 1 of the images (H). In image L, the anterior wall attenuation artifact offsets the inferior wall defect, causing what appears, falsely, to be normal uptake.

The MLAC images in Figure 6 were reconstructed from the same sets of projection data as were the corresponding FBP images in Figure 5. For all 3 anatomic classes, when perfusion was normal the reconstructed count density in the left ventricular wall was relatively uniform (top row), but a roughly 10% artifactual increase in count density in the inferoseptal wall occurred because of scatter and collimator-detector effects. When perfusion defects were present (bottom 2 rows), the reconstructed images generally showed a decrease in count density in the defect location, compared with other regions of the left ventricular wall. However, because of statistical noise, some images (image L) showed apparently uniform count density in the presence of a defect.

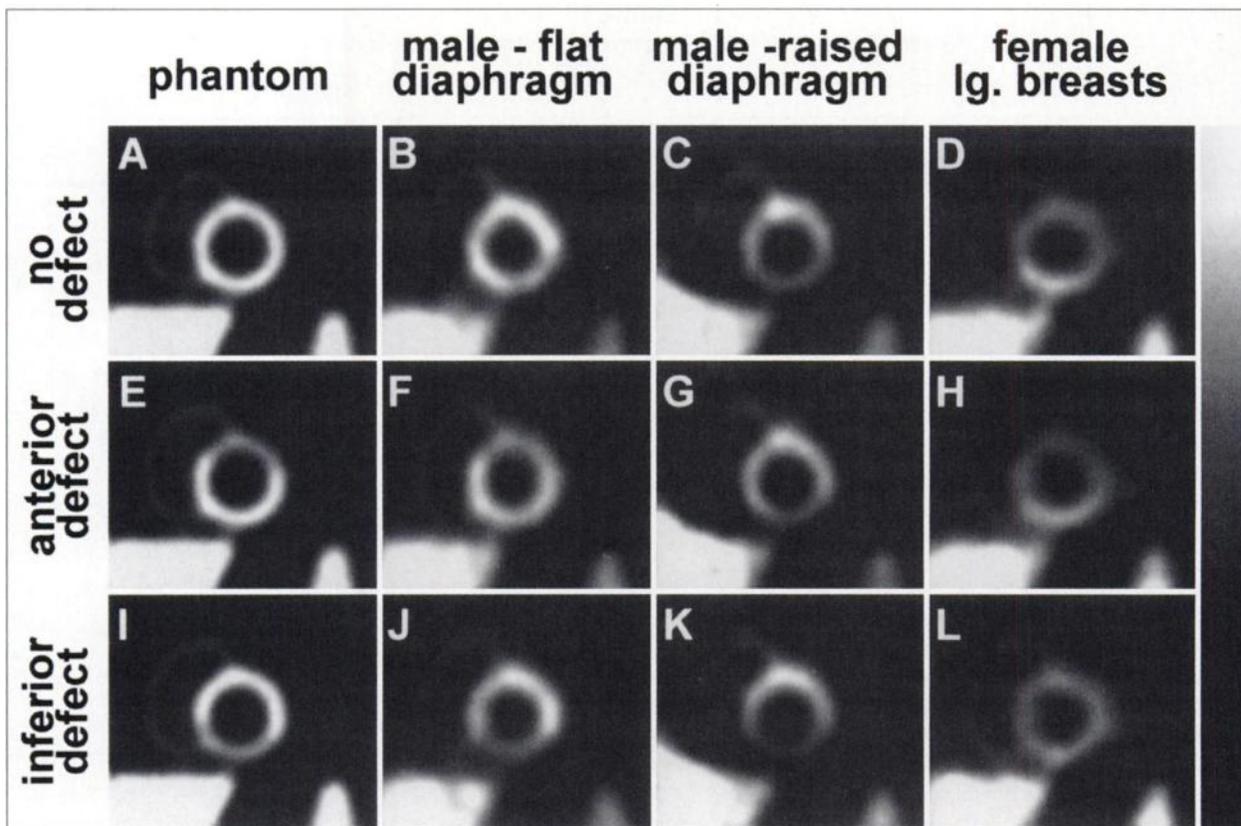


FIGURE 5. (A-L) Sample short-axis slices of images reconstructed using FBP without AC.

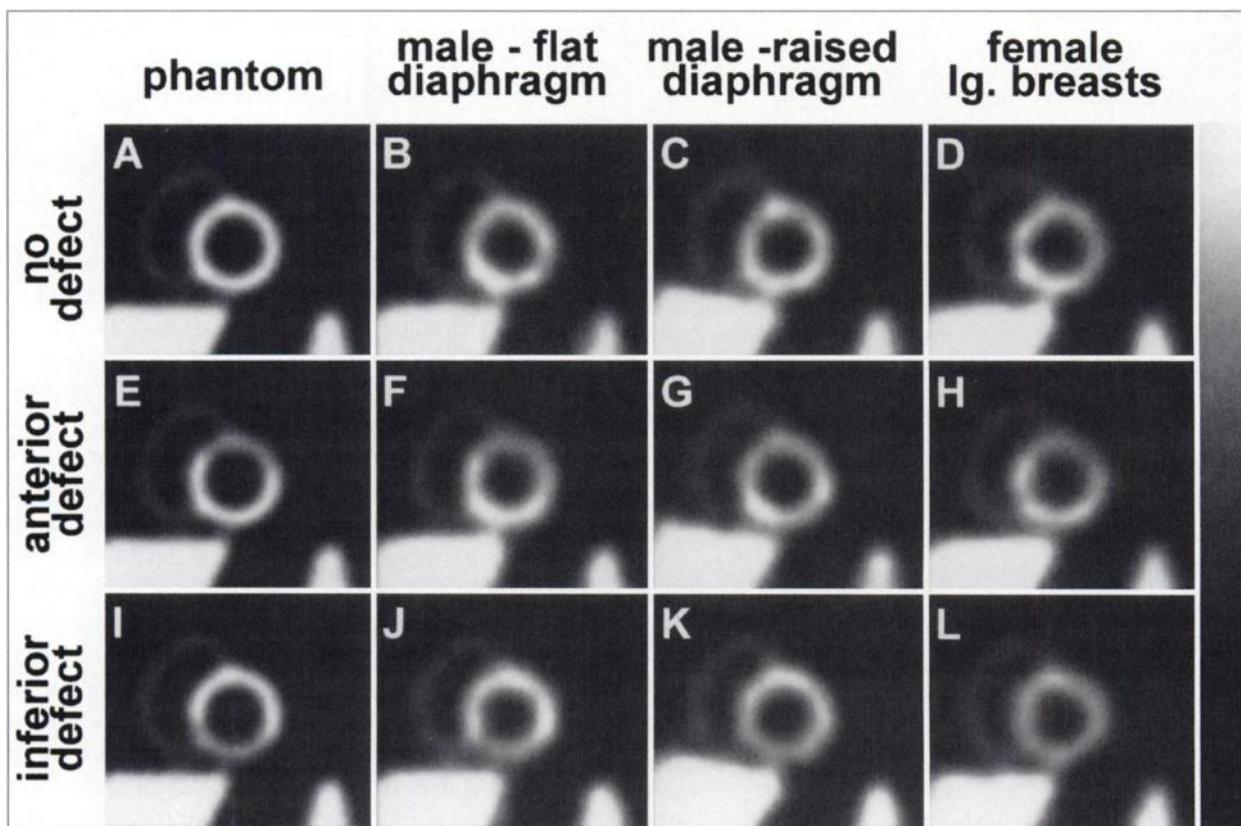


FIGURE 6. (A-L) Sample short-axis slices of images reconstructed using MLAC.

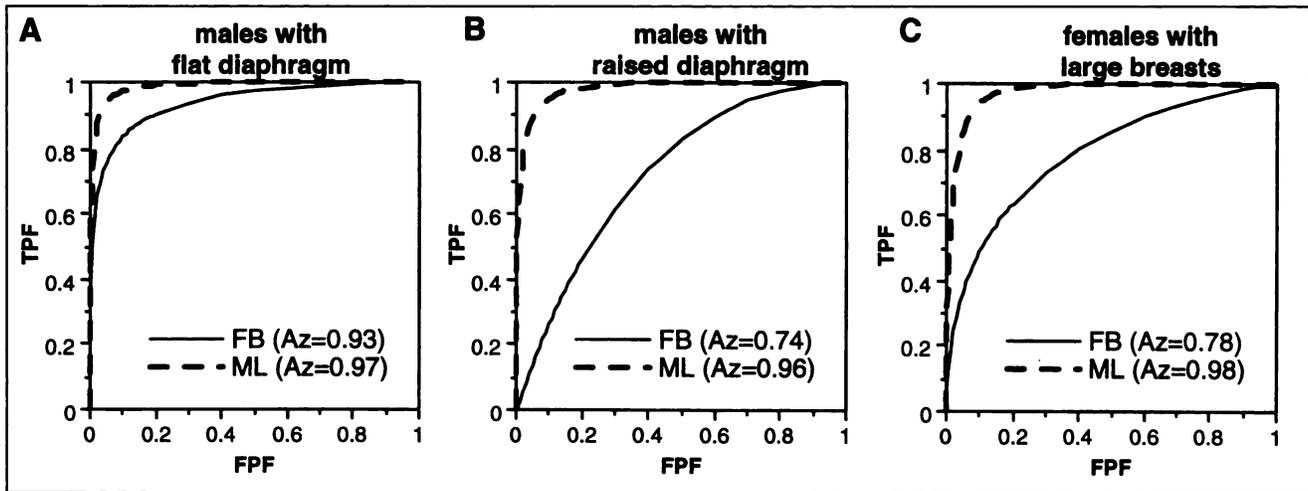


FIGURE 7. (A–C) ROC curves averaged across all 8 observers, for each anatomic class and reconstruction technique.

ROC Curves and Analysis

For each anatomic class, Figure 7 shows the average ROC curves, obtained from averaging the individual ROC curves of all observers. Table 2 lists the estimated binormal curve

parameters, A and B, of the fitted ROC curves for the individual observers and the average across observers. For all 3 anatomic classes, the MLAC ROC curve is higher and to the left of the corresponding FBP ROC curve. In addition,

TABLE 2
 Estimated Binormal ROC Curve Parameters, A and B

Type of phantom	Observer no.	FB				MLAC			
		A	SD _A	B	SD _B	A	SD _A	B	SD _B
Male, flat diaphragm	1 (MD)	1.654	0.286	0.356	0.131	4.737	1.851	1.556	0.957
	2 (MD)	1.127	0.231	0.820	0.171	1.583	0.360	1.565	0.359
	3 (GS)	1.888	0.330	0.895	0.220	1.854	0.313	0.607	0.175
	4 (PhD)	1.887	0.327	0.856	0.213	3.466	1.007	0.899	0.500
	5 (GS)	2.151	0.393	1.122	0.280	3.455	0.996	0.928	0.507
	6 (PhD)	2.122	0.375	0.646	0.208	3.927	1.640	0.915	0.766
	7 (PhD)	2.935	0.636	1.003	0.372	3.862	1.642	0.883	0.765
	8 (GS)	1.651	0.275	0.266	0.106	2.892	0.703	0.732	0.361
	Average	1.927	0.133	0.746	0.080	3.222	0.407	1.011	0.204
Male, raised diaphragm	1 (MD)	0.704	0.257	2.835	0.251	2.835	0.635	1.192	0.414
	2 (MD)	0.719	0.234	1.099	0.229	1.173	0.234	0.810	0.168
	3 (GS)	1.275	0.225	0.488	0.120	2.329	0.435	1.113	0.299
	4 (PhD)	0.573	0.184	0.601	0.113	4.056	1.674	0.974	0.755
	5 (GS)	0.822	0.199	0.668	0.127	2.833	0.585	1.055	0.353
	6 (PhD)	0.868	0.288	1.490	0.265	3.110	0.725	1.053	0.416
	7 (PhD)	1.421	0.305	1.293	0.254	4.215	1.992	1.045	0.925
	8 (GS)	1.220	0.314	1.448	0.280	3.944	1.109	1.595	0.667
	Average	0.950	0.092	1.240	0.076	3.062	0.328	1.105	0.176
Female, large breasts	1 (MD)	1.776	0.319	0.966	0.230	2.408	0.550	2.408	0.159
	2 (MD)	1.221	0.270	1.145	0.219	1.862	0.316	0.680	0.186
	3 (GS)	1.132	0.245	0.976	0.184	4.264	1.165	2.111	0.774
	4 (PhD)	1.076	0.216	0.678	0.137	2.400	0.522	0.170	0.191
	5 (GS)	0.793	0.197	0.649	0.124	2.570	0.501	1.086	0.326
	6 (PhD)	1.051	0.224	0.808	0.156	3.907	1.133	1.448	0.649
	7 (PhD)	0.868	0.207	0.743	0.138	2.654	0.648	0.315	0.299
	8 (GS)	0.624	0.206	0.877	0.149	6.931	3.862	3.071	2.068
	Average	0.950	0.092	1.240	0.076	3.062	0.328	1.105	0.176

MD = nuclear medicine physician; GS = graduate student in biomedical engineering; PhD = doctoral faculty or postdoctoral research associate in biomedical engineering.

TABLE 3
Estimated A_z and Results of Statistical Comparison of Differences in A_z between FBP and MLAC for Individual Observers

Type of phantom	Observer no.	FB		MLAC		FBP vs. MLAC comparison of A_z	
		A_z	SD_{A_z}	A_z	SD_{A_z}	$A_{z,ML} - A_{z,FB}$	P_z score
Male, flat diaphragm	1 (MD)	0.940	0.028	0.998	0.005	0.058	0.0527
	2 (MD)	0.807	0.044	0.803	0.047	-0.004	0.9467
	3 (GS)	0.920	0.027	0.943	0.023	0.023	0.4694
	4 (PhD)	0.924	0.026	0.995	0.005	0.071	0.0081*
	5 (GS)	0.924	0.026	0.994	0.005	0.070	0.0072*
	6 (PhD)	0.963	0.018	0.998	0.002	0.035	0.0507
	7 (PhD)	0.981	0.011	0.998	0.002	0.017	0.1083
	8 (GS)	0.945	0.028	0.990	0.008	0.045	0.1089
	Average	0.926	0.010	0.965	0.007	0.039	—
Male, raised diaphragm	1 (MD)	0.667	0.056	0.966	0.016	0.299	0.0000*
	2 (MD)	0.686	0.056	0.819	0.043	0.133	0.0488†
	3 (GS)	0.874	0.038	0.940	0.023	0.066	0.1105
	4 (PhD)	0.688	0.055	0.998	0.002	0.310	0.0000*
	5 (GS)	0.753	0.050	0.974	0.013	0.181	0.0000*
	6 (PhD)	0.686	0.054	0.984	0.010	0.298	0.0000*
	7 (PhD)	0.808	0.044	0.998	0.002	0.190	0.0000*
	8 (GS)	0.756	0.050	0.982	0.012	0.226	0.0000*
	Average	0.740	0.017	0.958	0.007	0.218	—
Female, large breasts	1 (MD)	0.899	0.031	0.992	0.012	0.093	0.0024*
	2 (MD)	0.789	0.046	0.938	0.024	0.149	0.0053*
	3 (GS)	0.791	0.045	0.966	0.019	0.175	0.0003*
	4 (PhD)	0.813	0.044	0.991	0.012	0.178	0.0000*
	5 (GS)	0.747	0.050	0.959	0.018	0.212	0.0001*
	6 (PhD)	0.793	0.045	0.987	0.009	0.194	0.0000*
	7 (PhD)	0.757	0.049	0.994	0.008	0.237	0.0000*
	8 (GS)	0.681	0.054	0.984	0.014	0.303	0.0000*
	Average	0.784	0.016	0.976	0.015	0.192	—

*Considered significant.

†Considered marginally significant.

MD = nuclear medicine physician; GS = graduate student in biomedical engineering; PhD = PhD faculty or postdoctoral research associate in biomedical engineering.

the FBP and the MLAC ROC curves do not cross, indicating that at any given sensitivity, the specificity for MLAC is higher than or equal to that for FBP without AC.

Table 3 lists the estimated A_z values of the individual observer and average ROC curves. For MLAC, the A_z was consistently high (>0.95) for all 3 anatomic classes; however, for FB, the A_z was relatively high (0.93) for male phantoms with a flat diaphragm but was relatively low (0.74–0.79) for male phantoms with a raised diaphragm and female phantoms with large breasts. On average, the differ-

ence in A_z between the FBP and MLAC ($A_{z,ML} - A_{z,FB}$) ROC curves was 0.04, 0.22, and 0.19 for the male, flat diaphragm; male, raised diaphragm; and female phantoms, respectively. These results are consistent with the apparent severity of the artifacts seen in the short-axis images in Figures 5 and 6.

Table 3 lists the results of statistical testing for the differences in A_z between the MLAC and FBP ROC curves for individual observers. Table 4 lists the results of statistical testing for the differences in A_z between the MLAC and FBP ROC curves averaged across all observers. For the indi-

TABLE 4
Results of Statistical Comparison of Differences in A_z between FBP and MLAC Across Observers

Type of phantom	Average ΔA_z	SD of ΔA_z	t statistic (df = 7)	P (df = 7)	95% confidence interval
Male, flat diaphragm	0.039	0.027	4.17	0.0042*	(0.0171, 0.0617)
Male, raised diaphragm	0.218	0.087	7.11	0.0002*	(0.1454, 0.2904)
Female, large breasts	0.192	0.062	8.79	0.0000*	(0.1408, 0.2444)

*Considered significant.

df = degrees of freedom; $\Delta A_z = A_{z,ML} - A_{z,FB}$.

vidual observer ROC curves for male phantoms with a flat diaphragm, the difference in A_z between FBP and MLAC ROC curves was not significant for 6 of 8 observers. In male phantoms with a raised diaphragm, the difference in A_z was significant for 6 of 8 observers, was marginally significant for 1 observer, and was not significant for 1 observer. In the case of female phantoms with large breasts, the difference in A_z was significant for all observers. When averaged across observers, the difference in A_z between the FBP and MLAC ROC curves was significant for all 3 anatomic classes.

DISCUSSION

This observer study showed an improvement in detection of myocardial perfusion defects in SPECT images reconstructed using MLAC compared with those reconstructed using FBP without AC. A separate comparison was performed for each of 3 anatomic classes: male patients with a flat diaphragm showing no significant diaphragmatic attenuation effect, male patients with a raised diaphragm showing the diaphragmatic attenuation effect, and female patients with large breasts and a flat diaphragm showing the breast attenuation effect. The MLAC reconstruction technique greatly improved defect detection for phantoms modeled with large breasts or with a raised diaphragm but only marginally improved defect detection for phantoms in the third class. These results suggest that the benefit of MLAC reconstruction is patient dependent and that some patient populations may benefit more than others from the MLAC reconstruction technique. The 2 reconstruction techniques were not compared for the population of phantoms as a whole because such a comparison requires a somewhat arbitrary assumption of what constitutes a typical patient population and cannot be generalized beyond the population studied. Instead, the potential benefit of the MLAC technique must be assessed given the results of this study and a knowledge of the prevalence of prominent breast or diaphragmatic attenuation artifacts in each reader's own patient population.

The application of AC in the reconstruction of SPECT images incurs costs from acquisition and maintenance of additional equipment, increased imaging time (for some imaging protocols), extra processing and archiving of images, modifications to existing imaging protocols, and additional training for technologists and radiologists. Thus, the degree of benefit one can expect from the MLAC reconstruction technique is important, because a marginal improvement in defect detection with MLAC may not offset these costs. This study can lead to no general recommendation on whether a clinic should use the MLAC technique, but the results suggest a potentially significant improvement in myocardial defect detection for clinics in which breast or diaphragmatic attenuation artifacts are prevalent.

A limitation of all observer studies is the difficulty in generalizing the results beyond the context of the study. The image display, image processing, data acquisition, and reconstruction methods chosen for this study are common,

but these methods vary considerably within the nuclear medicine community, and our results may not be applicable to all clinical settings. As in the studies of Ficaro et al. (11) and Gilland et al. (10), the results must be interpreted with respect to the task, context, and limitations of this study.

The acquired counts in the simulated data were much lower (and therefore the noise level was much higher) than those typically seen in ^{99m}Tc -sestamibi clinical data. The high level of noise in the simulated images in the final study did not fundamentally alter the task of the observers or undermine the relevance of the results. This lower count level was necessary in the final study, because in preliminary studies with images simulated at the clinical count levels (20), the task of the observers was trivial ($A_z = 1.0$) and thus not amenable to a meaningful statistical analysis using current techniques. The preliminary studies used the same types of images and methodology as the final study but were performed at several different acquired count levels. As the projection data were scaled from clinical count levels down to one tenth the clinical count level, the absolute values of A_z fell approximately 5% or less; however, the relative differences in A_z between the FBP and MLAC curves and the rank of the A_z values—the results of interest for this study—were essentially unchanged. The increased noise level simply scaled the A_z values to a more useful range of values for statistical analysis. For the FBP images with severe diaphragmatic and breast attenuation artifacts, detection was essentially unaffected by noise level (for the noise levels tested), indicating that the primary causes of degraded detection in those images were the attenuation artifacts.

In this study, the ^{99m}Tc -sestamibi MLAC images showed, on average, roughly a 10% artifactual increase in count density in the inferoseptal left ventricular wall because of the high liver uptake in combination with scatter and, possibly, collimator-detector response effects. However, this inferoseptal scatter artifact did not measurably degrade defect detection. Findings indicate that an inferoseptal scatter artifact can degrade defect detection in AC images; thus, this study may have underestimated the effect of scatter artifacts in AC images (25,26). The scatter in our simulation is modeled realistically, but some patient factors not modeled affect the inferoseptal artifact: a left ventricular wall in close proximity to the liver, high uptake in other subdiaphragmatic organs, and movement of the liver and heart during breathing. Further study of the efficacy of scatter correction in myocardial SPECT is needed.

The major physical factors relevant to this study, that is, the anatomy of the thoracic structures, the static radionuclide uptake model, and the basic physical effects of attenuation, scatter, collimator-detector response, and noise, are modeled realistically. Other patient-related factors, such as radionuclide uptake kinetics, variable myocardial wall thickness, and both voluntary and involuntary patient motion, were not modeled. The influence that these patient-related factors have on detection of perfusion defects in MLAC SPECT images is not clear, but patients with

prominent artifacts because of these factors may not benefit as much from AC. The inclusion of these factors would have undermined an important benefit of simulation: elimination of confounding effects that plague clinical images. The influence of these factors is, however, an important subject for future investigations, and the paradigm used in this study may be useful for those investigations.

Our results are clearly different from those of Gilland et al. (10). Defect size is likely a major factor. Detection of small defects, as in the study of Gilland et al. (10), appears to be more sensitive to the noise level in the images than to attenuation artifacts, whereas detection of large defects, as in our study, appears to be more sensitive to the severity of attenuation artifacts. Our choice of defect size, image display format, and phantom population related the task of our observers more closely to the detection of myocardial perfusion defects in a clinical setting. Thus, we believe our results more accurately predict the impact of the MLAC technique on detection of myocardial perfusion defects in clinical SPECT images. The results of Gilland et al. (10) may be more relevant to other clinical tasks. Taken together, both studies show that statements about the superiority of 1 reconstruction technique over another are meaningful only with regard to a specific task and imaging context.

Our results are consistent with those of Ficaro et al. (11) with respect to the ranking and the relative difference in detection between the FBP and the iteratively reconstructed AC images. For FBP images without AC, the primary cause of degraded detection in both this study and that of Ficaro et al. were attenuation artifacts caused by breast and diaphragmatic attenuation. Together, the results of both studies strongly suggest a significant potential benefit to using iterative reconstruction with AC for myocardial SPECT images.

Since the completion of this observer study, Jang et al. (38) published the results of an observer study and ROC analysis that evaluated, for females with large breasts, differences in myocardial defect detection between images reconstructed using FBP without AC and images reconstructed using a single iteration of the Chang algorithm with nonuniform AC. In the study of Jang et al., defect detection for images reconstructed with AC was not significantly different from—and in some cases was slightly worse than—defect detection for images reconstructed without AC. Several factors could have contributed to the different results between our study and that of Jang et al., but a major factor was probably the difference in the algorithm used to reconstruct the AC images.

In this study, the degree of benefit of MLAC over FBP without AC was found to depend on the anatomic characteristics of the patient. This finding has important implications for future observer studies using either simulated or clinical images. For simulation studies, these results underscore the need to apply new techniques to a variety of phantom anatomies, because different anatomies may yield different results. For clinical studies, these results underscore the need

to report the prevalence and severity of suspected artifacts in the sample population so that the results can be interpreted in the proper context.

CONCLUSION

An observer study and ROC analysis were performed to evaluate differences in myocardial defect detection between simulated ^{99m}Tc -sestamibi SPECT images reconstructed using FBP without AC and those reconstructed using MLAC. The results indicate that defect detection for MLAC is greater than or equal to that for FBP without AC; however, the difference in detection between MLAC and FBP varied with anatomy. The attenuation artifacts that degraded myocardial defect detection in the FBP images were primarily caused by breast or diaphragmatic attenuation. The phantoms with large breasts or a raised diaphragm showed a large increase in defect detection with MLAC, whereas the remaining class of phantoms showed a relatively small increase. The results suggest that the overall benefit of the MLAC technique for a given patient population depends on the prevalence of breast and diaphragmatic artifacts in that population. In addition, the results underscore the need to consider a variety of patient anatomies in simulation studies and the importance of the sample population in clinical studies.

ACKNOWLEDGMENTS

The authors thank Andrew Barclay (Crawford Long Hospital of Emory University) and Dr. Tim Turkington (Duke University Medical Center) for supplying PET patient images; Dr. William McCartney, Dr. J. Randolph Perry, Dr. David Lalush, Dr. Scott Wollenweber, Dr. Dan Kadrmas, Daniel Wessell, and Seemeen Karimi for being our observers; and Drs. Carol Lucas, Timothy Johnson, and Charles Metz for providing helpful and insightful comments. This work was supported by Public Health Grant CA39463. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Public Health Service.

REFERENCES

1. Wackers FJ. Myocardial perfusion imaging. In: Sandler MP, Coleman RE, eds. *Diagnostic Nuclear Medicine*. Vol 1, 3rd ed. Baltimore, MD: Williams and Wilkins; 1996:443–516.
2. Bellar GA. Myocardial perfusion imaging with thallium-201. *J Nucl Med*. 1994;35:674–680.
3. Zaret BL, Wackers FJ. Nuclear cardiology. Part 1. *New Engl J Med*. 1993;329:775–783.
4. Malko JA, Van Heertum RL, Gullberg GT, Kowalsky WP. SPECT liver imaging using an iterative attenuation correction algorithm and an external flood source. *J Nucl Med*. 1986;27:701–705.
5. Tsui BMW, Gullberg GT, Edgerton ER, et al. Correction of nonuniform attenuation in cardiac SPECT imaging. *J Nucl Med*. 1989;30:497–507.
6. Tsui BMW, Zhao XD, Frey EC, McCartney WH. Quantitative single-photon emission computed tomography: basics and clinical considerations. *Semin Nucl Med*. 1994;24:38–65.
7. King MA, Tsui BMW, Pan T-S. Attenuation compensation for cardiac single-photon emission computed tomographic imaging: part 1. Impact of attenuation and methods of estimating attenuation maps. *J Nucl Cardiol*. 1995;2:513–524.

8. Shepp LA, Vardi Y. Maximum likelihood estimation for emission tomography. *IEEE Trans Med Imaging*. 1982;MI-1:113-121.
9. Lange K, Carson R. EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr*. 1984;8:306-316.
10. Gilland DR, Tsui BMW, Metz CE, Jaszczak RJ, Perry JR. An evaluation of maximum likelihood-expectation maximization reconstruction for SPECT by ROC analysis. *J Nucl Med*. 1992;33:451-457.
11. Ficaro EP, Fessler JA, Shreve PD, Kritzman JN, Rose PA, Corbett JR. Diagnostic accuracy of attenuation-corrected ^{99m}Tc-sestamibi single-photon emission computed tomography. *Circulation*. 1996;93:463-473.
12. Fessler JA. Penalized weighted least squares image reconstruction for positron emission tomography. *IEEE Trans Med Imaging*. 1994;13:290-300.
13. DePuey EG, Garcia EV. Optimal specificity of thallium-201 SPECT through recognition of imaging artifacts. *J Nucl Med*. 1989;30:441-449.
14. Garcia EV, Van Train K, Maddahi J, et al. Quantification of rotational thallium-201 myocardial tomography. *J Nucl Med*. 1985;26:17-26.
15. Tamaki N, Yonekura Y, Kadota K, Kambara H, Torizuka K. Appearance of breast attenuation artifacts with thallium myocardial SPECT imaging. *Clin Nucl Med*. 1985;10:694-696.
16. Esquerré J, Coca FJ, Martinez SJ, Guiraud RF. Prone decubitus: a solution to inferior wall attenuation in thallium-201 myocardial tomography. *J Nucl Med*. 1989;30:398-402.
17. Segall GM, Davis MJ. Prone versus supine thallium myocardial SPECT: a method to decrease artifactual inferior wall defects. *J Nucl Med*. 1989;30:548-555.
18. Tsui BMW, Terry JA, Gullberg GT. Evaluation of cardiac cone-beam SPECT using observer performance experiments and ROC analysis. *Invest Radiol*. 1993;28:1101-1112.
19. Pretorius PH, Xia W, King MA, Tsui BMW, Pan TS, Villegas BJ. Evaluation of right and left ventricular volume and ejection fraction using a mathematical cardiac torso phantom for gated blood pool SPECT. *J Nucl Med*. 1997;38:1528-1534.
20. LaCroix KJ. *Evaluation of an Attenuation Compensation Method with Respect to Defect Detection in Tc-99m-MIBI Myocardial SPECT Images* [dissertation]. Chapel Hill, NC: The University of North Carolina at Chapel Hill; 1997.
21. Barclay AB, Eisner RL, DiBella EV. PET thorax model database. Available at: <http://www.emory.edu/CRL/abb/thoraxmodel>. Accessed Jan. 4, 2000.
22. Gilland DR, Jaszczak RJ, Hanson MW, Greer KL, Coleman RE. An experimental phantom based on quantitative SPECT analysis of patient MIBI biodistribution [abstract]. *J Nucl Med*. 1996;37(suppl):154P.
23. McCartney WH, Tsui BMW, Adams KF, et al. Clinical evaluation of attenuation and scatter compensation in TI-201 SPECT [abstract]. *J Nucl Med*. 1996;37(suppl):80P.
24. DePuey EG, Shahzad K. Does scatter correction significantly improve attenuation compensation for Tc-99m myocardial perfusion SPECT? [abstract]. *J Nucl Med*. 1998;39(suppl):74P.
25. Svane B, Bone D, Holmgren A, Landou C. Polar presentation of coronary angiography and thallium-201 SPECT: a method for comparing anatomic and pathologic findings in coronary angiography with isotope distribution in thallium-201 myocardial SPECT. *Acta Radiol*. 1989;30:561-574.
26. Frey EC, Ju Z-W, Tsui BMW. A fast projector-backprojector pair modeling the asymmetric, spatially varying scatter response function for scatter compensation in SPECT imaging. *IEEE Trans Nucl Sci*. 1993;NS-40:1192-1197.
27. Frey EC, Tsui BMW. Modeling the scatter response function in inhomogeneous scattering media for SPECT. *IEEE Trans Nucl Sci*. 1994;41:1585-1593.
28. Harrison R, Vannoy S, Haynor D. Preliminary experience with the photon history generator module of a public domain simulation system for emission tomography. In: Klaisner L, ed. *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference, San Francisco, CA, November 1993*. Piscataway, NJ: IEEE; 1993:1154-1158.
29. Ifitkar I, Koutelou M, Mahmarian JJ, Verani MS. Simultaneous perfusion tomography and radionuclide angiography during dobutamine stress. *J Nucl Med*. 1996;37:1306-1310.
30. Kiat H, Maddahi J, Roy LT, et al. Comparison of technetium 99m methyl isobutyl isonitrile and thallium 201 for evaluation of coronary artery disease by planar and tomographic methods. *Am Heart J*. 1989;117:1-11.
31. Taillefer R, DePuey EG, Udelson JE, Bellar GA, Latour Y, Reeves F. Comparative diagnostic accuracy of TI-201 and Tc99m sestamibi SPECT imaging (perfusion and ECG-gated SPECT) in detecting coronary artery disease in women. *J Am Coll Cardiol*. 1997;29:69-77.
32. Tung CH, Gullberg GT. A simulation of emission and transmission noise propagation in cardiac SPECT imaging with nonuniform attenuation correction. *Med Phys*. 1994;21:1565-1576.
33. Cao Z, Tsui BMW. Performance characteristics of transmission imaging using a uniform sheet source with parallel collimation. *Med Phys*. 1992;19:1205-1212.
34. Meikle S, Dahlbom M, Cherry S. Attenuation correction using count-limited transmission data in PET. *J Nucl Med*. 1993;34:143-150.
35. Tsui BMW, Zhao XD, Lalush DS, Frey EC, Johnston RE, Bujenovic LS. Pitfalls of attenuation compensation and their remedies in cardiac SPECT [abstract]. *J Nucl Med*. 1994;35(suppl):115P.
36. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol*. 1989;24:234-245.
37. Metz CE. Statistical analysis of ROC data in evaluating diagnostic performance. In: Herbert D, Myers R, eds. *Multiple Regression Analysis: Applications in the Health Sciences*. New York, NY: American Institute of Physics; 1986:365-384.
38. Jang S, Jaszczak RJ, Tsui BMW, et al. ROC evaluation of SPECT myocardial lesion detectability with and without single iteration non-uniform Chang attenuation compensation using an anthropomorphic female phantom. *IEEE Trans Nucl Sci*. 1998;45:2080-2088.