ROC and LROC Analyses of the Effects of Lesion Contrast, Size, and Signal-to-Noise Ratio on Detectability in PET Images

Thomas H. Farquhar, Jorge Llacer, James Sayre, Yuan-Chuan Tai, and Edward J. Hoffman

Department of Molecular and Medical Pharmacology, Division of Nuclear Medicine and Biophysics, and Department of Radiological Sciences, School of Medicine, University of California, Los Angeles; EC Engineering Consultants, Los Gatos, California

Image quality in PET is typically assessed using measures such as contrast recovery, noise variation, and signal-to-noise ratio (SNR). However, these criteria do not directly reflect performance in the clinical use of the images. Lesion detection is a critical task in the clinical interpretation of many PET studies. A receiver operating characteristic (ROC) study is an accepted method for quantitatively evaluating detection performance with respect to factors that influence image quality. ROC and localization ROC (LROC) analyses were conducted to investigate the effects of lesion contrast, SNR, and size on detectability of hot lesions in PET images. **Methods:** A thorax phantom was imaged with spheres of 3 sizes simulating lesions (0.45, 1.0, and 1.9 mL). The relative activity in the lesions and the total number of counts acquired were each varied by factors of 2 to ascertain the effects of contrast and SNR, respectively. Measured attenuation correction and a standard reconstruction protocol were used. Three nuclear medicine physicians and 6 medical physicists participated as readers, rating each image and indicating the suspected lesion location. The area under the calculated ROC and LROC curves (A_z and $A_{z,LROC}$) were used as measures of detection performance. **Results:** Detection performance was shown to increase from virtually random ($A_z \sim 0.5$, $A_{z,LROC} \sim 0.2$) to superior ($A_z > 0.9$, $A_{z,LROC} > 0.9$) as lesion contrast was increased by 50% and as lesion SNR was doubled. Detection performance was not seen to vary when comparison was made using image-based measures alone. **Conclusion:** This study quantitatively shows that moderate increases in the image-based measures of lesion contrast and SNR give a relatively large increase in the task-based measure of lesion detection as measured by ROC and LROC analyses. Thus, techniques that give modest increases in lesion contrast or SNR are expected to improve detection. Results will be useful in evaluating improvement in detection for various reconstruction, acquisition, and data analysis methods that enhance contrast or noise performance.

Key Words: receiver operating characteristic analysis; lesion detection; contrast; signal-to-noise ratio

The image quality of reconstructed PET data can be assessed using numerous objective measures. For example, resolution, contrast recovery, noise variation, bias, and signal-to-noise ratio (SNR) are widely cited measures of image quality (*1*). One or more of these measures are typically chosen to compare the relative performances of reconstruction algorithms. However, comparisons based on these quantitative, image-based criteria do not directly reflect the differences in performance in the ultimate use of the images, for example, clinical interpretation or research analysis. In the case of whole-body FDG PET images, the identification and localization of foci of increased FDG uptake (generically called a lesion) by an observer is the primary task in clinical interpretation.

A receiver operating characteristic (ROC) study is an accepted method for quantitatively evaluating the detection performance of an observer with respect to factors that influence image quality (*2,3*). ROC analysis has been applied to CT images to show that lesion contrast, size, and noise are closely related to measures of an observer's detection and localization ability (*4,5*). Similar work on lesion detection has been reported for SPECT imaging, although without formal ROC analysis (*6*). In this article, an ROC/localization ROC (LROC) study is described and identifies the effects of lesion contrast, SNR, and size on the detection of a small focus of elevated radioisotope activity. The study shows quantitatively the association between an increase in lesion contrast and lesion SNR and a corresponding increase in performance for the task of lesion detection, specifically for PET images.

A phantom of the human thorax with a heart and 2 lung inserts was used along with small, spherical "tumor" spheres of different sizes. The phantom approximated a hot lesion in an otherwise homogeneous background. (Cold lesions, which are foci of depressed radioisotope concentration and also of importance in clinical PET, were not considered in this study.) For specific lesion sizes, the range of contrast over which lesion detection improved from virtually undetectable to nearly perfect detection was determined. The narrow range over which detection improved suggested an optimal range for evaluation of new image

processing methods. By evaluating a proposed data correction technique or reconstruction algorithm with imaging conditions similar to those in this study, any measured improvement in contrast recovery can be translated into an expected improvement in lesion detection. In addition, the effect of SNR on detection is shown for a single lesion size. As with the studies of contrast, the degree to which a decrease in background noise and concomitant increase in lesion SNR result in improved lesion detection can be estimated from the results described here.

MATERIALS AND METHODS

Data Acquisition

The thorax phantom (Figure 1) consisted of an elliptical torso chamber containing 2 lung-equivalent foam inserts and a 3-chamber cylindrical heart insert. Fillable plastic spheres with volumes of 0.45 mL (radius = 4.7 mm), 1.0 mL (radius = 6.2 mm), and 1.9 mL (radius = 7.7 mm) were used as lesions. Data were acquired for the backgrounds and lesions separately. Using separate acquisitions, the sinograms from a background-only study and a lesion-only study could be added before reconstruction, to give a true-positive image. This allowed the activity concentration or count levels in the background and lesions to be determined precisely. This method also simplified acquisition of the many desired datasets.

For acquisitions of the thorax phantom only, the torso cavity and the 2 outer chambers of the heart insert were filled with a solution of FDG. The total activity used in the phantom was measured in a well counter and decay corrected to the start of the acquisition. Multiple frames of 7, 14, and 28 × 10⁶ total counts (over 63 planes) were collected in the 2-dimensional mode on an ECAT EXACT HR+ 962 scanner (CTI, Inc., Knoxville, TN). A 2-h transmission scan was obtained the day after the emission scans.

In subsequent acquisitions, each lesion sphere was filled with ^{13}N NH_3 solution and imaged dynamically at multiple positions. The background and heart of the phantom were filled with water but no radioisotope. Five dynamic frames covered 1 half-life of decay of ^{13}N, with relative activities of 1.0, 0.85, 0.71, 0.60, and 0.50. For each lesion acquisition, the entire phantom was repositioned to align with the previous study of the background and heart insert.
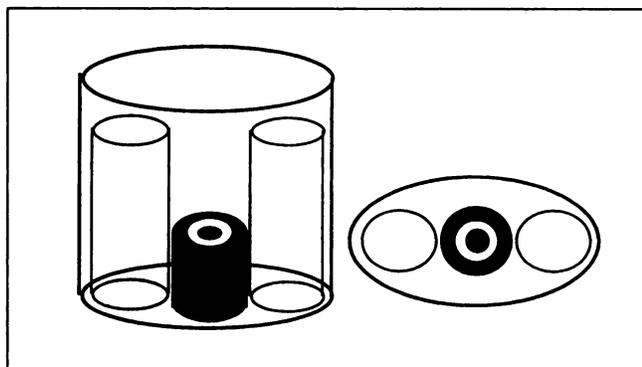
FIGURE 1. Schematic of thorax phantom used in ROC phantom study. Phantom consists of elliptical, background torso chamber enclosing 2 inserts with lung-equivalent foam, and 3-chamber fillable heart insert.

TABLE 1
Prepared Image Data Sets According to Lesion Size and Total Number of Counts in Background

Lesion size (mL)	Background counts	Lesions (n)	Total images (n)
0.45	14 × 10⁶	7	70
1.0	14 × 10⁶	7	70
1.9	14 × 10⁶	9	90
1.0	7, 14, and 28 × 10⁶	5	150

In each case, 5 contrast levels were used. In each set, half the images were true-negatives with no lesion present.

Image Preparation

Four sets of images were prepared (Table 1). Because statistical power in ROC studies was maximized when approximately half of the images were normal (7), half of the images in each group were true-negatives with no lesion present. Lesions of 5 contrast levels were used to evaluate the effect of contrast. For a single bed of a clinical 2-dimensional whole-body FDG PET scan in the region of the thorax, 14 × 10⁶ counts were observed to be typical Therefore, a background scan with a total of 14 × 10⁶ counts was used, with added lesions of different sizes and contrasts, to evaluate the effect of lesion contrast on detectability for the 3 lesion sizes.

The first 3 datasets in Table 1 vary contrast within a specific lesion size. Sinogram data for true-negative images consisted of 5 frames of 14 × 10⁶-count background collected without lesion data. Formation of true-positive sinogram data required the addition of lesion data. The same background scan was added to 5 dynamic frames from a particular lesion acquisition without background activity and then reconstructed.

The prepared true-positive and true-negative sinograms were reconstructed using measured attenuation correction and the standard clinical whole-body reconstruction filter parameters (Hann filter, cutoff frequency 0.42 of Nyquist in-plane [0.98 cycles/cm], 0.40 of Nyquist axially [0.86 cycles/cm]). Three consecutive transaxial image planes were displayed for each case to be rated, allowing the reader to evaluate the central plane in the context of adjacent planes. These sets of 3 consecutive images planes were referred to as image triplets. For true-positive image triplets, the center plane corresponded to the center of the lesion sphere.

The fourth dataset described in Table 1 used lesions of a single lesion size (1.0 mL), but the background count level was varied by factors of 0.5, 1.0, and 2.0. With differing count levels and therefore different noise levels, the effect of SNR apart from the effect of contrast could be evaluated. The images, with total count levels of 7, 14, and 28 × 10⁶ counts, were prepared in a method similar to that described previously, with the following modifications. For background studies with total counts of 7, 14, and 28 × 10⁶, the relative number of counts were 0.5, 1.0, and 2.0, respectively. The 5 contrast levels were obtained as described previously, with the dynamic frame offset by 1 half-life as necessary to generate relative numbers of counts of 0.5, 1.0, and 2.0. Reconstruction was identical to that in the previous datasets. Selection of image planes proceeded as previously described; the same planes were used for each true-positive or true-negative image triplet at each count level.

Image Presentation and Evaluation

Image evaluation was performed using an X Windows-based (Sun Microsystems, Palo Alto, CA) ROC program modified from code originally developed by Dr. Benjamin Tsui and colleagues at the University of North Carolina. A training session preceded the study. The nature of the study was explained and detailed instructions on use of the ROC evaluation software were provided. A representative sample of 2 true-negative and 2 true-positive triplets from each lesion size, lesion contrast level, and total count level were used. Using a special training mode of the software, the reader was able to reveal the true nature of the displayed image triplet as a true-negative image or a true-positive image along with the lesion location shown for a true-positive image.

Three nuclear medicine physicians and 6 medical physicists participated as readers. The order in which each of the 4 image sets was evaluated was randomized to minimize reader-order effects. Image evaluation was performed for each image triplet in an image set. Each triplet was rated, using a continuous rating scale from 1 to 5 to answer the question "Does the image show an abnormality?" The ratings were defined as: 1 = definitely no; 2 = probably no; 3 = possibly yes; 4 = probably yes; and 5 = definitely yes. In addition, readers indicated the most probable lesion locations. Readers were also able to adjust the color scale of each image to facilitate image viewing.

Measurement of Lesion Contrast

Lesion contrast was determined by 2 methods, representing measurement of the lesion contrast in the phantom and in the reconstructed image. Lesion contrast was calculated as lesion activity minus background activity divided by background activity. To measure the lesion contrast in the phantom, the integrated, decay-corrected activity concentration in the lesion and background were calculated from readings with a well counter. In the reconstructed images, the lesion activity was measured as the maximum pixel intensity in a region-of-interest (ROI) located at the lesion. Using images of the same background scans without the lesion data added, the average pixel intensity in an ROI surrounding the lesion location was used as the background activity.

Measurement of Lesion SNR

The noise levels in the images were also measured by 2 methods. First, the total number of true events in all planes of each acquisition frame was taken as the measure of relative noise in the original sinogram data. In both cases, a measure of lesion contrast represented the signal component. The signal strength was then divided by a measure of the noise to obtain the SNR. The first measure used the lesion contrast determined with well counter readings and the noise level estimated from the total number of acquired events. The second measure used the lesion contrast and noise variance determined from ROIs in the reconstructed images. These 2 measures reflected the SNR before and after the effect of image reconstruction.

ROC Analysis of Individual Readers

ROC analysis was performed using a modification of the CORROC2 program developed at the University of Chicago by Metz et al. (8). The CORROC2 algorithm computed binormal curve fits for the ratings of 2 sets of images in a matched-pair design. For each lesion size and noise level, a total of 4 pair-wise comparisons were made, comparing the ratings data for contrast levels 2–5 (higher contrast) with the ratings for contrast level 1 (lowest contrast). The ratings were compared as 2 random decision

variables under the assumption that each variable had a bivariate normal distribution. The CORROC2 analysis accounted for the underlying correlation of the matched-pair design (i.e., lesions differed only in contrast and had identical locations and background activity distributions) during computation of the curve fit (7). Using the resulting binormal curve fit, the area under the ROC curve (A_z) was also computed, as well as an SE estimate for the reported A_z. Four valves of A_z for contrast level 1 were obtained, 1 from each comparison between contrast level 1 and contrast levels 2–5. These multiple values for contrast level 1 were not seen to vary significantly. The values of A_z reported for contrast level 1 reflect the average of the 4 calculated values of A_z. Finally, a degenerate result was reported by the CORROC2 fitting algorithm in certain instances when the measured operating points could not be properly fitted by a maximum-likelihood method (3,7). In this study, the usual source of degeneracy was near-perfect detection, i.e., all of the true-positive images were rated higher than all of the true-negative images.

Jack-Knife Analysis of Pooled Reader Data

Because of the low number of sample images rated by each observer and the small differences in contrast between adjacent contrast levels, the SE of each measured A_z was expected to be relatively large. To establish a level of statistical significance for the observed improvement in detection, the multireader ROC data for each lesion volume were analyzed with a jack-knife technique (9,10). Using the jack-knife technique, ROC data can be pooled and both between-case and between-reader variations can be considered. The result was a reduction in the bias resulting from between-case and between-reader correlations and a reduction in SE of the estimated A_z (10).

The jack-knife analysis also provided a P to assess the statistical significance of a difference in detection observed between 2 sets of images. The null hypothesis assumed that the increase in contrast or SNR did not result in a change in detection performance. Comparisons in which $P < 0.05$ suggested that the null hypothesis could be rejected and that a difference in detection performance existed.

The jack-knife ROC procedure was even more sensitive to degenerate data than the CORROC2 algorithm, and, as such, the degenerate results from any reader were not incorporated into the pooled data for use with the jack-knife analysis. Because the small sample of rated images for each contrast and noise level (5 true-positive and 5 true-negative images) frequently gave degenerate results, the ratings data from the SNR image set were not amenable to analysis using the jack-knife method.

LROC Analysis of Individual Readers

The ratings and suspected lesion location recorded by each observer were also analyzed using an LROC analysis program developed by Swensson (2). In this analysis, the responses of each reader for all 5 contrast levels (and in the analysis of SNR, all 3 noise levels) were analyzed simultaneously, yielding a fitted LROC curve, the area under the fitted LROC curve ($A_{z,LROC}$), and the SE of the estimate for each contrast level (and noise level). A lesion was recognized as being properly localized when the pixel specified by the reader was within a 4-pixel radius (11.25 mm) of the center of the true lesion location. The 4-pixel radius accounted for the finite size of the lesions (larger than a single pixel) and mild inaccuracy in the positioning of the computer cursor on the center of the lesion. Although location information was included in the LROC analysis, correlations between cases and between readers were not accounted for as in the CORROC2 and jack-knife analyses of the

TABLE 2
Average Measured Contrast for Lesion Contrast Levels

| | 0.45 mL (r = 4.7 mm) | | 1.0 mL (r = 6.2 mm) | | 1.9 mL (r = 7.7 mm) | |
Level	Well counter mean (SD)	Image pixels mean (SD)	Well counter mean (SD)	Image pixels mean (SD)	Well counter mean (SD)	Image pixels mean (SD)
5	1.53 (0.17)	0.29 (0.08)	0.79 (0.28)	0.33 (0.08)	0.50 (0.11)	0.35 (0.06)
4	1.31 (0.15)	0.25 (0.07)	0.68 (0.24)	0.28 (0.05)	0.43 (0.10)	0.30 (0.05)
3	1.09 (0.12)	0.23 (0.07)	0.56 (0.20)	0.24 (0.04)	0.35 (0.08)	0.26 (0.05)
2	0.93 (0.10)	0.20 (0.07)	0.48 (0.17)	0.20 (0.03)	0.30 (0.07)	0.22 (0.04)
1	0.77 (0.09)	0.19 (0.07)	0.40 (0.14)	0.18 (0.03)	0.25 (0.06)	0.19 (0.03)

For each lesion size, lesion contrast at each level was determined by measuring activity concentration with a well counter and pixel intensity in an ROI in reconstructed image.

ROC data. To ascertain the statistical significance of difference between modalities observed across readers, a Student t test for paired data was performed. Unlike the jack-knife analysis, the t test assumed that the LROC of each observer was an independent sample from the same population. This assumption was not strictly satisfied, because each reader evaluated the same population of images. However, an LROC analysis program from correlated data was not available to the authors. The t statistic was converted to P to facilitate comparison with the ROC results.

RESULTS

Lesion Contrasts

The average lesion contrast for each lesion size (and SDs), as determined by the 2 methods described above, are shown in Table 2 and Figure 2. The reduced contrast observed in the image domain compared with the true activity concentration in the phantom can be attributed to physical effects in acquisition and reconstruction, including partial-volume effects (11,12) and blurring as a result of reconstruction with the smooth, whole-body protocol reconstruction filter.
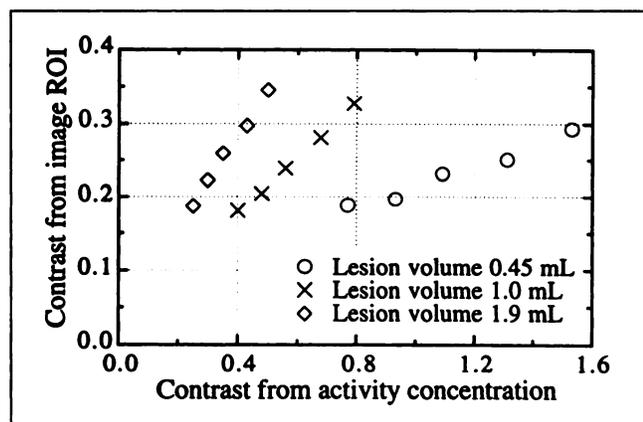
FIGURE 2. Average measured contrast for lesion contrast levels, by both methods of measurement. Contrast determined by activity concentration as measured with well counter is plotted versus contrast determined using ROI in reconstructed image.

Lesion SNRs

The results of the 2 measures of background noise levels are shown in Table 3. Because the sinogram data were essentially Poisson distributed, each 2-fold increase in the number of total number of counts should decrease the noise by a factor of $\sqrt{2}$, or 41%. Measured values from ROIs in the images showed a 20% decrease in noise variance for the 28×10^6-count data and a 20% increase in noise variance for the 7×10^6-count data, when compared with the 14×10^6-count studies. These data were combined with the lesion contrast results from Table 2 to create 15 different lesion SNRs for each method described previously.

ROC Analysis of Effect of Lesion Contrast and Lesion Size

The A_z of the fitted curves computed with the CORROC2 program for each reader by contrast level and lesion size are shown in Figure 3A. Readers 2–9 completed evaluation of the 0.45-mL lesion–sphere images. Readers 1–9 completed evaluation of the 1.0-mL lesion–sphere images. Readers 3–9 evaluated the images for the largest lesion–spheres with volumes of 1.9 mL. These results are summarized in Figure 4, with the average A_z of all readers plotted versus the contrast, as determined by both methods of measurement.
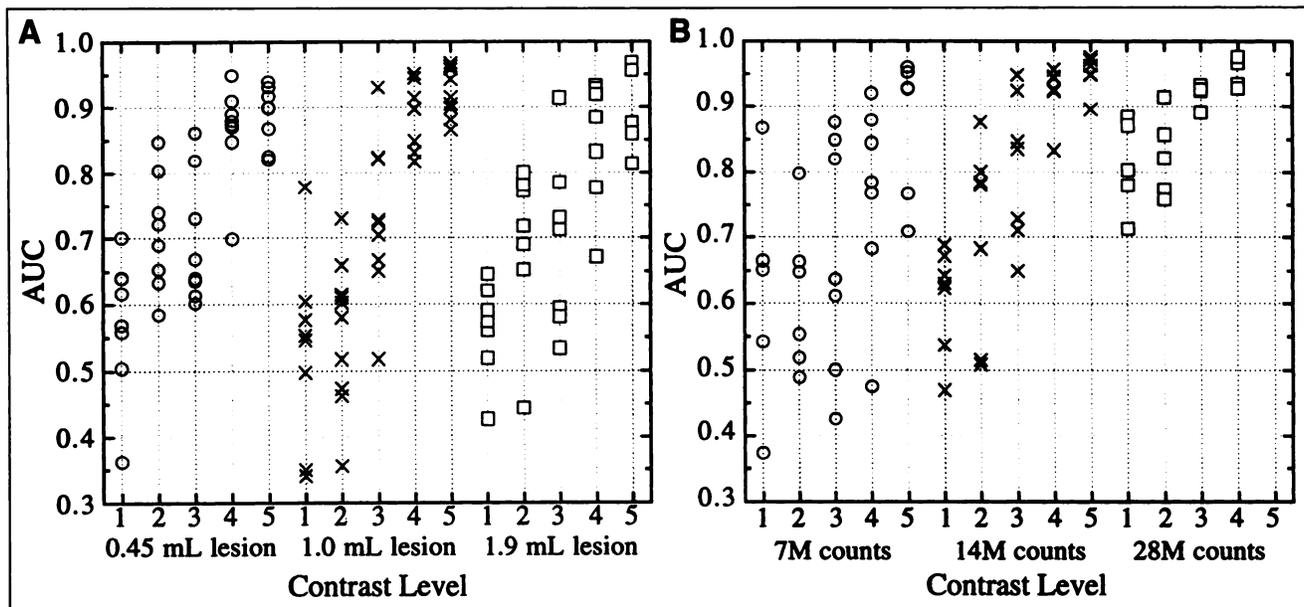
TABLE 3
Measurements of Noise Levels

Total no. of counts*	Noise by total counts† (%)	Noise by %SD of ROI‡ (%)
7×10^6	141	9.82
14×10^6	100	8.28
28×10^6	71	6.67

*Total number of counts for all 63 2-dimensional planes in single acquisition frame.
†Noise level calculation based on acquired counts is relative to 14×10^6-count dataset.
‡%SD of ROIs is averaged for ROIs of all 5 lesions in study.

FIGURE 3. (A) Area under ROC curve (AUC) for each reader by lesion size and contrast level. SE = 0.05–0.15. (B) AUC for each reader by total number of counts in acquisition and contrast level. SE = 0.04–0.18. Degenerate results are not included in either plot.

Figure 4 shows the improvement in detectability as lesion contrast was increased for each lesion volume in the study. At the lowest contrast, the average A_z was only marginally better than random performance (0.57, 0.54, and 0.56). The detection performance, as measured by the average A_z, improved steadily with increasing contrast towards virtually perfect detection (0.89, 0.92, and 0.89) for all of the lesion sizes.

The results of the pooled data analysis using the jack-knife methodology with each lesion volume are shown in Table 4. In each case, data at a given contrast level were

compared against 2 adjacent contrast levels to determine if a statistically significant difference existed. Using a Bonferroni correction for multiple comparisons, $P < 0.025$ was considered to be a significant difference with 95% confidence. Using this threshold, data for the 0.45-mL lesion supported the claim that a statistically significant difference in contrast was observed between contrast levels 2 and 4, 3 and 5, and, most likely, between 1 and 2. A reasonable interpretation of the high P value (0.616) of the contrast level 4-to-contrast level 5 comparison was that the lesion was nearly perfectly seen and that further
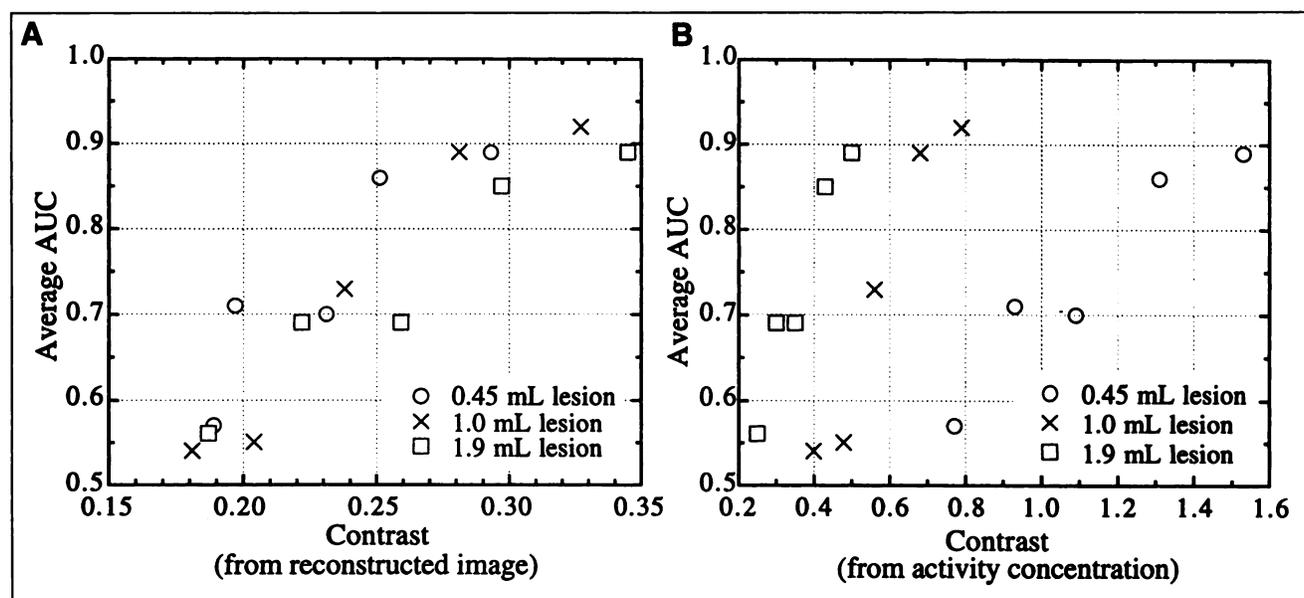
FIGURE 4. Average area under ROC curve (AUC) versus contrast. (A) Image contrast is measured from pixels in reconstructed image. (B) Image contrast is measured using activity concentration.

TABLE 4
Jack-Knife Analysis of Reader-Pooled Data by Contrast Level

Lesion size (mL)	Contrast level*					P†
	5	4	3	2	1	
0.45						
2 vs. 1				0.62 (0.06)	0.48 (0.06)	0.028
3 vs. 1			0.61 (0.06)		0.52 (0.06)	0.219
4 vs. 2		0.74 (0.04)		0.63 (0.06)		0.021
5 vs. 3	0.79 (0.04)		0.62 (0.06)			0.004
5 vs. 4	0.77 (0.04)	0.75 (0.04)				0.616
1.0						
2 vs. 1				0.49 (0.06)	0.50 (0.06)	0.950
3 vs. 1			0.65 (0.05)		0.49 (0.06)	0.008
4 vs. 2		0.77 (0.04)		0.51 (0.06)		0.000
5 vs. 3	0.79 (0.04)		0.63 (0.05)			0.000
5 vs. 4	0.80 (0.04)	0.77 (0.04)				0.260
1.9						
2 vs. 1				0.61 (0.06)	0.49 (0.06)	0.020
3 vs. 1			0.64 (0.06)		0.49 (0.06)	0.022
4 vs. 2		0.77 (0.04)		0.60 (0.06)		0.004
5 vs. 3	0.77 (0.05)		0.56 (0.07)			0.000
5 vs. 4	0.79 (0.04)	0.74 (0.05)				0.156

*Values are A_z with values in parentheses given as SD.

†P reported for each comparison can be used to establish level of statistical significance for observed trend of increasing detection over range of contrasts studied.

increases in contrast only marginally improved detection as measured by A_z.

For the 1.0-mL lesion volume, a statistically significant difference in detection was observed between contrast levels 1 and 3, 2 and 4, and 3 and 5. Again, the higher P values of the contrast level 1 versus contrast level 2 comparison and the contrast level 4 versus contrast level 5 comparison would be consistent with those pairs of levels each having similarly poor (levels 1 and 2) or similarly accurate (levels 4 and 5) detection.

Finally, in the pooled 1.9-mL lesion data, statistically significant differences were seen for the comparisons of contrast levels 1 and 2, 1 and 3, 2 and 4, and 3 and 5. As in the previous cases, detection at contrast levels 4 and 5 was most likely accurate enough that the increase in contrast did not result in a significant increase in detection.

ROC Analysis of Effect of Lesion SNR

The data from the SNR study were analyzed in a similar fashion to the contrast and size data already described. The A_zs for each reader by contrast level and total number of counts in the acquisition are shown in Figure 3B. As expected, the general trend of improving detection performance, as measured by A_z, was observed for both an increase in contrast and an increase in the total number of counts. In Figure 5, the average area under the ROC curve across all readers was plotted versus the described SNR measures for each noise level. The SNR was normalized to the lowest value (7×10^6 counts, contrast level 1). For both measures of SNR, the average area under the ROC curve showed an approximately linear increase with SNR for all noise levels considered. This trend was particularly evident in the image-based method in Figure 5A. In Figure 5A, detectability was seen to improve from random to superior performance as the relative SNR was doubled. In comparison, when plotted using the SNR measured before reconstruction in Figure 5B, an increase by a factor of 3.0–3.5 was required for the same change in detectability. This difference was attributed to the effects of the filtered backprojection reconstruction and low-pass filtering on the lesion contrast and background noise level.

LROC Analysis of Effect of Lesion Contrast

The LROC analysis described previously was applied to the ratings and localization information from each reader for each lesion size. Aggregate data of the average $A_{z,LROC}$ for each lesion size and contrast level versus contrast (measured by well counter and image pixels) were reported in Figure 6. These plots are similar to Figure 4, with average LROC results replacing average ROC results. The overall trends observed in the ROC analysis were largely confirmed by the inclusion of localization information in the LROC analysis. Regardless of the lesion size, a 2-fold increase in lesion contrast, determined by either method, resulted in an increase in the $A_{z,LROC}$ across all readers, from marginal detection performance ($A_{z,LROC} < 0.3$) to superior detection performance ($A_{z,LROC} > 0.8$). The statistical significance attributed to the differences in detection performance were determined by application of a Student t test and calculation of P values. P values were reported in Table 5 in the same format as Table 4. Because the SE for the estimates of $A_{z,LROC}$ was approximately twice that of the SE for the estimates of Az, P from the LROC analysis indicated lower levels of statistical significance than those calculated for the ROC results reported in Table 4.

LROC Analysis of Effect of Lesion SNR

LROC analysis was repeated using the ratings and localization data for the SNR dataset. Each of the 15 possible SNR combinations from the 5 levels of contrast and 3 levels of noise was analyzed for each reader. Degenerate
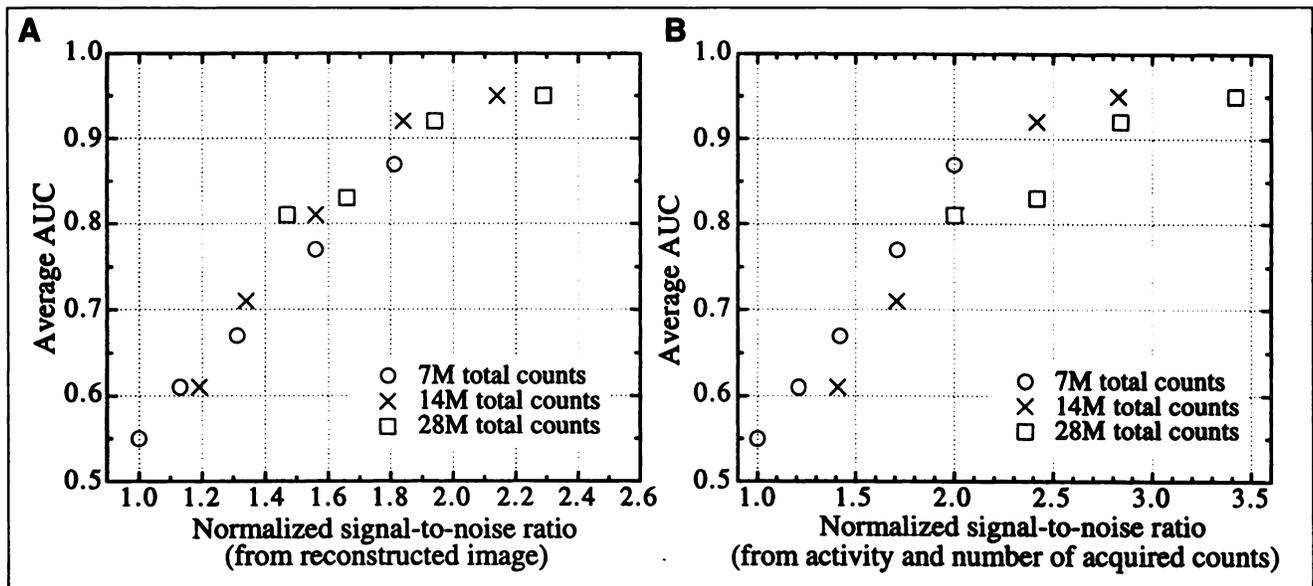
FIGURE 5. Average area under ROC curve (AUC) versus SNR. All SNR values were normalized to lowest (7 × 10⁶ counts; contrast level 1). (A) SNR was calculated as lesion contrast (measured from reconstructed images) divided by relative noise level (%SD of ROIs in reconstructed image). (B) SNR was calculated as lesion contrast (measured with well counter) divided by relative noise level expected from total number of acquired counts.

results were obtained for readers 3, 7, and 8; these results were not included in subsequent analysis. The average area under the LROC curve across the remaining readers was plotted versus the SNR measures for each noise level in Figure 7. As before, the SNR was normalized to the lowest value (7 × 10⁶ counts, contrast level 1). Again, the trend observed with the inclusion of location information in the LROC data supported the results found with the ROC data. For both measurements of SNR, a steady increase in detection performance from $A_{z,LROC}$ ~ 0.0 to $A_{z,LROC}$ ~ 1.0

was seen as the relative SNR increased from 1.0 to 3.0. Statistical tests on this data were not conducted because of the difficulties with degenerate results encountered with the SNR and LROC data.

DISCUSSION

Effect of Lesion Size on Detectability

The differences between the contrast measured in the reconstructed images and the contrast determined through readings with a well counter were well explained by known
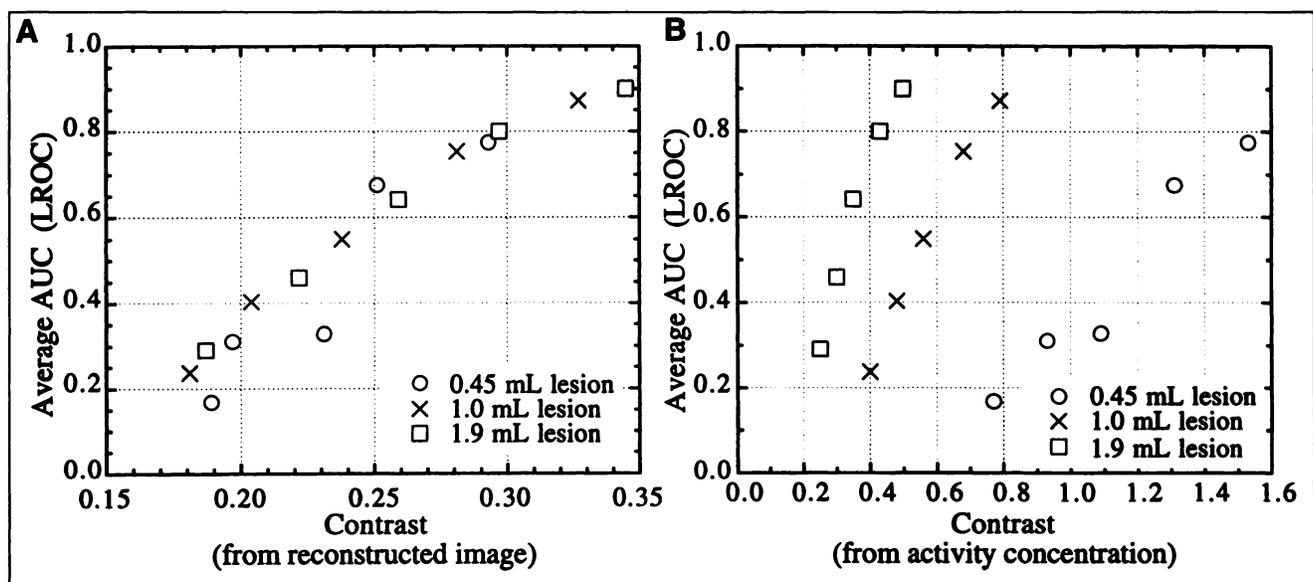
FIGURE 6. Average area under LROC curve (AUC) versus contrast. (A) Image contrast is measured from pixels in reconstructed image. (B) Image contrast is measured using activity concentration.

TABLE 5
P of Pooled LROC Results

Contrast levels compared	Lesion volume (mL)		
	0.45	1.0	1.9
2 vs. 1	0.39	0.37	0.41
3 vs. 1	0.42	0.12	0.09
4 vs. 2	0.14	0.06	0.10
5 vs. 3	0.06	0.08	0.12
5 vs. 4	0.52	0.46	0.45

Statistical comparison of $A_{z,LROC}$s between contrast levels for reader-pooled data analyzed with Student *t* test. *P*s are reported for same contrast level comparisons as in Table 4, for each of 3 lesion-sphere volumes. *P* can be used to establish level of statistical significance for observed trend of increasing detection over range of contrasts studied.

effects of object size, especially partial-volume effects (*11,12*). Measuring the contrast in the reconstructed images accounted for decreased image contrast as a result of smaller object size. When compared using the more directly relevant measure of contrast in the reconstructed image, detection performance was not seen to vary significantly between the lesion volumes in this study.

The resolution (given by full width at half maximum) measured for the Hann filter used in the filtered backprojection reconstruction of whole-body studies on the ECAT EXACT HR+ 962 was approximately 1.4 cm. Therefore, only the 1.9-mL lesion sphere, with a diameter of 1.54 cm, had a diameter greater than the resolution of the reconstructed images. Accordingly, data for the 3 lesion volumes were expected to reflect the usual effects of object size,

particularly reduced contrast recovery because of partial-volume effects, smoothing by the reconstruction filter, and finite sampling. Because of these effects, detection performance among the 3 lesion volumes was closely matched when contrast in the image domain was considered and not the contrast from the actual activity concentration.

Effect of Lesion Contrast on Detectability

The results of the individual-reader and pooled-reader ROC and LROC analyses showed that as the contrast measured in the reconstructed image was varied from 0.2 to 0.3, detection of the lesion spheres improved from nearly random detection ($A_z < 0.6$, $A_{z,LROC} < 0.3$) to superior detection ($A_z = 0.90$, $A_{z,LROC} > 0.75$). Although there was sizable statistical uncertainty for many of the individual measurements, the overall trend was established with sufficient statistical certainty. These results suggested that only a 50% improvement in contrast was necessary to make lesions at the borderline of detectability readily apparent. Even contrast differences of as little as 10%–20% resulted in statistically significant improvements in detection as measured by A_z. Thus, techniques that offer modest improvements in contrast recovery or related parameters such as resolution can be expected to noticeably improve detection performance.

In addition, the detection performance observed in this study can serve as a useful starting point for further investigation of lesion detection using PET. Two possible extensions include the preparation of datasets for additional ROC studies and for the validation of computer observers to approximate the performance of human observers. In the first instance, selection of a dataset that provides marginal
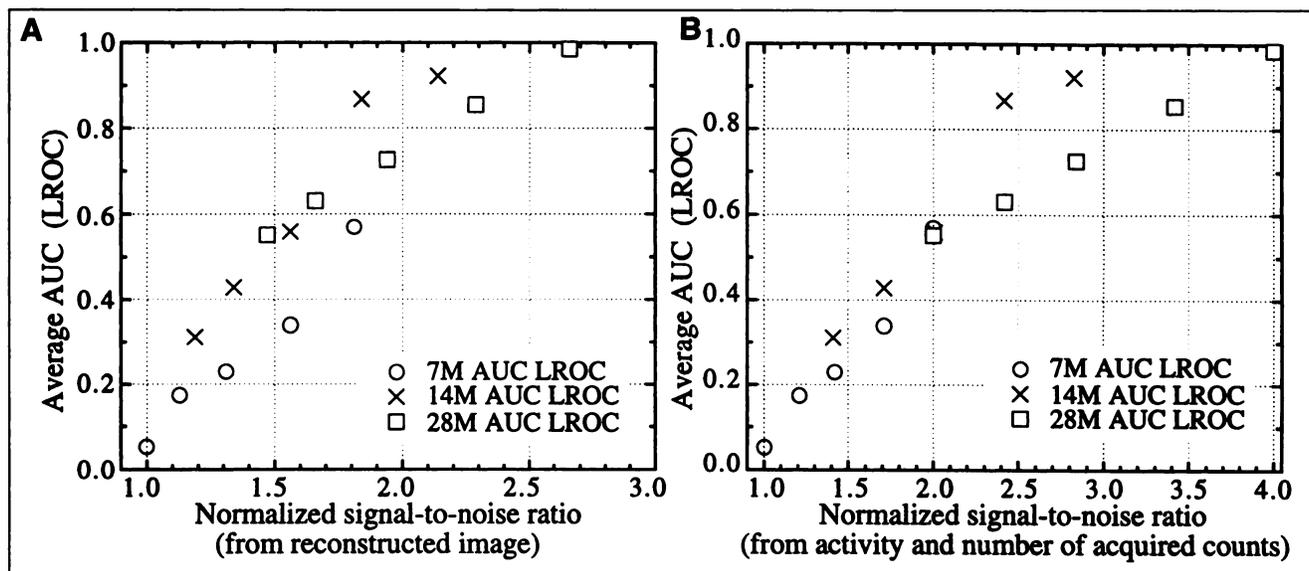
FIGURE 7. Average area under LROC curve (AUC) versus SNR. All SNR values were then normalized to lowest (7 × 10⁶, contrast level 1). (A) SNR was calculated as lesion contrast (measured from reconstructed images) divided by relative noise level (%SD of ROIs in reconstructed image). (B) SNR was calculated as lesion contrast (measured with well counter) divided by relative noise level expected from total number of acquired counts.

detection performance (in the range of $A_z = 0.75–0.80$) is desirable when comparing 2 proposed modalities (7). The results for the lesion sizes, contrast levels, and SNRs considered in this study can be used to guide proper selection of patient datasets or preparation of simulated datasets for use in an ROC or LROC study. Use of a computer observer has been suggested as an alternative to a full-scale ROC study comparing 2 possible methods (13). The results presented here could serve as a useful benchmark to calibrate the performance of a computer observer with the population of human observers participating in this study.

Effect of Lesion SNR on Detectability

This investigation showed the significant differences in detection performance of borderline lesions brought about by varying the SNR. The transition from random detection ($A_z \sim 0.5$, $A_{z,LROC} \sim 0.0$) to nearly perfect detection ($A_z \sim 1.0$, $A_{z,LROC} > 0.9$) was clearly delineated as the SNR was improved by a factor of 2–3, depending on the SNR measure chosen. The results provided an estimate of the detection performance that might be expected for borderline lesions with an improvement in the SNR. As with the contrast data, these results could be used to estimate the benefit from a method that reduces noise variance or improves SNR.

Limitations

Although several factors were considered in the design of this study, an even greater number of factors were fixed and their effects were not considered. Because of this, the results of this study are most relevant within the particular imaging environment considered. For example, the chosen paradigm, identifying a small focus of increased uptake, is applicable to whole-body PET for oncology but is less relevant for imaging of the heart or brain, where cold lesions with decreased uptake are both important and common. Also, the homogeneity and consistency in the "anatomy" of the phantom clearly simplifies the detection task compared with that in an actual patient scan. Moreover, image interpretation in a clinical setting is influenced by other sources of clinical information, such as patient history and presentation, laboratory results, and prior imaging studies, which could not be considered in this study. In addition, the lesion contrast, calculated in terms of the standardized uptake value, is often used in the determination of malignant versus benign lesions. This study only addressed the identification of a focal lesion from the background. The subsequent determination of clinical importance is beyond the scope of this particular investigation.

CONCLUSION

An ROC/LROC study was conducted with 9 observers, each with expertise reading PET images, who evaluated images of a thorax phantom. Less than a 2-fold increase in contrast was necessary to improve detection performance from poor to nearly perfect, suggesting that a relatively minor improvement in contrast (50%–75%) will noticeably

improve lesion detection performance. This improvement in detection with increasing contrast was established with adequate statistical significance. Lesion detection performance, as measured by average area under the ROC and LROC curves, was also shown to increase from marginal to nearly perfect as SNR was increased by a factor of 2 (measured in the reconstructed image), indicating that a relatively modest improvement in SNR can give a significant improvement in lesion detectability. Along with the results from varying the contrast, these results could be used to estimate the benefit from a method that improves resolution and contrast recovery, reduces noise variance, or increases SNR. Although the precise numerical results will almost certainly differ for an application with different acquisition parameters, viewing conditions, or reconstruction protocol, the general trends and relative importance of the factors considered in this study are expected to apply to a range of PET imaging applications.

REFERENCES

1. Kinahan PE, Karp JS. Figures of merit for comparing reconstruction algorithms with a volume-imaging PET scanner. *Phys Med Biol.* 1994;39:631–642.
2. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys.* 1996;23:1709–1725.
3. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol.* 1986;21:720–733.
4. Judy PF, Swensson RG, Szulc M. Lesion detection and signal-to-noise ratio in CT images. *Med Phys.* 1981;8:13–23.
5. Meaney TF, Raudkivi U, McIntyre WJ, et al. Detection of low-contrast lesions in computed body tomography: an experimental study of simulated lesions. *Radiology.* 1980;134:149–154.

6. Lim CB, Kyung SH, Hawman EG, Jaszczak RJ. Image noise, resolution, and lesion detectability in single photon emission CT. *IEEE Trans Nucl Sci.* 1982;29:500–505.

7. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol.* 1989;24:234–245.

8. Metz CE, Wang PL, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck H, ed. *Information Processing in Medical Imaging.* The Hague, Netherlands: Martinus Nijhoff; 1984:432–445.

9. Berbaum KS, Dorfman DD, Franken EA Jr. Measuring observer performance by ROC analysis: indications and complications. *Invest Radiol.* 1989;24:228–233.

10. Dorfman DD, Berbaum KS, Metz CE. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol.* 1992;27:723–731.

11. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr.* 1979;3:299–308.

12. Huang SC, Hoffman EJ, Phelps ME, Kuhl DE. Quantitation in positron emission computed tomography: 3. Effect of sampling. *J Comput Assist Tomogr.* 1980;4:819–826.

13. Chan MT, Leahy RM, Mumcouglu EU, Cherry SR, Czernin J, Chatziioannou A. Comparing lesion detection performance for PET image reconstruction algorithms: a case study. *IEEE Trans Nucl Sci.* 1997;44:1558–1563.