Effects of Scatter Subtraction on Detection and Quantitation in Hepatic SPECT

Daniel J. de Vries, Michael A. King, Edward J. Soares, Benjamin M.W. Tsui and Charles E. Metz

Department of Nuclear Medicine, University of Massachusetts Medical School, Worcester; Department of Mathematics, College of the Holy Cross, Worcester, Massachusetts; Department of Biomedical Engineering and Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; and Department of Radiology, University of Chicago Medical Center, Chicago, Illinois

The purpose of this investigation was to examine the effects of subtractive scatter compensation methods on lesion detection and quantitation. **Methods:** Receiver operating characteristic (ROC) methodology was used to measure human observer detection accuracy for tumors in the liver using synthetic images. Furthermore, ROC results were compared with mathematical models for detection and activity quantitation to examine (a) the potential for predicting human performance and (b) the relationship between the detection and quantitation tasks. Images with both low and high amounts of scatter were compared with the ideal case of images of primary photons only (i.e., perfect scatter rejection) and with images corrected by subtracting a scatter image estimated by the dual photopeak window method. **Results:** With low contrast tumors in a low count background, the results showed that scatter subtraction improved quantitation but did not produce statistically significant increases in detection accuracy. However, primary images did produce some statistically significant improvements in detection accuracy when compared with uncorrected images, particularly for high levels of scatter. **Conclusion:** Although scatter subtraction methods may provide improved activity quantitation, they may not significantly improve detection for liver SPECT. The results imply that significant improvement in detection accuracy for the conditions tested may depend on the development of gamma cameras with better scatter rejection.

Key Words: lesion detection; receiver operating characteristic analysis; scatter correction

J Nucl Med 1999; 40:1011–1023

Scattering of photons within the patient is one of the several sources of degradation in image quality and quantitative accuracy in SPECT. Scattered photons primarily reduce contrast and also degrade spatial resolution. Quantitation is affected by the presence of scattered photons that are mispositioned in the image with respect to the location of the emission of the photons. The problems presented by scatter are particularly important for low energies, such as those of ^{201}Tl and ^{99m}Tc, where the change in energy due to Compton scatter is often small enough to prevent adequate discrimination of scattered photons from primary (i.e., nonscattered) photons using a conventional NaI gamma camera.

Various methods have been proposed to compensate for the adverse effects of scatter, and most evaluations have focused on measures of the fidelity of image pixel values and quantitative accuracy. Several reviews and comparative assessments of some of these methods have been presented (*1–5*).

A common strategy used for scatter compensation involves the subtraction of an estimate of the scatter within the photopeak energy window from the photopeak image. The assumption is that scatter represents an error with respect to the primary image, and methods for reducing or removing scatter represent an attempt to correct an image. Various methods for scatter subtraction have been proposed. Some approaches use scatter images acquired outside of the photopeak region of the energy spectrum to estimate the scatter distribution in the photopeak image. Increased quantitative accuracy and contrast were reported for the dual window subtraction method (*6*), where pixels of an image from a Compton scatter window were scaled by a factor, k, and subtracted from the photopeak image. However, the subtraction resulted in increased noise fluctuations, which gave a signal-to-noise ratio (SNR) that was comparable to that in the uncorrected image (*6*). A similar method, proposed for obtaining accurate SPECT quantitation, was based on the assumption that the shape of the scatter energy spectrum can be approximated at each pixel with a triangle if there is a single emission peak or with a trapezoid if scatter spills down from multiple peaks (*7*). By using scaled, narrow windows abutted to either side of the photopeak window, the triple energy window method gives an improved estimate of the scatter distribution. However, the low number of counts detected in the narrow windows may present a disadvantage in comparison to the dual window subtraction method (*8*).

Other proposed subtraction methods estimate scatter using only the data contained within the photopeak image. A convolution-subtraction method was proposed, which estimated the scatter image by convolving a decaying exponential function with the photopeak image (*9*). Other methods have been based on dividing the photopeak into two

subwindows, where either a direct difference of the two images was used to remove scatter from one (*10,11*) or a ratio of counts in the two windows was used to subtract an estimate of scatter from the summed subwindow images (*12,13*). For any of these corrections, a gain in contrast or in quantitative accuracy comes at the expense of reducing the counts in the projection images set and altering the noise characteristics of the reconstructed SPECT images.

The references cited for reviews and assessments of scatter compensation methods provide a fairly good representation of the means that have been used to evaluate the methods. They include comparison of total counts in images with and without correction to a reference image of the source in air (*2*), spatial resolution before and after correction (*2*) and the corrected image to a reference image using the normalized mean square error (*3,5*). Additional means used to evaluate the methods include measurements of lesion contrast with and without correction (*2,3*), calculations of SNRs using lesion contrast and the standard deviation of counts in the background (*2,4*), activity recovery ratios (*2,3*), the root mean square of the percent relative error in the corrected image (*5*), plots of concentration ratios from corrected images versus the true concentration ratios (*4*) and plots of counts in corrected images versus the true counts (*4*).

Given that scatter correction by any method will change the contrast and the magnitude and texture of noise in an image, it cannot be assumed that the result will necessarily provide improved detection accuracy for human observers. A thorough examination of the effect of scatter correction must include an assessment of detection accuracy. Receiver operating characteristic (ROC) analysis is the widely accepted methodology for testing human observer detection accuracy in medical imaging (*14,15*).

ROC analysis was used in a study that had implications for the effect of scatter on detection in planar images (*16*). Rolland et al. (*16*) evaluated the effect on detection of deconvolution filtering of long-tailed point spread functions, which can arise from physical processes such as scatter. The authors reported significant improvement in detection accuracy as a result of the filtering.

The effect of scatter reduction on detection in SPECT images was studied by Staff et al. (*17*), with a comparison of uncorrected images with spatial filtering by the energy-weighted acquisition method (*18*). Three human observers were required to locate and detect "cold" lesions in a cylindrical phantom that presented an unstructured, hot background and in a Hoffman brain phantom. The results were pooled over observers and were analyzed with methods typically used for ROC studies of the simple detection task (i.e., where localization is not part of the task). Examination of the ROC curves reportedly demonstrated a statistically significant increase in the area under the curve for energy-weighted acquisition compared to no correction, but only for the cases with structured background (i.e., the brain phantom).

Similarly, we have shown with ROC studies that the removal of scatter may increase the accuracy of the simple detection task only for particular conditions (*19*). For low contrast, "cold" tumors in synthetic hepatic SPECT images, our dual photopeak window (DPW) scatter subtraction method (*13*) did not produce a statistically significant increase in detection accuracy. However, the ideal case of primary images showed significantly higher detection accuracy in comparison to the uncorrected photopeak images.

Given that ROC experiments can be costly in terms of time and resources, there has been an interest in finding ways to predict human observer performance for various tasks. Based on signal detection theory, mathematical algorithms have been derived to serve as "model observers" for the assessment of image quality (*20–22*). These mathematical observers can produce task-dependent, physical SNRs based on image parameters (*23,24*) and have been used to evaluate or optimize the design of imaging systems or to evaluate or predict human observer performance (*25,26*).

In the context of medical imaging, detection of a signal (e.g., a lesion) and activity quantitation are examples of classification and estimation tasks, respectively. Barrett (*23*) has presented several mathematical models for calculating figures of merit for these tasks in terms of SNRs. Furthermore, mathematical relations between the SNRs of several analogous pairs of classification and estimation tasks were derived (*23*).

With a need for additional information regarding the effect and utility of scatter correction in SPECT, we conducted experiments that were designed to study the effects of scatter subtraction on detection and quantitation. By using synthetic images, we intended for the experiments to approximate clinical imaging conditions, while isolating the effects of scatter subtraction. The effects on the simple detection task were evaluated with human observers using ROC analysis. Furthermore, detection and quantitation were evaluated with one pair of analogous mathematical models. The object was a focal lesion in the liver, and both cold and hot contrasts were considered.

For evaluation of quantitative accuracy, the region-of-interest (ROI) estimator, although suboptimal, represents a common approach to extracting quantitative information from image data (*23*). Counts within an ROI applied to a reconstructed image are simply summed. The related detection model is the non-prewhitening (NPW) matched filter, a quasi-ideal model observer. Whereas the ideal observer detects all information needed for a given task and maximizes the sensitivity at any prescribed level of specificity (*21*), the accuracy of both the human and NPW observers is degraded by noise correlations (*24,26*), which can either mimic or mask the lesions to be detected in SPECT images. We compared ROC results from human observers with the results from the NPW observer and ROI estimator to examine the potential for predicting human detection perfor-

mance and the relationship between the accuracy of detection and of quantitation.

MATERIALS AND METHODS

Simulated Clinical Conditions

Synthetic SPECT images were used for the experiments, because they enabled us to determine the truth regarding the presence of a lesion and to separate the primary and scatter components of an image. The SIMIND Monte Carlo simulation software (27) was used to propagate photons through a digitized, anthropomorphic phantom obtained from CT images (28). High-count projections were produced with primary and scatter images saved separately (26 million photons were emitted per SPECT projection set).

The biodistribution of FO23C5 anti-carcinoembryonic antigen (CEA) antibody fragments was approximated (29). From averaging over 7 patients at 3 to 5 h postinjection, the percent dose (and standard deviation) in the whole organ was 9.1 ($\pm$ 2.2), 1.3 ($\pm$ 0.8) and 18.7 ($\pm$ 6.1) for the liver, spleen and kidneys, respectively. Motivated by results from preliminary ROC experiments, we increased scatter into the liver region by increasing the dose in the spleen to a level 3 SDs higher than the mean. The simulated activity distribution had 22.8% of the total activity in the liver, 8.8% in the spleen and 46.6% in the kidneys, with the remaining 21.8% distributed throughout other organs as an approximately uniform background. Given that the specificity of the FO23C5 anti-CEA antibody fragment could result in either cold or hot tumors relative to the uptake in the normal liver, each contrast polarity was studied in separate experiments.

To produce images representing multiple patient cases, several locations in the liver were used for simulated tumors and multiple, independent Poisson noise realizations were added to the projection images. Tumors were represented by 2.5-cm-diameter spheres, a size that was considered sufficient to allow changes in contrast— which were produced by varying the amount of scatter—to be perceived, even though partial volume effects could reduce the measured activity contrast by about 20% (30). Locations were selected such that the distance to any liver boundary was greater than the tumor radius plus twice the full width at half maximum (FWHM) of the camera system to avoid having a tumor easily recognized by a change in the appearance of the edge of the liver. (This constraint required elimination of normally cold regions, such as hepatic veins and arteries, and the biliary tree.) Tumor projection sets were scaled to obtain the desired contrast and were then either added to or subtracted from the liver projections to produce hot or cold tumors, respectively.

The locations studied are illustrated in Figure 1, which shows high-count (i.e., approximately noise-free) primary SPECT images of both contrast polarities. The locations were selected to represent a range of location-dependent SNRs, as measured in the projection data by the ideal observer (19). The contrast was specified by the ratio:

$$\text{contrast} = \frac{N_{tmr} - N_{lvr}}{N_{lvr}}, \qquad \text{Eq. 1}$$

where N_{tmr} and N_{lvr} were the numbers of photons emitted per voxel from the tumor and liver, respectively. A contrast of 13% (-13% for cold tumors and $+13\%$ for hot tumors) was determined in preliminary experiments to produce an appropriately difficult task (19).

The simulated SPECT camera system approximated the characteristics of a Picker PRISM camera (Picker International, Inc., Cleveland, OH) with a low-energy, ultra-high-resolution (LEUHR) parallel-hole collimator, a circular radius of rotation at 21.5 cm, 9.4% FWHM energy resolution at 140 keV, and an intrinsic resolution of 0.28-cm FWHM. With a pixel size of 0.36 cm, 128 $\times$ 128 projection images were produced at 128 viewing angles.

For both contrast polarities, five treatments of SPECT images were evaluated: ideal scatter subtraction, represented by photopeak images containing only primary (i.e., nonscattered) photons; ^{99m}Tc images with no scatter correction and scatter fractions typically in the range of 0.4–0.5 (denoted as low scatter); images of a hypothetical radioisotope with scatter fractions in the range of 1.0–1.2 (denoted as high scatter), which could arise from cases of multiple emission energies (e.g., ^{201}Tl and ^{111}In); and images with DPW scatter subtraction applied to both the low- and high-scatter cases. The DPW method, which was being considered for clinical use when our initial ROC studies were being designed, was selected as a clinically feasible method that was fairly representative of the various energy-based, scatter subtraction methods that use multiple energy windows to estimate scatter in the photopeak region. The five treatments are listed in Table 1, with the total counts in the projection sets used to obtain the SPECT images.

Acquisition of the projection images was simulated with a 20% energy window centered on 140 keV. Projections for the high-scatter conditions were produced by scaling the scatter component of the ^{99m}Tc projections by a factor of 2.5. The DPW correction of projections was implemented as described by de Vries and King (13), using simulated calibration experiments. The dual windows were 5% and 15% in width and abutted at 7 keV below the photopeak (i.e., W_L, 126–133 keV; W_U, 133–154 keV). The scatter estimate for the total 20% window was obtained from a power function that related the scatter-to-total count ratio (STR) in the total window to the lower-to-total window count ratio. For the ith
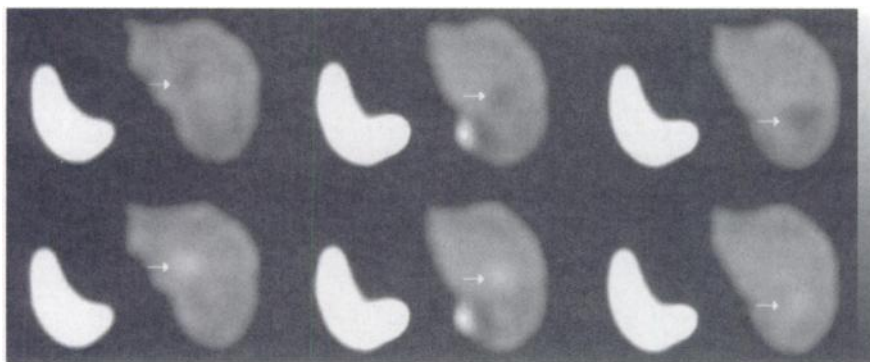
FIGURE 1. From left to right, the tumor locations that gave lowest, mid-range and highest SNRs, which were calculated using ideal observer applied to noise-free primary projection set. Arrows point to lesion for both cold (upper row) and hot (lower row) contrast polarities. (Thresholding was applied to increase the visibility.)

pixel of the total photopeak image, the calculated STR was:

$$STR[i] = A \times \left(\frac{W_L[i]}{W_L[i] + W_U[i]}\right)^B + C, \qquad \text{Eq. 2}$$

where the values of the coefficients, A, B and C, were A = 2.792, B = 1.015 and C = −0.2749 for the low-scatter condition, and A = 2.539, B = 0.6865 and C = −0.5486 for the high-scatter condition. The product of the STR and counts in the total image at each pixel produced the scatter estimate, which was smoothed with a Wiener low-pass filter before subtraction (13).

The synthetic images provided the ability to increase the statistical power of the ROC analysis through the use of correlated images (31). For a particular tumor location and noise realization (i.e., a "case"), an image for each treatment was constructed by adding the appropriate scatter image to the primary image to produce uncorrected and DPW-corrected images for both levels of scatter.

SPECT images (128 × 128, 0.36-cm pixels) were reconstructed with techniques that were typically used in our clinic with commercially available software. Ramp-filtered backprojection was used, with noise-suppression prefiltering by a two-dimensional Butterworth low-pass filter (order 4; cutoff frequency 0.25 cycles/ cm). The SPECT images, which contained at most one tumor, were reconstructed through the center of the tumor. Multiplicative Chang attenuation correction was applied (32), using an elliptical attenuation map that approximated the cross section of the slice. The narrow-beam attenuation coefficient at 140 keV, used by the Monte Carlo program (μ = −0.1546 cm^{-1}), was applied to primary and DPW-corrected images. Effective attenuation coefficients (μ_E) and build-up factors (B_0) for broad-beam attenuation were calculated (33) and applied to the uncorrected images (μ_E = −0.1163 cm^{-1} and B_0 = 1.079 for low scatter; μ_E = −0.0831 cm^{-1} and B_0 = 1.177 for high scatter).

Human Observer Detection Accuracy

The detection experiments were designed for signal-known-exactly (SKE) conditions, in which lesion size and shape were constant and the location to be evaluated was indicated by cross-hairs, which could be toggled on and off by the observer. Software developed at the University of North Carolina at Chapel Hill was used for the display and rating of images (34).

Before display, each SPECT image was magnified by a factor of 3 using bilinear interpolation, and the central 256 × 256 pixel region was extracted to fill an area of 25 cm^2 on the display monitor. Furthermore, each image was scaled such that the background (i.e., liver) would be displayed at nearly the same gray level for both cold and hot tumor experiments. Pixel values that were negative or zero were mapped to a gray level of zero; the mean liver pixel value (excluding the tumor) was mapped to the mid gray-scale level (out of 127 levels). The scaling determined an

TABLE 1
Treatments Studied and Count Levels

Treatment	Total counts in projection set
A: Primary (ideal correction)	8.77 × 10^6
B: Uncorrected low-scatter	12.50 × 10^6
C: Uncorrected high-scatter	21.82 × 10^6
D: DPW-corrected low-scatter	8.55 × 10^6
E: DPW-corrected high-scatter	8.25 × 10^6

upper threshold, above which pixels values were clipped. Given the high activity in the spleen and kidneys, the clipping of high values did not occur in liver pixels.

An image contained either one lesion or none, and the locations evaluated by the observers in the lesion-present images were the same locations evaluated in the lesion-absent images. As each image was presented sequentially, a continuous scale from "definitely absent" to "definitely present" was used to rate an observer's confidence regarding the decision about the absence or presence of a tumor (35).

For each treatment, there were 240 images for rating in the study sessions. These images were produced from the three tumor locations, each having two signal conditions (i.e., present or absent) and 40 different Poisson noise realizations for each signal condition. The 240 images were divided into two subsets of 120 images each, with the reading order of the images in the subsets randomized for each observer. The subsets for each of the five treatments were arranged in 10 study sessions, such that the subsets were read in a different order by each observer.

Before each study session for a particular treatment, an observer was trained by first viewing the noise-free signal-present and signal-absent images for each tumor location and then rating 60 noisy images (10 noise realizations for each of two signal conditions at three tumor locations). When the observer entered a confidence rating, the corresponding noise-free image was displayed adjacent to the noisy image, and the observer was told whether or not a tumor was present. The 60 images were presented in a different order before the reading of the second subset of study images for a treatment. From this experience, an observer was expected to learn to identify lesions and to set rating strategies.

Images from each treatment were read independently by seven observers, who were members of the medical physics research group of the Department of Nuclear Medicine at the University of Massachusetts Medical School. Readings took place in a dark room, with no constraints on either the time allotted for reading images or the viewing distance. However, observers were not allowed to change the brightness and contrast levels of the display monitor, given that a constant transfer function for mapping gray-scale pixel values to luminance was desired.

The transfer function of the monitor was modified with a mapping that ensured that equal steps in gray-scale values would be perceived as equal differences in luminance, given that the human visual system response as described by Weber's Law shows a logarithmic relationship between intensity and perceived brightness (36). The modified mapping produced the desired log-linear transfer function and, thus, the effects of monitor characteristics on displayed contrast were reduced. The measured luminance ranged from 0.1 to 57.4 foot-lamberts (0.343–196 cd/m^2). The mean background luminance in the liver was in the range from 3.0 to 3.5 foot-lamberts. To eliminate background luminance from all regions except the region used for reading and rating the images, we placed a mask on the monitor.

For each observer and treatment, a fitted ROC curve and the area under the curve (AUC), A$_z$, were estimated using the LABROC1 program (35). The average ROC curve for each treatment was produced by averaging the curve parameters, a and b, from the seven observers. The perceptual SNR of the human observers was then calculated by converting the AUC to the detectability index, d$_a$, using the relation

$$AUC = \frac{1}{2} + \frac{1}{2}\,\text{erf}\left(\frac{d_a}{2}\right), \qquad \text{Eq. 3}$$

where erf(·) is the error function. The value of d_a was regarded as the human observer SNR.

To perform multiple comparisons of treatment SNRs while limiting type I errors (i.e., rejection of a true null hypothesis), we conducted a two-way analysis of variance (ANOVA) to determine whether or not there was evidence of a real, nonrandom difference between the treatment means (averaging over observers). The null hypothesis, H_0, was that all treatments were equal, whereas the alternative hypothesis, H_A, was that there were real differences between the treatments. The null hypothesis was rejected when the value of the calculated test statistic was greater than the critical point of the upper 5% of the F distribution with degrees of freedom of 4 and 24, that is, $F_{4,24} = 2.78$ for $\alpha = 0.05$.

When the null hypothesis of the ANOVA was rejected, Scheffé's test for multiple paired comparisons was applied (at the 5% level) to examine the significance of seven differences in mean SNRs between treatments: primary versus uncorrected low scatter (A versus B); primary versus uncorrected high scatter (A versus C); primary versus DPW-corrected low scatter (A versus D); primary versus DPW-corrected high scatter (A versus E); uncorrected low scatter versus uncorrected high scatter (B versus C); DPW-corrected low scatter versus uncorrected low scatter (D versus B); and DPW-corrected high scatter versus uncorrected high scatter (E versus C). The null hypothesis for Scheffé's test was

$$H_0: \sum_{i=1}^{c} k_i SNR_i = 0, \qquad \text{Eq. 4}$$

where there were c treatments; SNR_i was the mean SNR for treatment i (averaged over observers) and the k_is were constants assigned the values -1, 0 or $+1$, which summed to zero and produced the desired differences between pairs of SNRs. The Scheffé test statistic was given by

$$S = \frac{\left(\sum\limits_{i=1}^{c} k_i SNR_i \right)^2}{(c - 1)s^2 \left(\sum\limits_{i=1}^{c} k_i^2/r \right)}, \qquad \text{Eq. 5}$$

where s^2 was the residual mean square from the ANOVA and r was the number of human observers. Furthermore, the bounds on the 95% confidence interval were calculated from

$$F_{(c-1),(c-1)(r-1)} = \frac{[r/(c - 1)](\Delta - L)^2}{s^2 \sum\limits_{i=1}^{c} k_i^2}, \qquad \text{Eq. 6}$$

where Δ was the difference between a pair of SNRs. Equation 6 was solved for L, and the lower and upper bounds were equal to the minimum and maximum roots, respectively. A statistically significant difference between treatments at the 5% level required a test statistic greater than $F_{4,24} = 2.78$ and 95% confidence intervals that did not include zero. The calculations required for two-way ANOVA, Scheffé's test and the bounds on the confidence interval were obtained from Pollard (37).

Comparison of Human and Non-Prewhitening Observers

An imaging system can be mathematically described by

$$g = Hf + n, \qquad \text{Eq. 7}$$

where g is raw image data with M pixels arranged as an $M \times 1$ vector, the $N \times 1$ vector f is the discrete representation of the object (i.e., the pixelized distribution of activity in the patient), H is an $M \times N$ matrix that represents the imaging system and n is an $M \times 1$ vector representing noise in the image (23). An estimate, $\hat{f}$, of the object is obtained from g by the reconstruction process. An observer, whether human or mathematical, performs the simple detection task by comparing a test statistic, $\lambda(\hat{f})$, to a threshold to classify images as either signal present or signal absent.

For comparison to the human observer, the NPW observer was implemented as a matched filter. Filters, or templates, were formed by reconstructing a noise-free image through the center of the tumor for each of the three locations and five treatments. The appropriate template, w, was cross-correlated with the SPECT images, producing a linear combination of the pixel values from $\hat{f}$. This value, the test statistic, is given by the weighted sum:

$$\lambda(\hat{f}) = \sum_{i=1}^{N} w_i \hat{f}_i. \qquad \text{Eq. 8}$$

With w having the profile of the tumor, the highest values of w were placed on the locations in $\hat{f}$ where the tumor was expected to be and the values of w decreased to zero with increased distance from the center of the profile. By using for $\hat{f}$ the same images that were rated by the human observers in the ROC experiments, the effects of scaling the images for display were taken into account.

With the lesion-present images having a signal in the presence of noise and the lesion-absent images having only noise, the NPW observer produced a distribution of test statistics for both signal-present (sp) and signal-absent (sa) images. For each of the five treatments and each lesion location, the mean and variance of the sp and sa distributions were used to calculate a physical SNR from the detectability index, d_a:

$$SNR_{NPW}^2 = d_a^2 = \frac{[E(\lambda(\hat{f})|sp) - E(\lambda(\hat{f})|sa)]^2}{\frac{1}{2}[var(\lambda(\hat{f})|sp) + var(\lambda(\hat{f})|sa)]}, \qquad \text{Eq. 9}$$

where the number of sp and sa images for a given tumor location and imaging treatment were equal. The means of the test-statistic distributions are denoted $E(\lambda(\hat{f})|sp)$ and $E(\lambda(\hat{f})|sa)$, and the variances are denoted by $var(\lambda(\hat{f})|sp)$ and $var(\lambda(\hat{f})|sa)$. The numerator and denominator in equation 9 are measures of the square of the signal and of the noise, respectively. For a single lesion location and additive Gaussian noise, the NPW SNR is $\sqrt{d_a^2}$. Given that the noise in the liver of the SPECT images was a good approximation of additive Gaussian noise, that three different tumor locations were rated by the human observers in a random order and that the location to be evaluated was specified, the signal and the noise were first calculated for each location and then were averaged over location. The average signal was divided by the average noise for each treatment to calculate an NPW SNR (25).

Using the average human SNR, the statistical efficiency of the human observer with respect to the NPW observer was calculated for each treatment with the ratio:

$$\text{Efficiency} = \left(\frac{SNR_{human}^2}{SNR_{NPW}^2} \right). \qquad \text{Eq. 10}$$

For many tasks, this efficiency has been in the range of about 0.5 ± 0.2, when the displayed contrast was sufficient for the specified task (21).

The relation between the SNRs of the human and NPW

observers was examined using nonparametric correlation analysis for two reasons: (a) Knowledge of the probability distribution functions from which the SNRs were drawn was unavailable; and (b) with a relatively small number of SNRs to be examined, the assumption of normal distributions for the SNRs was questionable. By assigning integer ranks to the SNRs, a correlation between SNRs was examined using the linear correlation coefficient of the ranks. Spearman's rank correlation coefficient, ρ_S was used, with a null hypothesis of no correlation between the human and NPW SNRs (i.e., $H_0:\rho_S = 0$). The alternative hypothesis, H_A, was that there was a direct correlation of human SNRs with NPW SNRs (i.e., a tendency for large human and NPW SNR values to be paired), so a one-tailed test of the significance of ρ_S was performed.

Comparison of Detection and Quantitation

Activity quantitation is an example of a task where the strength (or magnitude) of a parameter, θ (e.g., the number of counts in a region), is desired. The strength of the parameter in a region of an object, f, may be defined as the sum of the elements that are contained within a region defined by a template, w. With the object and the template represented as arrays of N elements, the true value of θ is given by:

$$\theta = \sum_{i=1}^{N} w_i f_i. \qquad \text{Eq. 11}$$

When ROI estimation is used to measure activity in an estimate of the object (i.e., an image), the same template is applied directly to the estimate (23), $\hat{f}$, which was a reconstructed SPECT image in our experiments. The ROI estimate of θ is given by:

$$\hat{\theta}_{ROI} = \sum_{i=1}^{N} w_i \hat{f}_i. \qquad \text{Eq. 12}$$

In contrast to the NPW template, which had the profile of the lesion, the template for the ROI estimator was a circular disk, having the diameter of the lesion and pixel values of either one or zero for pixels located entirely inside or outside of the ROI, respectively. If a fraction of a pixel was included in the ROI, subsampling of the edge pixels of the disk provided pixel values between zero and one. Thus, the ROI estimator simply summed the counts detected within the ROI.

Barrett (23) recommended the use of the ensemble mean square error (EMSE) as a figure of merit for the ROI estimator to account for both bias and variance. Using notation similar to Barrett's, the bias of the estimator was given by the difference between the averaged estimates and the true value of θ:

$$b_{ROI} = E(\hat{\theta}_{ROI})_{n|f} - \theta, \qquad \text{Eq. 13}$$

where $E(\hat{\theta})_{n|f}$ is the estimate of activity in the ROI averaged over n noise realizations for a particular object, f (i.e., a lesion at a particular location). The mean square bias was defined as the squared bias averaged over all objects: $E(b_{ROI}^2)_f$. The variance of the estimator was given by

$$\text{var}(\hat{\theta}_{ROI}) = E([\hat{\theta}_{ROI} - \bar{\theta}_{ROI}]^2)_{n,f}, \qquad \text{Eq. 14}$$

where the average of the squared differences between estimates and the mean estimate is taken over all noise realizations and over all lesion locations. The EMSE was then defined as the average squared difference of the estimates and the true parameter value, θ:

$$\text{EMSE}(\hat{\theta}_{ROI}) = E([\hat{\theta}_{ROI} - \theta]^2)_{n,f}, \qquad \text{Eq. 15}$$

which is equal to the sum of the mean square bias and the variance. Finally, the squared SNR of the ROI estimator was defined as the reciprocal of the EMSE normalized by the square of the object strength (23):

$$\text{SNR}_{ROI}^2 = \frac{E(\theta^2)_{n,f}}{\text{EMSE}(\hat{\theta}_{ROI})}. \qquad \text{Eq. 16}$$

The estimation of activity in ROIs was performed using the tumor-present images of the ROC experiments before the interpolation and scaling to gray-level values for display. For both cold and hot tumors, the calculation of the "true" value of activity, θ, was done using the noise-free primary SPECT image for each tumor location. These images were corrected for attenuation but did not have the Butterworth low-pass filtering that was applied to the noisy images.

With the noise-free primary SPECT image of the tumor as the standard for the EMSE calculations, activity estimates from the primary images were affected only by noise. Therefore, the ROI SNR for the primary images was by definition the best case. This is similar in concept to the detection experiments, where the primary image was defined as the ideal scatter subtraction.

The ROI SNR of the activity estimates was calculated for each treatment and then was compared with the detection SNRs from the human and NPW observers. Spearman's rank correlation coefficient was used to test for correlation between SNRs of the detection and estimation tasks. The null hypothesis, H_0, was that $\rho_S = 0$; the alternative hypothesis, H_A, was that there was a direct correlation of the estimation SNR values with the detection SNR values.

RESULTS

Human Observer Detection Accuracy

The average ROC curves for each treatment are shown in Figure 2, for the cold and hot tumor experiments, respectively. For both lesion contrasts, comparison of the curves from primary images with curves from uncorrected images of both low- and high-scatter conditions showed an obvious decrease in detection accuracy (i.e., the AUC or SNR) as the amount of scatter increased. The effect of DPW scatter correction on detection was not as clear.

The results of the two-way ANOVA of the SNRs gave strong evidence for the existence of real, statistically significant differences between treatments. As shown in Table 2, the calculated F statistics were 13.45 ($P = 0.7 \times 10^{-5}$) and 20.26 ($P = 0.2 \times 10^{-6}$) for cold and hot tumors, respectively. With $F_{4,24} = 2.78$ for $\alpha = 0.05$, the null hypothesis that the differences between average SNRs of the treatments were equal to zero was rejected for both contrast polarities.

The results of applying Scheffé's multiple comparisons test for paired samples to the differences in the average human SNR between selected pairs of treatments are presented in Table 3 and in Figure 3 for cold and hot tumors, respectively. The primary images gave a statistically significant ($P < 0.05$) increase in detection accuracy over the uncorrected low-scatter images (A versus B) only for hot lesions. Although the ROC curves for the cold lesions appear to have a real difference, the null hypothesis (i.e., the difference is equal to zero) could not be rejected. However,
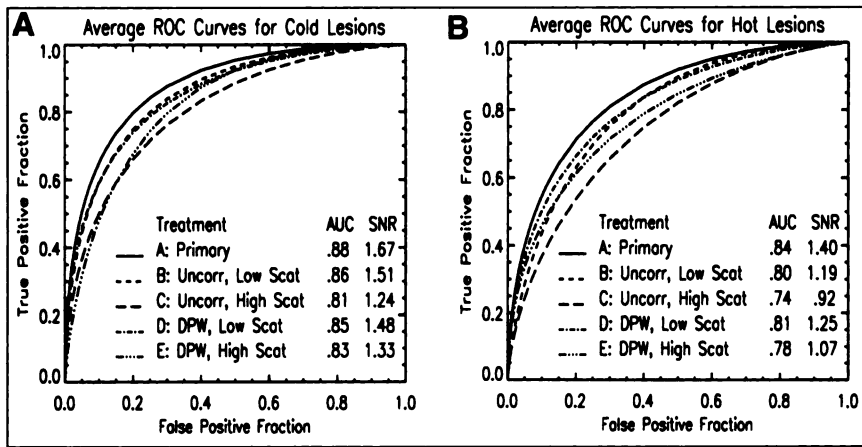
FIGURE 2. Fitted, average receiver operating characteristic (ROC) curves from seven independent human observers for experiments that included primary images, uncorrected images for low- and high-scatter conditions and dual photopeak window (DPW)-corrected images for low- and high-scatter conditions. Curves are shown for images with cold lesions (A) and images with hot lesions (B). Task was detection of 2.5-cm-diameter tumor, which had contrast of 13% (−13% cold and +13% hot) and was located at one of three possible sites. AUC and corresponding SNR are reported. (For clarity, error bars were omitted.)

for both contrast polarities, primary images gave a significant ($P < 0.05$) increase in detection accuracy in comparison to uncorrected high-scatter images (A versus C), and uncorrected low-scatter images gave significantly greater ($P < 0.05$) detection accuracy than uncorrected high-scatter images (B versus C).

Visual inspection of the ROC curves for the comparison of DPW scatter correction with uncorrected low-scatter images (D versus B) showed little difference in detection accuracy. For the high-scatter condition, curves for the comparison of DPW-corrected images versus uncorrected images (E versus C) showed more noticeable separation as the difference in the SNRs (derived from the AUCs) gradually increased. However, given the variances of the curve fits and resulting AUCs, as well as the sacrifice of power with the conservative approach to reducing type I errors inherent in Scheffé's multiple comparisons test, the difference in SNRs for any comparison of DPW scatter correction with uncorrected images was too small to demonstrate statistical significance (i.e., the null hypothesis could not be rejected).

For the comparison of primary images with DPW-corrected low-scatter images (A versus D), the null hypothesis could not be rejected. However, detection accuracy was significantly greater for primary versus the DPW-corrected high-scatter images (A versus E).

Comparison of Human and Non-Prewhitening Observers

The average human SNR, the NPW SNR and the human efficiency are reported for each treatment and contrast polarity in Table 4. Both the human and NPW observers produced lower SNRs for the hot lesions in comparison to the cold lesions.

A positive correlation between the human and NPW observers is evident in Figure 4, with an apparent separation of the points associated with the DPW-processed images (D and E) from those of uncorrected images (B and C). Linear fits to the primary and DPW-corrected data (slope = 0.41, intercept = 0.42) and to the primary and uncorrected data (slope = 0.71, intercept = −0.40), which are shown in Figure 4, further emphasize the apparent difference in the way the human and NPW observers responded to the DPW-corrected images.

With cold and hot tumor SNRs pooled, Spearman's rank correlation coefficient gave a strong and significant correlation between the ranks of human and NPW observer SNRs associated with the primary and non-DPW treatments (A, B and C) and the primary and DPW treatments (A, D and E). The correlation remained significant but decreased when all treatments were considered together. Therefore, correlation between the human and NPW SNRs was demonstrated,

TABLE 2
Two-Way ANOVA of Average Human Observer SNR

Source of variation	Cold lesions					Hot lesions				
	Sum of squares	df	Mean square	F	P	Sum of squares	df	Mean square	F	P
Between observers	0.35	6	0.058	3.94	0.007	0.25	6	0.041	3.72	0.009
Between treatments	0.79	4	0.197	13.45	0.7×10^{-5}	0.90	4	0.225	20.26	0.2×10^{-6}
Residual	0.35	24	0.015			0.26	24	0.011		
Total	1.49	34				1.41	34			

ANOVA = analysis of variance; SNR = signal-to-noise ratio; df = degree of freedom.

TABLE 3
Scheffé's Pairwise Multiple Comparisons Test for ROC Results*

	Cold lesions			Hot lesions		
Comparison	Difference of mean SNRs	Test statistic†	95% Confidence interval	Difference of mean SNRs	Test statistic†	95% Confidence interval
A versus B	0.16	1.61	−0.05, 0.38	**0.20**	3.20	0.01, 0.39
A versus D	0.19	2.16	−0.03, 0.41	0.15	1.71	−0.04, 0.33
D versus B	−0.03	0.04	−0.24, 0.19	0.05	0.23	−0.13, 0.24
B versus C	**0.27**	4.31	0.05, 0.48	**0.27**	5.81	0.08, 0.46
A versus C	**0.43**	11.19	0.22, 0.65	**0.47**	17.64	0.29, 0.66
A versus E	**0.34**	7.02	0.13, 0.56	**0.32**	8.30	0.14, 0.51
E versus C	0.09	0.48	−0.13, 0.31	0.15	1.74	−0.04, 0.34

*Boldface differences are significant ($P < 0.05$).
†Critical value: $F_{4,24} = 2.78$, $P < 0.05$.
ROC = receiver operating characteristic; SNR = signal-to-noise ratio.

and the null hypothesis that $\rho_s = 0$ was rejected with $P < 0.05$. The results are reported in Table 5. The P values were from the one-sided test with H_A specifying direct correlation.

Comparison of Detection and Quantitation

The results of activity quantitation using the ROI estimator are summarized for both cold and hot tumors in Table 6 and Figure 5. Note that the magnitude of the EMSE is inversely proportional to the SNR for the ROI estimator. The bias and variance for the primary images (A) were affected only by the random noise due to counting statistics, given that the noise-free primary image corrected for attenuation was used as the standard against which all activity estimates were compared. Bias increased as the amount of scatter in the uncorrected images increased from low (B) to high (C), with a decrease in variance as the number of counts in the images increased. The DPW-corrected images for both low and high scatter (D and E, respectively) showed the decreased bias obtained by scatter subtraction, with the cost of increased variance due to a reduction in the counts in the projection image set and an alteration of the noise characteristics of the reconstructed SPECT images.

The SNR of the ROI estimator reported in Table 6 is shown in Figure 6, which provides a visual comparison with the average human SNR and the NPW SNR. Like the detection SNRs, the ROI estimator SNR was greatest for the primary images and decreased as the amount of scatter in the images increased. Unlike the detection SNRs, a notable increase in the ROI SNR was produced by the DPW scatter subtraction.

With cold and hot tumor SNRs pooled, the comparison of the average human SNR with the ROI SNR using Spearman's rank correlation coefficient resulted in rejection of the null hypothesis (i.e., $\rho_s = 0$) at the $P < 0.05$ level only when SNRs from all treatments were considered together. For the comparison of the NPW SNR with the ROI SNR, the null

hypothesis was rejected at the $P < 0.05$ level only when the SNRs of the non-DPW treatments (i.e., A, B, and C) were considered. The results are reported in Table 5. The P values were from the one-sided test with H_A specifying direct correlation.

DISCUSSION

The ROC experiments indicated that human detection accuracy in SPECT images decreased as the amount of scatter present increased. For both contrast polarities, uncorrected low-scatter images gave significantly higher detection accuracy than the uncorrected high-scatter images. Furthermore, although the higher detection accuracy was statistically significant only for hot lesions when primary images were compared with uncorrected [99m]Tc (i.e., low scatter) images, the increase in accuracy was significant for both hot and cold lesions when primary images were compared with uncorrected high-scatter images.

However, the subtraction of estimated scatter by means of the DPW method failed to demonstrate statistically significant improvement in detection accuracy over uncorrected images. Moreover, for the high-scatter images of both contrast polarities, primary images gave significantly better detection accuracy than the DPW-corrected images. Whereas the primary and DPW-corrected images had count levels that were lower than the levels in the uncorrected images, the DPW-corrected images generally exhibited more prominent correlated noise blobs than either the primary or the uncorrected images for both low- and high-scatter cases.

The noise characteristics of the DPW images may have differed from those of the other treatments due to overcorrection of the projection sets (as seen in Table 1) and to the subtraction of smoothed estimates of scatter from the projections, as opposed to the rejection of the true scatter before image acquisition. It is known that noise correlations have an adverse effect on the ability of the human observer
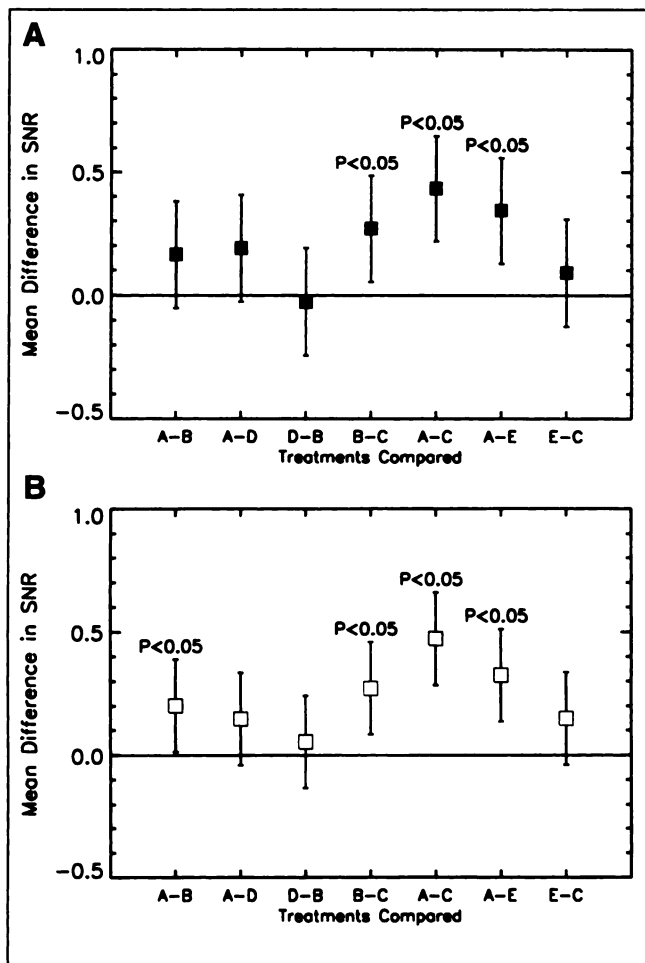
FIGURE 3. Comparisons of human observer signal-to-noise ratios (SNRs) from ROC experiments for selected pairs of treatments are presented for cold (A) and hot lesions (B). Between-treatment comparisons were made using Scheffé's method of multiple comparisons of paired samples (test statistics are reported in Table 3). Mean difference in detection SNRs, averaged over seven observers, was plotted for each comparison. Bars indicate 95% confidence intervals. Null hypothesis was that mean difference was equal to zero. Statistically significant differences are labeled with $P < 0.05$.

to detect lesions in a background (26). The liver presented a large, fairly uniform, region of activity in the SPECT images, in which noise blobs provided the primary source of distraction for the observers.

Both the human and the NPW observers indicated that detection accuracy was somewhat lower for hot lesions in comparison to cold lesions. However, the difference in magnitude did not affect the conclusion that increased scatter in the SPECT images decreased the detection SNR. Nor did the difference affect the fact that the DPW scatter subtraction produced inconclusive results with respect to detection.

The correlation of the NPW and human SNRs indicated that the NPW observer can be a useful predictor of how human observer detection accuracy is affected by scatter in SPECT images under certain conditions. There was a strong, direct correlation between human and NPW observer SNRs when uncorrected results and DPW-corrected results were considered separately. However, the NPW observer appeared to be somewhat less effective for predicting the human observers' performance when DPW-corrected and uncorrected images were considered together. The DPW correction generally caused an increase in the human SNR with respect to the uncorrected images, while the NPW SNR decreased.

The discrepancies between the two types of observers may indicate that they responded differently to the effect of the DPW correction on noise correlations. It is possible that human observers were able to do some decorrelation of image noise with partial prewhitening (38). Perhaps a model observer that included the effects of spatial frequency-selective channels in the human visual system would provide an improved prediction of human performance (24,26).

As expected, for the task of estimating activity, the results showed that quantitative accuracy decreased as the amount of scatter present increased. In addition, the SNR of the ROI estimator indicated that scatter subtraction improved the accuracy of quantitative measurements on average. Although the ROC experiments did not demonstrate a significant improvement in detection accu-

TABLE 4
Average Human Observer SNR, NPW Observer SNR and Human Efficiency

| Treatment | Cold lesions | | | Hot lesions | | |
	Average human SNR (SE)	NPW SNR	Human efficiency	Average human SNR (SD)	NPW SNR	Human efficiency
A	1.67 (0.04)	2.97	0.31	1.40 (0.02)	2.61	0.28
B	1.51 (0.02)	2.45	0.38	1.19 (0.02)	2.26	0.28
C	1.23 (0.02)	2.29	0.29	0.92 (0.01)	1.98	0.22
D	1.48 (0.02)	2.39	0.38	1.25 (0.01)	1.98	0.40
E	1.33 (0.01)	2.02	0.43	1.07 (0.03)	1.81	0.35

SNR = signal-to-noise ratio; NPW = non-prewhitening.

FIGURE 4. For both cold and hot lesions, scatter plot of average human observer signal-to-noise ratio (SNR) versus non-prewhitening (NPW) quasi-ideal observer SNR shows correlation between detection accuracy of two types of observers. Differences in dashed and dotted lines fit to primary and uncorrected data (A, B and C) and to primary and DPW-corrected data (A, D and E), respectively, suggested that human and NPW observers may respond differently to effects of DPW method on image noise.
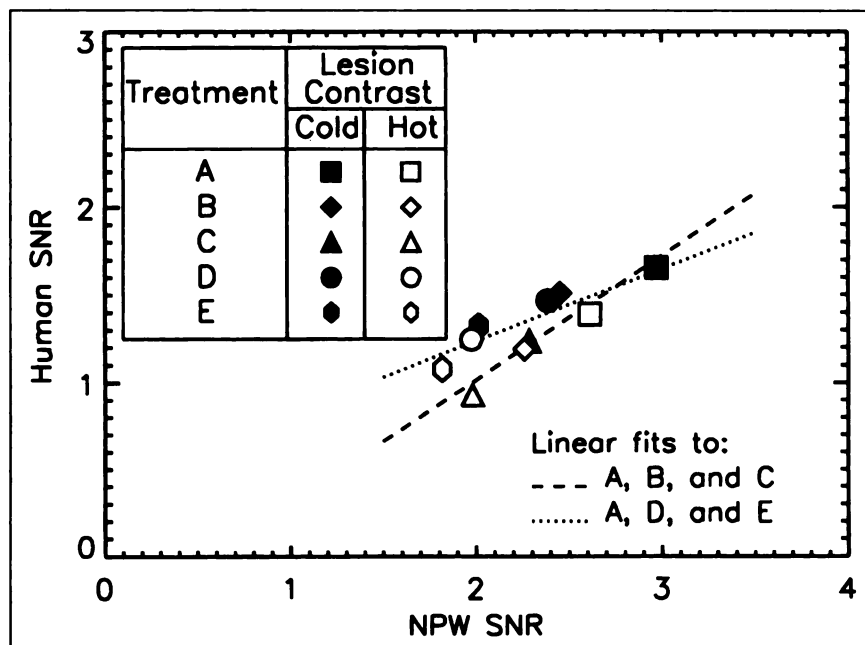
racy with DPW scatter subtraction, DPW substantially decreased the bias at the cost of increased noise, as evidenced by the increased variance. Furthermore, whereas the ROI SNR for DPW-corrected images was better than that of uncorrected images (comparing treatments D to B for low scatter and E to C for high scatter), it was always lower than the ROI SNR of primary images (treatment A).

A strong direct correlation was seen between the human and NPW detection SNRs, but the correlation of the ROI estimator SNRs with both detection SNRs (i.e., human and NPW) was weaker. Barrett (23) gave the relationship between the NPW SNR and the ROI SNR as

$$\frac{SNR_{NPW}^2}{SNR_{ROI}^2} = Q,$$ Eq. 17

where Q is a product of factors, which accounts for the different effects that characteristics of images (e.g., bias and correlated noise) have on the detection and estimation tasks. Some of these factors may explain the difference in the

strength of the correlation between the SNRs for the two tasks.

Before generalizing the results of the ROC experiments, several considerations regarding the conditions under which they were conducted should be considered. First, for the clinical application considered, the attenuating medium was nearly uniform, unlike locations such as the thorax and certain regions of the head. Second, the experiments tested only the simple detection task under SKE conditions. This required the use of images with low contrast and low counts (i.e., high noise) to obtain suitable areas under the ROC curves. Other tasks that are part of clinical decision making were not considered, such as the ability to correctly locate or to correctly determine the size of a lesion. Third, filtered backprojection was used to reconstruct the images, which produces particular artifacts and noise characteristics that differ from other algorithms, such as the iterative methods. Fourth, observers were not allowed to adjust the transfer function of the display monitor to their liking. The display monitor was carefully controlled throughout all of the

TABLE 5
Spearman's Rank Correlation Coefficient for Comparisons of Detection and Estimation SNRs*

Treatments	Sample size	Human SNR versus NPW SNR	Human SNR versus ROI SNR	NPW SNR versus ROI SNR
Primary and uncorrected (A, B and C)	6	**0.94** (P = 0.002)	0.66 (P = 0.08)	**0.77** (P = 0.04)
Primary and DPW corrected (A, D and E)	6	**0.94** (P = 0.002)	0.49 (P = 0.16)	0.66 (P = 0.08)
All treatments (A, B, C, D and E)	10	**0.81** (P = 0.002)	**0.60** (P = 0.03)	0.38 (P = 0.14)

*Boldface correlation coefficients are significant (P < 0.05).
SNR = signal-to-noise ratio; NPW = non-prewhitening; ROI = region of interest.

TABLE 6
Region-of-Interest (ROI) Estimation of Activity Using the Ensemble Mean Square Error (EMSE)

| Treatment | Cold lesions (counts in ROI = 74.1) | | | | | Hot lesions (counts in ROI = 87.3) | | | | |
	Mean estimated counts	Mean-square bias	Variance	EMSE	ROI SNR	Mean estimated counts	Mean-square bias	Variance	EMSE	ROI SNR
A	73.9	0.6	4.7	5.3	32.3	86.9	0.4	5.2	5.6	37.0
B	71.9	11.2	3.6	14.8	19.3	81.3	42.7	3.2	45.9	12.9
C	87.3	200.6	2.9	203.5	5.2	95.1	88.3	2.4	90.7	5.2
D	74.7	0.8	9.0	9.8	23.7	88.5	1.7	10.0	11.7	25.5
E	72.0	5.7	14.8	20.5	16.4	86.9	0.2	15.7	15.9	21.9

SNR = signal-to-noise ratio.

experiments to maintain a consistent log-linear transfer function and thus to minimize effects of the display on the outcome of the experiments. Fifth, pooling the results from the several lesion locations had an averaging effect that obscured the differences in detection and estimation SNRs that can arise due to location (39). Finally, statistically significant differences are not necessarily equivalent to clinically significant differences.

CONCLUSION

For both hot and cold tumors in the liver, primary images produced detection and quantitation SNRs that were superior to the other treatments. The greatest improvement was seen in the cases where the scatter fraction was high. Scatter subtraction with the DPW method produced an increase in quantitative accuracy that was greater than the increase in detection accuracy. The NPW observer model was fairly effective at predicting the trends of the average human

observer performance; the ROI estimator was less effective in predicting detection accuracy. The results suggest that to achieve improved accuracy for both detection and quantitation either scatter correction algorithms must produce results that approach the ideal conditions of primary images or detectors must have improved energy resolution for scatter rejection.

ACKNOWLEDGMENTS
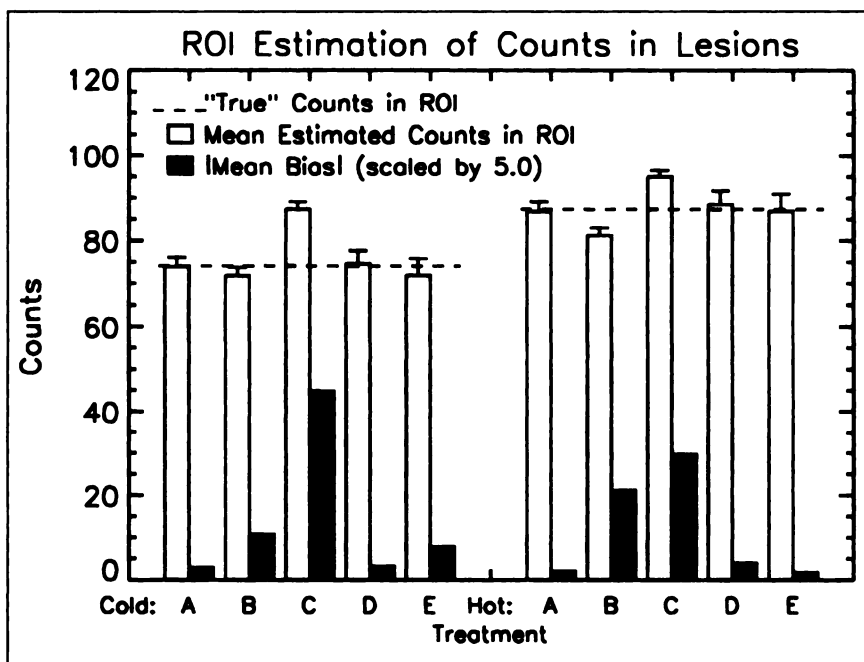
FIGURE 5. Region of interest (ROI) estimator activity estimates are plotted (light bars) with variance indicated by error bars (+1 SD). Absolute value of bias (dark bars) is also shown. Note different vertical scales. Dashed lines indicate "true" activity in ROI, as calculated from noise-free primary image corrected for attenuation. Results for both cold and hot lesions are shown for each treatment.
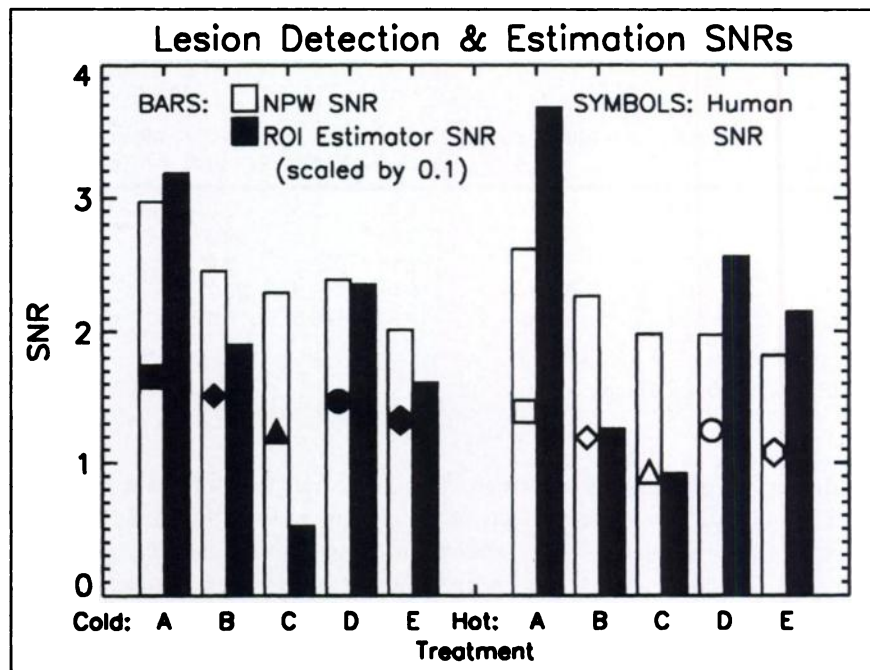
FIGURE 6. For both cold and hot lesions, average human observer signal-to-noise ratio (SNR) (symbols), non-prewhitening (NPW) observer SNR (light gray bars) and region of interest (ROI) estimator SNR (dark gray bars) are shown for each treatment. The symbols used to show human SNR correspond to those in correlation plot in Figure 4.

REFERENCES

1. Jaszczak RJ, Floyd CE, Coleman RE. Scatter compensation techniques for SPECT. *IEEE Trans Nucl Sci.* 1985;32:786–793.
2. Gilardi MC, Bettinardi V, Todd-Pokropek A, Milanesi L, Fazio F. Assessment and comparison of three scatter correction techniques in single photon emission computed tomography. *J Nucl Med.* 1988;29:1971–1979.
3. Ljungberg M, King MA, Hademenos GJ, Strand SE. Comparison of four scatter correction methods using Monte Carlo simulated source distributions. *J Nucl Med.* 1994;35:143–151.
4. Buvat I, Rodriguez-Villafuerte M, Todd-Pokropek A, Benali H, DiPaola R. Comparative assessment of nine scatter correction methods based on spectral analysis using Monte Carlo simulations. *J Nucl Med.* 1995;36:1476–1488.
5. Haynor DR, Kaplan MS, Miyaoka RS, Lewellen T. Multi-window scatter correction techniques in single-photon imaging. *Med Phys.* 1995;22:2015–2024.
6. Jaszczak RJ, Greer KL, Floyd CE, Harris CC, Coleman RE. Improved SPECT quantification using compensation for scattered photons. *J Nucl Med.* 1984;25:893–900.
7. Ogawa K, Harata Y, Ichihara T, Kubo A, Hashimoto S. A practical method for position-dependent Compton-scatter correction in single photon emission CT. *IEEE Trans Med Imaging.* 1991;10:408–412.
8. Moore SC, Kijewksi MF, Müller SP, McIlvain TR, Zimmerman RE. Evaluation of SPECT scatter compensation methods by their effects on parameter estimation tasks [abstract]. *J Nucl Med.* 1996;37:29P.
9. Axelsson B, Msaki P, Israelsson A. Subtraction of Compton scattered photons in single-photon emission computed tomography. *J Nucl Med.* 1984;25:490–494.
10. Logan KW, McFarland WD. Single photon scatter compensation by photopeak energy distribution analysis. *IEEE Trans Med Imaging.* 1992;11:161–164.
11. Kojima A, Tsiju A, Takaki Y, et al. Correction of scattered photons in Tc-99m imaging by means of a photopeak dual-energy window acquisition. *Ann Nucl Med.* 1992;6:153–158.
12. Pretorius PH, van Rensburg AJ, van Aswegen A, Lötter MG, Serfontein DE, Herbst CP. The channel ratio method of scatter correction for radionuclide image quantitation. *J Nucl Med.* 1993;34:330–335.
13. de Vries DJ, King MA. Window selection for dual photopeak window scatter correction in Tc-99m imaging. *IEEE Trans Nucl Sci.* 1994;41:2771–2778.
14. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol.* 1979;14:109–121.
15. Metz CE. ROC methodology in radiological imaging. *Invest Radiol.* 1986;21:720–733.
16. Rolland JP, Barrett HH, Seeley GW. Ideal versus human observer for long-tailed point spread functions: does deconvolution help? *Phys Med Biol.* 1991;36:1091–1109.

17. Staff RT, Gemmel HG, Sharp PF. Assessment of energy-weighted acquisition in SPECT using ROC analysis. *J Nucl Med.* 1995;36:2352–2355.
18. Hamill JJ, DeVito RP. Scatter reduction with energy-weighted acquisition. *IEEE Trans Nucl Sci.* 1989;36:1334–1339.
19. de Vries DJ, King MA, Soares EJ, Tsui BMW, Metz CE. Evaluation of the effect of scatter correction on lesion detection in hepatic SPECT imaging. *IEEE Trans Nucl Sci.* 1997;44:1733–1740.
20. Tsui BMW, Metz CE, Atkins FB, Starr SJ, Beck RN. A comparison of optimum detector spatial resolution in nuclear imaging based on statistical theory and on observer performance. *Phys Med Biol.* 1978;23:654–676, 1203–1205.
21. Wagner RW, Brown DG. Unified SNR analysis of medical imaging systems. *Phys Med Biol.* 1985;30:489–518.
22. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci USA.* 1993;90:9758–9765.
23. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. *J Opt Soc Am A.* 1990;7:1266–1278.
24. Barrett HH, Gooley T, Girodias K, Rolland J, White T, Yao J. Linear discriminants and image quality. In: Colchester ACF, Hawkes DJ, eds. *Information Processing in Medical Imaging.* Berlin, Germany: Springer-Verlag; 1991:458–473.
25. Judy PF, Swensson RG, Szulc M. Lesion detection and signal-to-noise ratio in CT images. *Med Phys.* 1981;8:13–23.
26. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Am A.* 1978;12:2447–2457.
27. Ljungberg M, Strand SE. A Monte Carlo program for the simulation of scintillation camera characteristics. *Comput Meth Prog Biomed.* 1989;29:257–272.
28. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerized three-dimensional segmented human anatomy. *Med Phys.* 1994;21:299–302.
29. Hnatowich DJ, Mardirossian G, Ruscowski M, et al. Pharmacokinetics of the FO23C5 anti-CEA antibody fragment labeled with Tc-99m and In-111: a comparison in patients. *Nucl Med Comm.* 1993;14:52–63.
30. Kessler RM, James R, Ellis J, Eden M. Analysis of emission tomographic scan data: limitations imposed by resolution and background. *J Comput Assist Tomogr.* 1984;8:514–522.
31. Metz CE, Wang PL, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F, ed. *Information Processing in Medical Imaging.* The Hague, The Netherlands: Nijhoff; 1984:432–445.
32. Chang LT. A method for attenuation correction in radionuclide computed tomography. *IEEE Trans Nucl Sci.* 1978;25:638–642.
33. King MA, Glick SJ, Penney BC. Activity quantitation in SPECT: a comparison of three attenuation correction methods in combination with pre-reconstruction restoration filtering. *IEEE Trans Nucl Sci.* 1991;38:775–760.

34. Tsui BMW, Terry JA, Gullberg GT. Evaluation of cardiac cone-beam single photon emission computed tomography using observer performance experiments and receiver operating characteristic analysis. *Invest Radiol.* 1993;28:1101–1112.

35. Rockette HE, Gur D, Metz CE. The use of continuous and discrete confidence judgments in receiver operating characteristic studies of diagnostic imaging techniques. *Invest Radiol.* 1992;27:169–172.

36. Nawfel RD, Chan KH, Wagenaar D, Judy PF. Evaluation of video gray-scale display. *Med Phys.* 1992;19:561–567.

37. Pollard JH. *A Handbook of Numerical and Statistical Techniques.* Cambridge, UK: Cambridge University Press; 1979:184–188, 191–192, 226–228.

38. Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am.* 1997;14:2420–2442.

39. de Vries DJ. *Development and Evaluation of Scatter Subtraction for SPECT Imaging* [doctoral thesis]. Worcester, MA: Worcester Polytechnic Institute; 1997.