
A Lesion Detection Observer Study Comparing 2-Dimensional Versus Fully 3-Dimensional Whole-Body PET Imaging Protocols

Carole Lartizien, PhD¹; Paul E. Kinahan, PhD²; and Claude Comtat, PhD³

¹Centre d'Exploration et de Recherche Médicales par Émission de Positons, Hôpital Neurologique, University Claude Bernard, Lyon, France; ²Department of Radiology, University of Washington, Seattle, Washington; and ³Department of Medical Research, Service Hospitalier Frédéric Joliot, Commissariat à l'Énergie Atomique, Orsay, France

We compared the impact of 2-dimensional (2D) and fully 3-dimensional (3D) acquisition modes on the performance of human observers in detecting and localizing tumors in whole-body ¹⁸F-FDG images. **Methods:** We selected protocols based on noise equivalent count (NEC) rates derived from a series of 2D and fully 3D whole-body patient and phantom acquisitions on a dual-mode PET scanner. The fully 3D peak NEC value for a standard 70-kg patient was achieved for an injected dose of approximately 444 MBq (12 mCi) assuming a 90-min delay before acquisition, whereas the 2D peak value was never reached. The protocols were therefore set to those corresponding to a 444-MBq injected dose in fully 3D and 2D and a 740-MBq (20 mCi) injected dose in 2D that was considered as the maximum allowable dose. We used a non-Monte Carlo simulator to generate multiple realizations of whole-body PET data based on the geometry of the mathematic cardiac torso phantom (MCAT) with accurate noise properties. Two-dimensional and fully 3D acquisition times were set to 5 min per bed position. Spherical 1-cm-diameter lesions (targets) with random locations and contrasts were distributed in different organs. The simulated 2D datasets were reconstructed using attenuation-weighted ordered-subsets expectation maximization ((AW)OSEM) and the fully 3D datasets were reconstructed with FORE+(AW)OSEM (FORE = Fourier rebinning). Five human observers located and ranked the targets using a volumetric display of the whole-body PET data to replicate the clinical practice. An alternate free-response operating characteristic (AFROC) analysis of the human observer reports was performed for each protocol and each organ separately. **Results:** The 2D protocol corresponding to 740-MBq injected dose allowed the overall best detection performance. It was followed by the fully 3D acquisition at the peak fully 3D NEC rate from a 444-MBq injected dose. A 2D acquisition corresponding to a 444-MBq injected dose was ranked last. Differences in detection performance were organ specific. **Conclusion:** This study showed that, for this patient size and scanner type, the fully 3D acquisition mode allowed better or equivalent detection performance than the 2D mode for an injected dose corresponding to the peak fully 3D

NEC rate. The 2D acquisition protocol combined with a higher injected dose resulted in the highest detectabilities.

Key Words: detection performance; model observer; whole-body PET imaging

J Nucl Med 2004; 45:714–723

Multibed ¹⁸F-FDG PET whole-body imaging (1) is increasingly used to stage cancer and metastases in many regions of the body. Whole-body scans, however, typically have short acquisition times for each bed position, resulting in images with high levels of statistical noise.

One method of reducing statistical noise is the use of fully 3-dimensional (3D) imaging to improve sensitivity (2,3). Experimental studies have predicted higher noise equivalent count (NEC) rates in fully 3D as compared with standard 2-dimensional (2D) imaging at low activity levels (4,5). The use of fully 3D mode PET has demonstrated significant advantages for brain imaging compared with 2D mode (6). The relative advantage of fully 3D versus 2D mode for whole-body imaging, however, is less clear and is currently the focus of considerable debate as the number of scanners that only operate in fully 3D mode is increasing with the advent of dual-modality PET/CT scanners (7). Raylman et al. (8) performed 2D and fully 3D whole-body PET 10-min acquisitions of an anthropomorphic phantom with inserted spheric lesions at the same injected activity corresponding to 250 MBq ¹⁸F-FDG at scan start. They found similar visual quality and no statistically significant difference of signal-to-noise ratios measured over the spheres. In a recent study, Lodge et al. (9) performed interleaved 7-min 2D and 6-min fully 3D acquisitions on 10 oncology patients 60 min after the injection of approximately 200 MBq ¹⁸F-FDG. They found similar image noise in 2D and fully 3D images reconstructed with the ordered-subsets expectation maximization (OSEM) algorithm and a 15% higher contrast in 2D acquisitions. More recently, numeric and human observer studies have compared detection performance for 2D and fully 3D PET using clinical

Received Jun. 30, 2003; revision accepted Dec. 16, 2004.

For correspondence or reprints contact: Paul E. Kinahan, PhD, University of Washington Medical Center, Box 356004, 1959 N.E. Pacific St., Seattle, WA 98195-6004.

E-mail: kinahan@u.washington.edu

and phantom data (10–12). These studies came to different conclusions that may result from differences in the choice of the acquisition parameters (injected dose, scan duration) and the patient or phantom habitus. There are several possible reasons for the contradiction between these results and the NEC-based comparison of the acquisition modes at low injected dose: The processing of fully 3D datasets is more complicated and there may be small artifacts in the image resulting from imperfect corrections for attenuation, detector efficiencies, and random and scattered coincidences.

Most importantly, it is difficult to perform a fair comparison between 2D and fully 3D acquisition modes with patient or phantom studies, since the different modes are typically collected in an interleaved manner, as done, for example, by Lodge et al. (9) and El Fakhri et al. (11). With this type of procedure the same activity is in the patient or phantom for both acquisition modes, thus precluding optimal activity levels for one of the two modes. One approach that avoids this problem is the use of simulation studies with accurate noise characteristics, if appropriate count levels for optimal activity levels for both 2D and fully 3D acquisition modes can be determined.

In a previous study, we performed extensive 2D and fully 3D whole-body acquisitions of an extended anthropomorphic phantom (5). The NEC rate (13) was measured as a function of the activity concentration in the phantom for different bed positions centered on the head, the thorax, and the abdomen and was correlated to single-photon and coincident count rates measured with a series of whole-body patient scans. This study demonstrated that the NEC rates varied significantly with position in the body and with patient habitus, which could be accounted for by using the body mass index (BMI).

In this work, we extend our previous study by comparing the impact of the acquisition mode (2D vs. fully 3D) on the performance of human observers in detecting and localizing tumors in a standard 70-kg, 170-cm-tall patient. The 2D and fully 3D acquisition protocols were both optimized based on the previously determined NEC rates (5). This avoids the problem described above of suboptimal activity levels with simultaneously acquired 2D and fully 3D data. A non-Monte Carlo simulator was used to rapidly generate PET sinogram data with all corrections applied and with accurate noise properties, thus avoiding the confounding effect of residual bias (14). The simulation studies were based on an extended version of the volumetric mathematic cardiac torso phantom (MCAT) (15) with added spherical lesions. Data acquired in 2D mode were reconstructed using the attenuation-weighted OSEM ((AW)OSEM) algorithm, whereas fully 3D data were reconstructed with the FORE+(AW)OSEM algorithm (FORE = Fourier rebinning) (16), which has been shown to improve detection performance as compared with OSEM when the effect of attenuation is ignored (17). To compare the acquisition modes, we performed human observer studies using a volumetric display of PET images based on clinical software

(18,19). The task was defined as the detection, localization, and ranking of multiple targets per image volume. The choice of using a volumetric display mode instead of the single transverse image plane traditionally used for receiver operating characteristic (ROC) studies was motivated by results from Wells et al. (20) showing differences in the ability to localize lesions in thoracic SPECT scans when using either a single 2D or multiple contiguous 2D image display. The design of this study was not compatible with a standard ROC analysis as we used multiple targets per volume and did not include any volumes without targets. Multiple-target studies, however, can be readily analyzed using the alternate free-response operating characteristic (AFROC) method (21,22). The procedures used for simulation, reconstruction, and analysis procedures are summarized below and are described in more detail in a previous study comparing the impact of reconstruction algorithm on lesion detectability (17).

MATERIALS AND METHODS

For the study reported here, we used simulated 2D and fully 3D whole-body PET data replicating 3 acquisition protocols of the MCAT phantom for different injected doses. Spherical 1-cm-diameter lesions (targets) were randomly located in the phantom subject to the constraint of maintaining a minimum distance of 1 cm from another target sphere or organ boundary. Contrast levels were randomly selected from a predetermined range of values that sampled the range of human detectability from 10% to 90%. Human observers were asked to locate targets and for each location rate the likelihood of there being an actual target. This procedure used a modified version of the display software used for routine clinical studies. Results of the detection and localization studies were analyzed with an AFROC method and reduced to a signal-to-noise (SNR) detectability index as a function of acquisition mode, general location, and target contrast level.

Selection of 2D and Fully 3D Acquisition Conditions

To provide a fair comparison between 2D and fully 3D whole-body imaging, it is necessary to define the optimal protocol for each mode. To this end, we performed 2D and fully 3D whole-body acquisitions of an extended anthropomorphic phantom (Radiologic Support Device, Inc.) on a Siemens/CTI ECAT HR+ scanner (5). The NEC rate was measured as a function of the activity concentration in the phantom for different bed positions centered on the head, the thorax, and the abdomen and using accurate 2D and fully 3D estimations of the scatter fractions at each bed position. The use of the anthropomorphic phantom replicated realistic FDG imaging conditions of different organs of the upper torso and abdomen of a standard 70-kg, 170-cm-tall patient (23) and accounted for the contamination from activity sources outside the field of view. The acquisition parameters were based on standard clinical protocols. The 2D and fully 3D data were acquired with an energy window between 350 and 650 keV. The maximum ring difference was set to 22 in fully 3D and 7 in 2D, which is equivalent to maximum acceptance angles of 12.5° and 4°, respectively. The lines of responses (LORs) were compressed axially into groups of 4 or 5 LORs in fully 3D mode and 7 or 8 LORs in 2D mode. The count rates obtained in the phantom were compared with clinical whole-body data acquired on an HR+

scanner in order to validate the consistency of using such a model of the human thorax. Figure 1 shows the NEC curves for bed positions centered on the thorax and the abdomen. These curves show that, for the ECAT HR+ scanner, the peak value in fully 3D for the abdomen was achieved for an injected dose of 444 MBq (12 mCi) when followed by a 90-min uptake period, whereas the 2D peak value was never reached. Based on these curves, we considered 3 acquisition protocols:

- 3D: A fully 3D acquisition protocol at the peak NEC rate that corresponded to an injected dose of 444 MBq;

- 2Dmax: A 2D protocol corresponding to a 740-MBq (20 mCi) injected dose, which was considered as the maximum allowable dose;
- 2D: A 2D acquisition protocol at the injected dose of the fully 3D peak NEC value.

Table 1 reports the mean values of the true, random, and scatter coincidences for the 3 simulated 5-min emission scan protocols for an acquisition centered on the abdomen. The scatter fractions in the torso were estimated to 20% and 55%, respectively, in the 2D and fully 3D acquisition modes.

Data Simulation

We used the analytic simulation method (ASIM) (14) to allow for the generation of multiple noisy realizations of whole-body sinogram datasets with statistically accurate noise properties. The whole-body simulator accounts for effects that are important in whole-body PET imaging: attenuation, random, and scattered coincidence arising from the activity inside and outside the field of view; detector efficiency variations; activity decay between bed positions; system dead time; and noise arising from the transmission scan. ASIM first calculates the analytic projections of the activity distribution based on a user-specified scanner geometry and adds noise to these projections accounting for the detection efficiency estimated from a real normalization scan. The transaxial profiles of the scatter and random distribution are based on measured scanner data and a 1-dimensional Monte Carlo simulation is used to estimate the scatter and random amplitude as a function of the axial position. Finally, the raw data are corrected for attenuation, random, and scatter coincidences, etc. with the same techniques used in clinical practice, assuming that the corrections, although they can be noisy, are accurate. This technique was validated using experimental data (14).

We generated data that reproduce the FDG distribution in the torso, with a geometry based on the volumetric MCAT phantom (15) with the addition of a head, arms, and a bladder. The dimensions of this phantom correspond to a standard 70-kg, 170-cm-tall patient as defined in MIRD Pamphlet No. 5 (24) and match the size of the anthropomorphic phantom used to measure the NEC rate. Due to the significantly different scatter distributions in 2D and fully 3D imaging, the 2D and 3D scatter profiles used in the simulation tool were estimated from the scatter correction technique developed by Watson et al. (25). Seven spherical 1-cm-diameter lesions with varying contrasts were inserted at randomly generated locations within the lungs, the liver, and the background soft tissues of the MCAT phantom, respecting a minimal distance of 1 cm from the edge of an organ or another target to avoid confusion. Target contrast was defined as the target-to-background activity concentration ratio—that is (concentration in the target/concentration in the background).

Emission data were generated in the fully 3D and 2D modes of acquisition with the total number of true, scattered, and random coincidences based on those given in Table 1. The acquisition parameters (energy window, maximum ring difference) were similar to the values used for the anthropomorphic phantom experiment.

Data Reconstruction

Simulated 2D mode sinograms were reconstructed using (AW)OSEM, whereas fully 3D data were reconstructed with FORE+(AW)OSEM (16). We used 16 subsets and 4 iterations for FORE+(AW)OSEM and (AW)OSEM algorithms followed by a

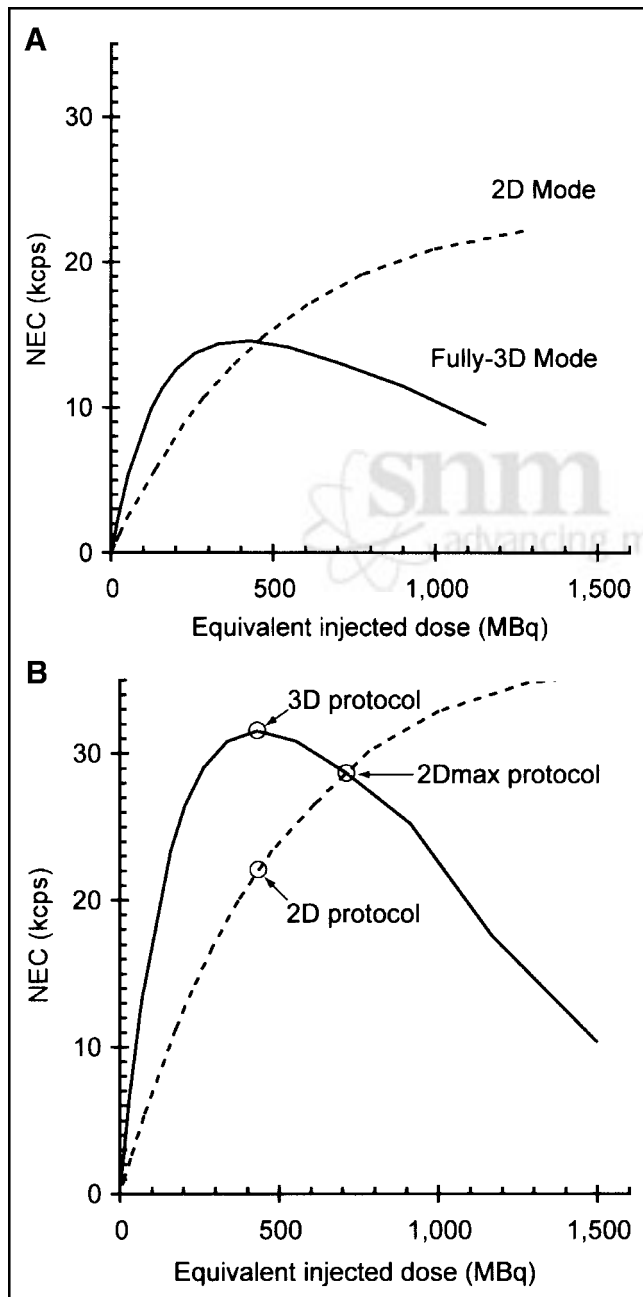


FIGURE 1. 2D and fully 3D NEC curves on EXACT HR+ scanner as function of injected dose to 70-kg, 170-cm-tall patient. Curves were obtained for a bed position centered on thorax (A) and abdomen (B).

TABLE 1
Average Number of True, Scatter, and Random Coincidences (in Million Counts) for 3 Acquisition Conditions

Coincidence type	2D	2Dmax	3D
True	10.0	16.3	37
Scatter	2.5	4.1	45
Delayed	4.5	12.0	90

postreconstruction smoothing with a 3D gaussian filter to control the contrast-to-noise ratio (CNR). The full width at half maximum (FWHM) of the gaussian filters for the 2D and fully 3D acquisitions were determined as those that maximized the CNR for 1-cm-diameter spheres. The CNR is closely related to the nonprewhitening matched filter (26) and was calculated as:

$$CNR = \frac{\langle(T - B)/B\rangle}{\sqrt{\sigma^2(T) + \sigma^2(B)}}, \quad \text{Eq. 1}$$

where T and B are the measured activity concentrations for the target and background regions in the reconstructed image volumes, $\langle \rangle$ represents the ensemble average, and $\sigma^2(T)$ and $\sigma^2(B)$ are the variances of these activities estimated across multiple realizations. Figure 2 plots the variation of the CNR as a function of the FWHM of the gaussian filter for spheres with a contrast of 4:1 and 8:1. The CNRs for the low- and high-contrast spheres were maximal with a filter of 10-mm FWHM and 12-mm FWHM, respectively, although the variations were not sharp. As a consequence, the FWHM was set to 10 mm allowing for a consensus of subjective visual preferences.

Calibration

An important parameter is the distance threshold used to determine when a target has been correctly reported. The distance threshold needs to be large enough to allow for observer localization error, but small enough to minimize the chance identification of a true target. Based on calibration studies, a target was considered as correctly reported when the 3D position reported by the observer was within a 15-mm (3 voxels) distance from the true target location.

When designing an observer detection performance study, it is important to select appropriate values of the internal or true target contrast (as opposed to the apparent contrast in the reconstructed image) to sample the range of detectability. Based on the distance threshold established above, the fraction of targets correctly reported (i.e., the fraction of targets found or “fraction found”) was estimated as a function of true contrast. Table 2 reports the true contrast values needed for fraction found values to be uniformly sampled between 0.1 and 0.9 for each tissue type. These values were used for the final observer performance study.

Observer Detection Performance Study

The dataset used for the human observer studies consisted of 2D and fully 3D acquisitions of the MCAT phantom over a 55-cm axial extent corresponding to 4 bed positions. Fifty noise-free scans of the phantom containing 7 targets each were simulated in both acquisition modes. The targets were randomly distributed in the 3 organs of interest (lungs, liver, and background soft tissue) from a multinomial probability distribution with an average of 2.5 targets each in the lungs and the liver and 2 targets in other soft

tissues throughout the thorax and abdomen. The target contrast was randomly selected from the predetermined set of 5 values based on the preliminary calibration study, thus leading to approximately 25 targets of the same activity ratio for the liver and the lungs (2.5 targets per organ \times 50 volumes/5 contrast levels). Three noisy realizations (corresponding to the 2D, 2Dmax, and 3D protocols) per noise-free scan were generated with noise levels determined from the anthropomorphic phantom experiment, thus resulting in 150 whole-body images (50 noisy scans \times 3 acquisition protocols). The 2D and 3D datasets were reconstructed with the (AW)OSEM and FORE+(AW)OSEM algorithms, using the parameters described above. Interpretation of the simulated whole-body PET images was performed with a “volumetric” display (linked display of the 3 primary views: coronal, sagittal, and transverse) to replicate typical clinical practice.

Five observers participated to the study (2 nuclear medicine physicians and 3 experienced physicists working in PET facilities). Four of the 5 observers read the whole set of 150 whole-body images corresponding to the 3 protocols, whereas the remaining observer only read the image volumes from protocols 2D and 2Dmax. All observers undertook a short training session, using noise-free images to check their reported locations. For both the training and the actual studies, no time limit was imposed and the observers could control color scales and threshold as in clinical practice. They were told the mean number of targets per organ and were asked to report and rate 7 locations per volume on a 5-point ordinal scale of confidence (5 = definite strong target, 4 = medium-strong target, 3 = medium-weak target, 2 = weak target, and 1 = probably not a target). The analysis assigned a default rating of zero to all unreported targets. The reconstructed images of each acquisition protocol were split into subsets of 10 image volumes that were presented to the observers in a random sequence to reduce reading order effects.

Observer Detection Performance Analysis

The choice of using multiple targets per image volume increases the sensitivity of the study (27), thus reducing the number of image

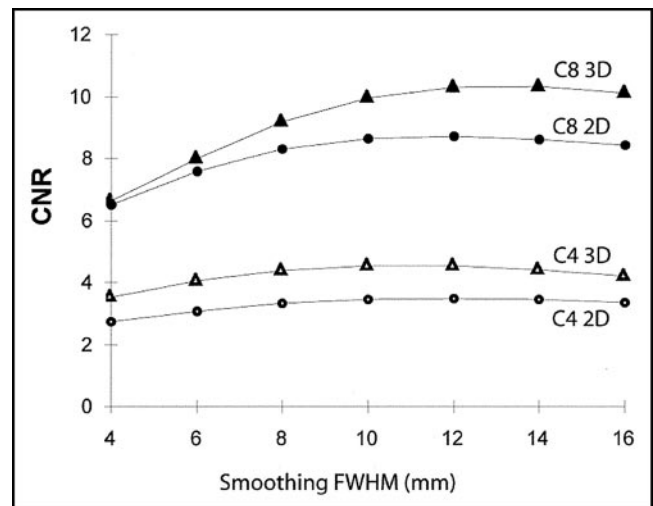


FIGURE 2. Measured CNR in image volumes vs. FWHM of postreconstruction gaussian smoothing filter. C4 3D = fully 3D acquisition with target contrast of 4:1; C4 2D = 2D acquisition with target contrast of 4:1; C8 3D = fully 3D acquisition with target contrast of 8:1; C8 2D = 2D acquisition with target contrast of 8:1.

TABLE 2
True Contrast Values of Targets Inserted
in MCAT Phantom

Fraction of targets found	0.1	0.3	0.5	0.7	0.9
Lungs	5.5	6.5	7	7.5	9
Liver	2.5	3	3.25	3.75	4.75
Soft tissue	6.5	7	8	9	10.5

volumes that must be read by observers in order to achieve a statistically significant result, which is an important consideration when mimicking the standard clinical procedure or reviewing image volumes with linked views of the 3 principal views: coronal, sagittal, and transverse (19). This design, however, was not compatible with a standard ROC analysis but could be readily analyzed using the AFROC method (21,22). This method was proposed as a technique to evaluate observer detection performance in more complicated and realistic tasks than single-target ROC studies by measuring localization accuracy with multiple targets per image. The AFROC curve plots the probability of a correct target report at each rating cutoff, as a function of the probability that the observer will also report one or more false targets at the same rating cutoff. The area below an AFROC curve may be interpreted as the probability that a specified target would be either (a) rated higher than the most suspicious nontarget location (21) or (b) correctly localized (by first choice) on an image containing only that single target (22).

An AFROC analysis was performed for each observer and each acquisition protocol (2D, 2Dmax, 3D). A set of AFROC curves was derived for all targets pooled across contrasts and a second set of AFROC curves was generated for each contrast separately. The parameters of the AFROC curves with data pooled across contrast were estimated by the CORROC program developed by Metz et al. (28) for pair-wise comparison of correlated ratings from 2 conditions that presented the same case. The CORROC program also provided the area A_L under the AFROC curve and statistical z -score tests for intermodal differences in areas below pairs of fitted curves. Rating correlations induced by the use of the same set of noise-free MCAT sinograms for the 3 protocols were accounted for. A default rating of zero was assigned to any unreported target. We note that since the area under the AFROC curve corresponds to the fraction of targets rated above the most-suspicious nontarget

(22), it is possible for the area A_L to be <0.5 . This is unlike a standard ROC analysis, where an area under the ROC curve of $A_Z = 0.5$ indicates random guessing.

In addition, the series of AFROC curves for each individual contrast level were fitted based on a multiple alternative procedure that accounts for reports of different target classes (each class corresponding to a different value of the target contrast here) and assuming that the observer uses a common perceptual rating criteria for all target classes (29). The bounded area $0 \leq A_L \leq 1$ below the AFROC curve was then converted into an unbounded detectability index, d_L , through the relation (30):

$$d_L = 2 \operatorname{erf}^{-1}(2A_L - 1), \quad \text{Eq. 2}$$

where $\operatorname{erf}(\)$ denotes the error function.

RESULTS

Coronal and transaxial sections of images reconstructed from one noise-free scan of the MCAT phantom with noise levels corresponding to the 3 acquisition protocols are shown in Figure 3.

Results of the AFROC analysis indicated that the ranking order of the 3 protocols was similar for all observers and that the absolute detection performances were also homogeneous except for 1 observer (observer 3), whose average performances were lower. Figure 4 plots the AFROC curves obtained for 1 observer (observer 1) for the 3 acquisition protocols and for the 3 organ types. Figure 5 plots the AFROC curves obtained by averaging the linear parameters of the fitted AFROC curves for the 4 observers who read the whole set of 150 whole-body images, including the lower performance from observer 3. The fraction of targets correctly reported (fraction found) on these plots was determined from all targets pooled across contrast within an organ. These curves show that the 2D protocol corresponding to a 740-MBq injected dose (2Dmax) allowed the overall best detection performances. It was followed by the fully 3D peak NEC acquisition protocol corresponding to a 444-MBq injected dose and, then, the 2D mode at the same injected activity of the fully 3D peak NEC value. These differences in detection performance were region specific.

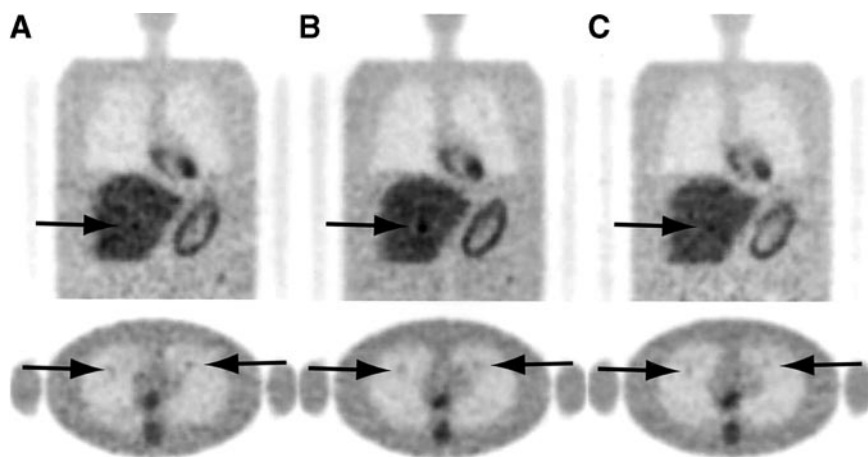


FIGURE 3. Coronal and transaxial sections of typical simulated scans of MCAT phantom with noise levels corresponding to 3 protocols: (A) 2D; (B) 2Dmax; and (C) 3D. Data were reconstructed with FORE+(AW)OSEM. This example shows 3 targets (arrows).

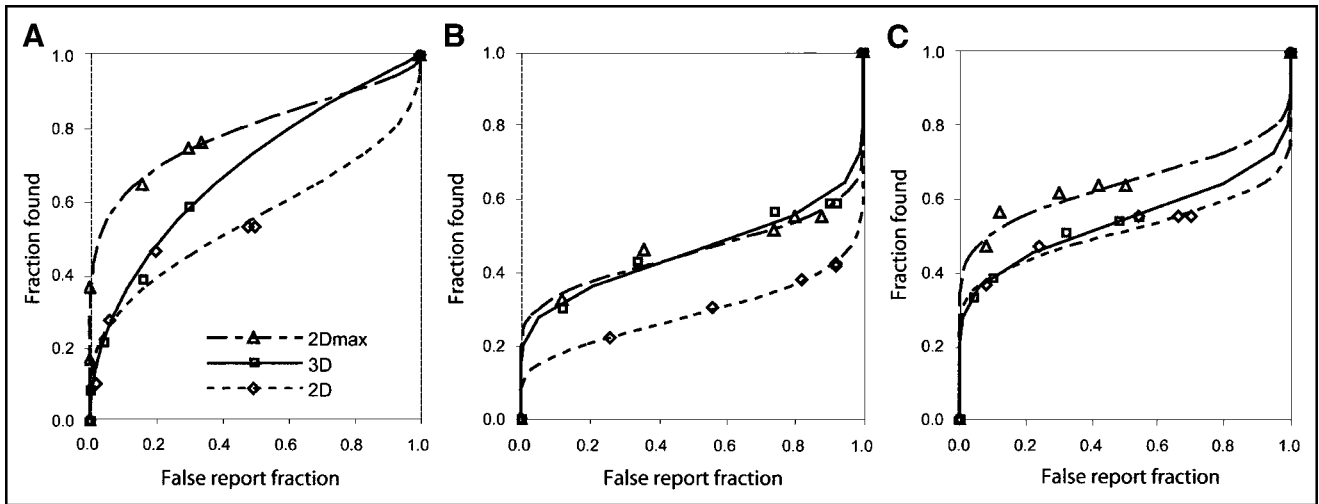


FIGURE 4. AFROC curves obtained for 1 observer for 3 acquisition protocols and for detection of all targets in lungs (A), liver (B), and soft tissues (C). 3D = fully 3D NEC peak acquisition; 2D = 2D acquisition at injected dose corresponding to fully 3D NEC peak value; 2Dmax = 2D acquisition at maximum injected dose of 740 MBq.

The AFROC curves for the 3 protocols show little cross-over, thus indicating that the area below the curve, A_L , is a reasonable figure of merit for comparing the different protocols. Tables 3–5 report the estimated values of A_L for each individual observer, each organ, and each protocol, together with the results from correlated z -score tests of differences in A_L for each pair of protocols. The SEs for individual estimates of A_L were obtained from the CORROC maximum-likelihood fitting procedure and were similar for all observers and all organs with values of approximately 0.045. The numeric results of Tables 3–5 are summarized in Table 6 by averaging the areas under the AFROC curves for all observers and all organs.

The 2Dmax acquisition protocol at 740-MBq injected dose led to an improvement in detection performance as compared with the 2D acquisition protocol at 444-MBq injected dose. This difference was statistically significant

for all observers for targets located in the lungs, for 4 of 5 observers for the liver, and for 3 of 5 observers for targets located in the soft tissues.

Comparison of fully 3D images with 2D images for the same injected dose of 444 MBq indicates that the fully 3D mode yielded equivalent or better detection performances for all observers and for all organ types. This difference was statistically significant for 2 of 4 observers for the liver but failed to achieve statistical significance for targets located in the lungs and the soft tissues.

Finally, improvement in detectability in all organs was found with 2D images corresponding to 740-MBq injected dose as compared with fully 3D images at 444-MBq injected dose. This improvement was not strongly significant for targets located in the liver and soft tissues (1 of 4 observers), but it was statistically significant for 3 of 4 observers in the lungs.

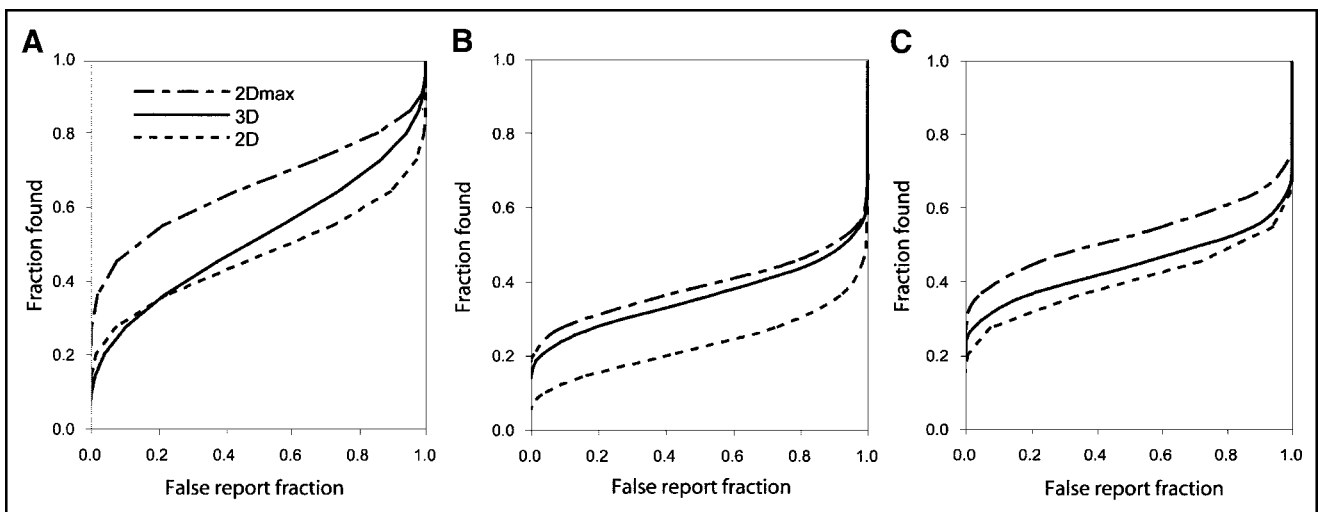


FIGURE 5. AFROC curves averaged over 4 observers for lungs (A), liver (B), and soft tissues (C).

TABLE 3
Areas Under Fitted AFROC Curves for Targets
Located in Lungs

Protocol	Parameter	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
2D	A_L	0.54	0.51	0.29	0.52	0.42
2Dmax	A_L	0.79	0.69	0.45	0.64	0.59
3D	A_L	0.69	0.54	0.30	0.49	
2D vs. 2Dmax	P	<0.001	0.002	0.002	0.030	0.005
2D vs. 3D	P	0.044	0.665	0.832	0.507	
2Dmax vs. 3D	P	0.112	0.005	0.005	0.007	

Obs. = observer; A_L = area below AFROC curve; P = probability value of z-score test for each intermodality comparison.

We also derived the series of AFROC curves for each target contrast using the multiple alternative procedure of Kijewski et al. (29). Figure 6 plots the variations of the unbounded SNR index d_L averaged over 3 of the observers as a function of target contrast for the 3 organ types. Reports from observer 3 produced degenerate results and were not included. This degeneracy originated from a complete separation of the distributions of correct and false reports for this observer. Results from Figure 6 confirm the rank ordering derived from the AFROC curves for all targets pooled across contrast in Figures 4 and 5.

DISCUSSION

For brain imaging it is well established that a fully 3D acquisition mode leads to improved image quality relative to a 2D acquisition mode. In whole-body imaging the relative trade-offs are less clear. A difficulty in comparing between 2D and fully 3D acquisition modes with patient or phantom studies is that the different modes are typically collected in an interleaved manner. With this type of procedure the same activity is in the patient or phantom for

TABLE 4
Areas Under Fitted AFROC Curves for Targets
Located in Liver

Protocol	Parameter	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
2D	A_L	0.29	0.22	0.20	0.21	0.17
2Dmax	A_L	0.45	0.42	0.29	0.38	0.27
3D	A_L	0.45	0.39	0.30	0.28	
2D vs. 2Dmax	P	0.004	<0.001	0.080	<0.001	0.023
2D vs. 3D	P	0.004	<0.001	0.0570	0.117	
2Dmax vs. 3D	P	0.947	0.630	0.856	0.024	

Obs. = observer; A_L = area below AFROC curve; P = probability value of z-score test for each intermodality comparison.

TABLE 5
Areas Under Fitted AFROC Curves for Targets
Located in Soft Tissues

Protocol	Parameter	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
2D	A_L	0.51	0.52	0.23	0.36	0.16
2Dmax	A_L	0.63	0.57	0.26	0.64	0.29
3D	A_L	0.54	0.50	0.30	0.44	
2D vs. 2Dmax	P	0.020	0.273	0.580	<0.001	0.004
2D vs. 3D	P	0.599	0.779	0.137	0.212	
2Dmax vs. 3D	P	0.092	0.131	0.277	0.004	

Obs. = observer; A_L = area below AFROC curve; P = probability value of z-score test for each intermodality comparison.

both acquisition modes, thus precluding optimal activity levels for one of the two modes. One approach that avoids this problem is the use of simulation studies with optimal activity levels for both 2D and fully 3D acquisition modes. Simulation studies, however, are never completely realistic in the replication of bias and resolution effects. Our approach is to assume no bias in the simulations, which corresponds to perfect correction for effects such as random and scattered coincidences, attenuation, detector efficiency variations, and dead time. Considerable effort has been, and continues to be, devoted to reducing the bias for these corrections. Our simulation is based on the assumption that for whole-body oncology imaging, the dominant factor for image quality is noise, which is accurately simulated by our procedure (14).

In a previous study, we performed extensive 2D and fully 3D whole-body acquisitions of an extended anthropomorphic phantom (5). The NEC rate (13) was measured as a function of the activity concentration in the phantom for different bed positions centered on the head, the thorax, and the abdomen and was correlated to single-photon and coincident count rates measured with a series of whole-body patient scans. This study demonstrated that the NEC rates varied significantly with position in the body and with patient habitus, which could be accounted for by using the BMI. As we noted in that study, however, NEC rates are only one factor that affect image quality. The NEC figure of merit does not account for the effects of detector resolution or image noise covariance, both of which can differ between 2D and fully 3D imaging modes. The covariance of image noise is also strongly affected by the choice of image

TABLE 6
Summary of Tables 3–5 of A_L Averaged
Over All Observers and All Regions

Protocol	Lungs	Liver	Soft tissues	Average over all regions
2D	0.46	0.22	0.36	0.34
2Dmax	0.63	0.36	0.48	0.49
3D	0.51	0.36	0.45	0.43

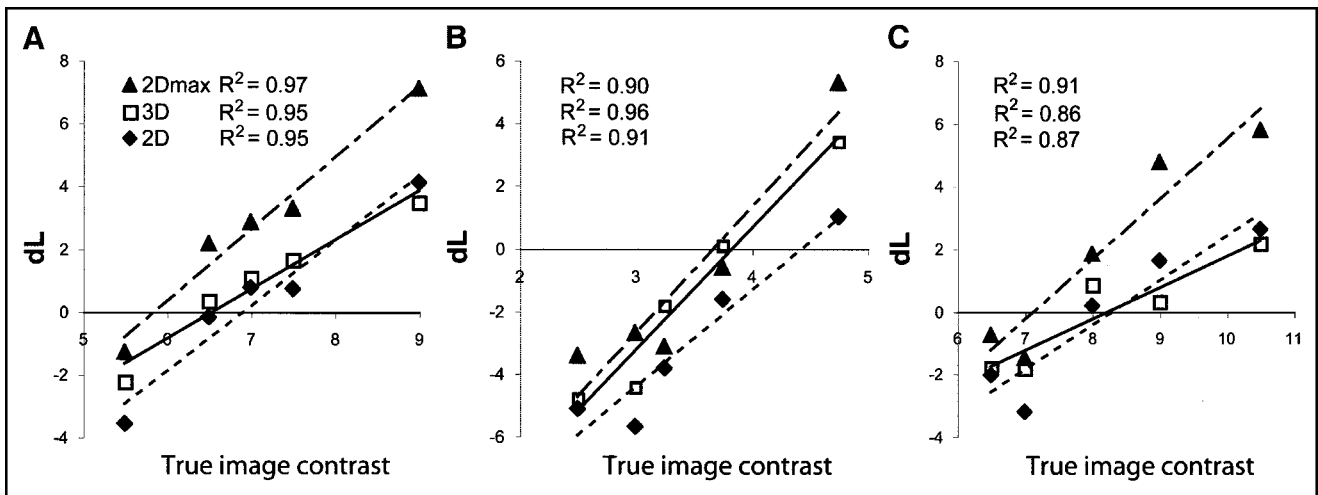


FIGURE 6. Detectability index d_L averaged over 3 observers as function of theoretic target contrast for 3 acquisition protocols and for lungs (A), liver (B), and soft tissues (C).

reconstruction algorithm. Image quality can be more directly assessed using a quantitative task-based measurement such as lesion detectability.

In this study, we compared the impact of 2D versus fully 3D whole-body PET imaging protocols on the performance of human observers in detecting and localizing spherical 1-cm-diameter lesions of different contrasts in a standard 70-kg, 170-cm-tall patient. Three acquisition protocols were selected for this comparison based on the NEC index and evaluated by 5 observers. The results indicate that the best overall human detection performances were achieved for images acquired in 2D mode with count rates corresponding to the maximum allowable (in our study) injected dose of 740 MBq (20 mCi) as summarized in Table 6. This acquisition protocol led to an improvement in detectability as compared with the 2D and fully 3D acquisition protocols with an injected dose of 444 MBq (12 mCi). The 444-MBq dose corresponded to the fully 3D peak NEC rate at acquisition for a standard 70-kg patient. These differences were region specific with statistically significant differences in the lungs and lower differences in the liver between the 2D protocol at maximum dose and the fully 3D peak NEC protocol.

We also found improved detectability for fully 3D images compared with 2D images when both protocols used the same injected dose of 444 MBq (12 mCi). This improvement was statistically significant for targets located in the liver. This result suggests that for a detection task images acquired in fully 3D are superior or equivalent to those acquired in 2D at low injected activity. In this case it is likely that detectability benefits from the gain in sensitivity in fully 3D mode without being too affected by the noise correlations introduced by the processing of the fully 3D data. This improvement in detectability using the fully 3D mode as compared with the 2D mode for the same dose and acquisition time is consistent with results reported by Moore

et al. based on a phantom experiment also using the ECAT HR+ scanner (31). They used the nonprewhitening matched filter (NPWMF) (32) to estimate detectability of spheres embedded in an anthropomorphic phantom as a function of the acquisition mode and counting rate characteristics. The NPWMF is a numerical observer that has been shown to be related to human observer performance in simple detection tasks of signal known exactly in homogeneous background (33). Moore et al. found that fully 3D images corrected for attenuation and reconstructed with the FORE+FBP (FBP = filtered backprojection) algorithm provided better detectability than 2D images corrected for attenuation and reconstructed with FBP for the same injected dose and scan duration. In a more recent study, El Fakhri et al. used another numerical model, the channelized Hotelling observer (CHO), to perform a similar evaluation of the acquisition mode based on clinical data with superimposed spherical lesions (11). Thirty-six patients were scanned on an ECAT HR+ scanner 3 h after injection of 740 MBq ^{18}F -FDG. This corresponded to a mean activity of 240 MBq at scan start time, which is equivalent to our fully 3D peak NEC protocol (250 MBq at scan start time). This study showed that the mean CHO detectability index was significantly better in fully 3D for patients with a BMI of <33 compared with the 2D mode, but the 2D mode improved detection in large patients (BMI, >33) and the difference was not as significant as for average size patients. Similar results were also recently presented by Kadrmas et al., who performed a human observer study using an anthropomorphic phantom with and without added chest overlay to simulate different patient habitus (10). These results indicate that our conclusions, though valid for medium-sized patients, should be further investigated for larger patients. Conclusions of our work differ from those of the human ROC and localization ROC observer studies presented by Farquhar et al. (12) based on 2D and fully 3D

clinical scans of healthy patients with added simulated lung tumors. Twenty-five volunteers were scanned on an ECAT HR+ scanner. The mean dose in patients at the beginning of the 2D scan was about 245 MBq, which is equivalent to our fully 3D NEC peak protocol at the start of acquisition. The dose at the beginning of the fully 3D scan was about 33% decreased relative to the 2D scan and the fully 3D scan duration (5 min) was also reduced by half of the 2D scan duration (9 min). Results show a decrease in detection performance with fully 3D acquisitions as compared with 2D acquisitions based on data that were not corrected for attenuation and were reconstructed with FBP in 2D and FORE+FBP in fully 3D. This comparison of 2D and fully 3D imaging is thus different from the design of Moore et al. and our design since we chose the same scan duration for both modes.

We showed that the 2D protocol for the highest injected dose of 740 MBq was ranked higher for detection performance than the 2D protocol corresponding to an injected dose of 444 MBq. This result is also in good agreement with results presented by Moore et al. based on the NPWMF (31).

The NEC and AFROC curves for the lungs (Figs. 1A and 5A) and the liver (Figs. 1B and 5B) indicate that the rank order of detection performance tracked the rank order NEC rates. The studies reported here, however, did not allow for variations in background biodistributions; further investigations would be needed to determine if the correlation between rank ordering of detection performance and NEC rates generalizes in practice. Even if NEC rates are equal between 2D and fully 3D acquisition modes, the lesion detectability would not necessarily be the same for both modes. The noise levels and noise correlations as a function of axial position differ between images reconstructed with the 2D (AW)OSEM and the fully 3D FORE+(AW)OSEM algorithms. The noise variations are due to 2 factors: differences in the axial sensitivity profiles between the 2D and fully 3D datasets (34,35) and the different processing steps applied to the 2D and fully 3D sinograms during reconstruction (16). In general, the results will change for different image reconstruction algorithms as the noise correlations are strongly affected by the choice of algorithm (36).

A limitation of our study is that it only applies to a 70-kg, 170-cm-tall patient imaged on a Siemens/CTI ECAT HR+ PET scanner. Changing the patient size or the PET scanner will potentially change the lesion detectability results. In addition, the peak NEC rates for 2D and fully 3D acquisition modes do not necessarily correspond to peak lesion detectability for either mode. In other words, a suboptimal NEC rate could potentially have a higher lesion detectability. We note that, with all other factors being equal for the same scanner, the detectability with the different protocols is rank-order consistent with the corresponding NEC rates. It also seems likely that for different scanners with equal NEC rates and energy resolution, improved detectability would be rank-order consistent with spatial resolution if equally accurate data processing methods were used. With

scanners of different spatial and energy resolutions and NEC rates, however, the relative impact of these factors would need to be reevaluated. Finally, we note that lesion detectability for the system considered here can potentially be increased with injected activity levels of >740 MBq (20 mCi).

CONCLUSION

In this study, we performed an AFROC analysis to evaluate the impact of the acquisition mode (2D vs. fully 3D) on human observer detection performances. Three acquisition protocols were selected to provide a fair comparison between the acquisition modes. Results showed that the fully 3D acquisition mode allowed better or equivalent detection performance than the 2D mode for a same injected dose typical of the clinical practice (about 440 MBq) in a standard patient. The 2D acquisition protocol combined with higher injected doses (about 740 MBq) resulted in higher detectability than those achieved with the fully 3D acquisition mode for approximately half the injected dose. Changing the patient size or the PET scanner model will potentially change the lesion detectability results of this study.

ACKNOWLEDGMENTS

We thank Dr. Christian Michel and Dr. Michel Defrise for providing the reconstruction software; Dr. Ben Tsui and Dr. Karen LaCroix for help with the MCAT phantom; and Dr. Michael A. King, Dr. Howard Gifford, and Dr. Harrison H. Barrett for helpful discussions. We gratefully acknowledge the extensive contributions to this project by the late Dr. Richard Swensson. We also thank Dr. Matthew Heller and Dr. Subhash Chander for their assistance with the observer studies. This work was supported by National Cancer Institute grant CA-74135. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

REFERENCES

1. Dahlbom M, Hoffman EJ, Hoh CK, et al. Whole-body positron emission tomography. Part I. Methods and performance characteristics. *J Nucl Med.* 1992;33:1191-1199.
2. Cherry SR, Dahlborn M, Hoffman EJ. High sensitivity, total body PET scanning, using 3D data acquisition and reconstruction. *IEEE Trans Nucl Sci.* 1992;39:1088-1092.
3. Cutler PD, Xu M. Strategies to improve 3D whole-body PET image reconstruction. *Phys Med Biol.* 1996;41:1453-1467.
4. Badawi RD, Marsden PK, Cronin BF, et al. Optimization of noise-equivalent count rates in 3D PET. *Phys Med Biol.* 1996;41:1755-1776.
5. Lartzien C, Comtat C, Kinahan PE, et al. Optimization of injected dose based on noise equivalent count rates for 2- and 3-dimensional whole-body PET. *J Nucl Med.* 2002;43:1268-1278.
6. Cherry SR, Woods RP, Hoffman EJ, et al. Improved detection of focal cerebral blood flow changes using three-dimensional positron emission tomography. *J Cereb Blood Flow Metab.* 1993;13:630-638.
7. Beyer T, Townsend DW, Brun T, et al. A combined PET/CT scanner for clinical oncology. *J Nucl Med.* 2000;41:1369-1379.
8. Raylman RR, Kison PV, Wahl RL. Capabilities of two- and three-dimensional FDG-PET for detecting small lesions and lymph nodes in the upper torso: a dynamic phantom study. *Eur J Nucl Med.* 1999;26:39-45.
9. Lodge MA, Lowe J, Lowe G, et al. A comparison of 2D and 3D PET images

- using an interleaved septa-in/septa-out acquisition protocol [abstract]. *J Nucl Med.* 2002;43(suppl):206P.
10. Kadrmas DJ, Christian PE, Wollenweber SD, et al. Comparative evaluation of 2D and 3D lesion detectability on a full-ring BGO PET scanner [abstract]. *J Nucl Med.* 2002;43(suppl):204P.
 11. El Fakhri G, Holdsworth C, Badawi RD, et al. Impact of acquisition geometry and patient habitus on lesion detectability in whole-body FDG-PET: a channelized Hotelling observer study. *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference* [CD-ROM]. Norfolk, VA; 2002.
 12. Farquhar TH, Llacer J, Hoh CK, et al. ROC and localization ROC analyses of lesion detection in whole-body FDG PET: effects of acquisition mode, attenuation correction and reconstruction algorithm. *J Nucl Med.* 1999;40:2043–2052.
 13. Strother SC, Casey ME, Hoffman EJ. Measuring PET scanner sensitivity: relating count rates to image signal-to-noise ratios using noise equivalent counts. *IEEE Trans Nucl Sci.* 1990;37:783–788.
 14. Comtat C, Kinahan PE, Defrise M, et al. Simulating whole-body PET scanning with rapid analytical methods. *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference.* Seattle, WA; 1999;3:1260–1264.
 15. Tsui BMW, Terry JA, Gullberg GT. Evaluation of cardiac cone-beam SPECT using observer performance experiments and ROC analysis. *Invest Radiol.* 1993;28:1101–1112.
 16. Comtat C, Kinahan PE, Defrise M, et al. Fast reconstruction of 3D PET data with accurate statistical modeling. *IEEE Trans Nucl Sci.* 1998;45:1083–1089.
 17. Lartizien C, Kinahan PE, Swensson R, et al. Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging. *J Nucl Med.* 2003;44:276–290.
 18. Lartizien C, Kinahan PE, Comtat C, et al. Performing ROC studies with the same volume visualization software tools used in whole-body 3D PET oncology imaging [abstract]. *J Nucl Med.* 2000;41(suppl):102P.
 19. Lartizien C, Kinahan PE, Comtat C, et al. Evaluating image reconstruction methods for tumor detection performance in whole-body PET oncology imaging. *Proceedings of the Society of Photo-Optical Instrumentation Medical Imaging Conference.* San Diego, CA; 2000;3981:49–59.
 20. Wells RG, King MA, Gifford HC, et al. Single-slice versus multi-slice display for human-observer lesion-detection studies. *IEEE Trans Nucl Sci.* 2000;47:1037–1044.
 21. Chakraborty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology.* 1990;174:873–881.
 22. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys.* 1996;23:1709–1725.
 23. Doshi NK, Basic M, Cherry SR. Evaluation of the detectability of breast cancer lesions using a modified anthropomorphic phantom. *J Nucl Med.* 1998;39:1951–1957.
 24. Snyder WS, Ford MR, Warner GG, et al. Estimates of absorbed fractions for monoenergetic photon sources uniformly distributed in various organs of a heterogeneous phantom. *Nuclear Medicine/MIRD Pamphlet No. 5.* New York, NY: Society of Nuclear Medicine; 1969.
 25. Watson C, Newport D, Casey ME. Evaluation of simulation-based scatter correction for 3D PET cardiac imaging. *IEEE Trans Nucl Sci.* 1997;44:90–97.
 26. Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. *Phys Med Biol.* 1985;30:489–518.
 27. Chakraborty DP. Statistical power in observer-performance studies: comparison of the receiver operating characteristic and free-response methods in tasks involving localization. *Acad Radiol.* 2002;9:147–156.
 28. Metz CE, Wang PL, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. *Proceedings of the Information Processing in Medical Imaging.* The Hague, The Netherlands; 1984:432–445.
 29. Kijewski MF, Swensson RG, Judy PF. Analysis of rating data from multiple-alternative tasks. *J Math Psychol.* 1989;33:428–451.
 30. Burgess A. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Med Phys.* 1995;22:643–655.
 31. Moore SC, El Fakhri G, Badawi R, et al. Relative lesion detectability in 3D vs. 2D dedicated multi-ring PET. *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference.* Lyon, France; 2000;3:18–22.
 32. Barrett HH, Yao J, Rolland JP, et al. Model observers for assessment of image quality. *Proc Natl Acad Sci USA.* 1993;90:9758–9765.
 33. Myers KJ, Barrett HH, Borgstrom MC, et al. Effect of noise correlation on detectability of disk signals in medical imaging. *J Opt Soc Am.* 1985;A2:1752–1759.
 34. Kinahan PE, Karp JS. Figures of merit for comparing reconstruction algorithms with a volume-imaging PET scanner. *Phys Med Biol.* 1994;39:631–638.
 35. Defrise M, Geissbuhler A, Townsend DW. A performance study of 3D reconstruction algorithms for positron emission tomography. *Phys Med Biol.* 1994;39:305–320.
 36. Pajevic S, Daube-Witherspoon ME, Bacharach SL, et al. Noise characteristics of 3-D and 2-D PET images. *IEEE Trans Med Imaging.* 1998;17:9–23.