Performance Standards in Positron Emission Tomography

Joel S. Karp, Margaret E. Daube-Witherspoon, Edward J. Hoffman, Thomas K. Lewellen, Jonathan M. Links, Wai-Hoi Wong, Richard D. Hichwa, Michael E. Casey, James G. Colsher, Richard E. Hitchens, Gerd Muehllehner, and Everett W. Stoub

Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania; PET Department, National Institutes of Health, Bethesda, Maryland; Division of Nuclear Medicine and Biophysics, UCLA School of Medicine, Los Angeles, California; Department of Radiology, University of Washington, Seattle, Washington; Division of Nuclear Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland; MD Anderson Cancer Center, University of Texas, Houston, Texas; Department of Nuclear Medicine, University of Iowa, Iowa City, Iowa; CTI PET Systems, Knoxville, Tennessee; General Electric Medical Systems, Milwaukee, Wisconsin; Positron Corporation, Houston, Texas; UGM Medical Systems, Philadelphia, Pennsylvania; and Nuclear Science Consulting Service, Hillsboro, Missouri

A standard set of performance measurements is proposed for use with positron emission tomographs. This set of measurements has been developed jointly by the Computer and Instrumentation Council of the Society of Nuclear Medicine and the National Electrical Manufacturers Association. The measurements include tests of spatial resolution, scatter fraction, sensitivity, count rate losses and randoms, uniformity, scatter correction, attenuation correction, and count rate linearity correction.

J Nucl Med 1991, 32:2342–2350

A few years ago the Computer and Instrumentation Council of the Society of Nuclear Medicine (SNM) recognized the need to standardize the measurement of performance of positron emission tomography (PET) scanners as the number of PET centers was starting to grow rapidly. A committee was organized to standardize performance measurements for the users of PET scanners for such purposes as acceptance testing. The National Electrical Manufacturers Association (NEMA) also formed a committee, shortly after, to develop performance measurement standards for PET instruments in a manner similar to their effort for scintillation cameras (*1*). The two groups decided to cooperate and exchange information in an effort to unify the needs of both groups, since many of the tests needed to specify an instrument in the factory are the same tests that are used for acceptance testing at a user's facility. At the same time, the European Economic Community (EEC) Concerted Action of Cellular Regeneration and Degeneration organized a task group on PET instrumentation to evaluate and standardize performance measurements. There has been ongoing communication and discussion among the three groups, however, there remain some differences in philosophy between the effort of the SNM and NEMA groups (henceforth referred to as the SNM/NEMA tests) and the European effort (EEC tests).

There are three important tasks to accomplish in standardizing performance measurements. The first is to establish a common methodologic language that clearly defines the experimental measurements which are to be performed. The second is to develop tools for performance evaluation and to provide reliable tests that can be used to evaluate different scanners, despite the differences that exist among them. The third is to structure the tests so that one can use them to understand and predict the accuracy of data from realistic in-vivo studies. The first task is relatively straightforward, but the latter two are more difficult. In evaluating scanners, it is clear that each scanner has its own strengths and weaknesses, and the tests should be able to illustrate both. The tests need to be designed to be applicable to all types of scanners, including brain and body, conventional (non-TOF) and time-of-flight (TOF), slice-oriented imaging with septa, and volume imaging without septa. Realistic predictions of in-vivo performance are difficult to make for any scanner, since a phantom can only approximate the activity distribution in a subject.

This paper will briefly describe the individual SNM/NEMA tests and discuss the rationale for each one, including what the test does and does not measure. Rather than show data taken with specific PET scanners, examples of the measurement procedures will be illustrated with simulated data. A comparison between the SNM/NEMA protocol and that of the EEC will be pointed out for each measurement. The EEC document has been completed and is available from the Commission of the European Communities (*2*).

Received Jul. 25, 1991; revision accepted Aug. 7, 1991.

For reprints contact: Joel S. Karp, PhD, Physics and Instrumentation, Nuclear Medicine, Dept. of Radiology, Donner 110, H.U.P., 3400 Spruce St., Philadelphia, PA 19104.

It is important to recognize that this paper does not represent an official standard endorsed by NEMA, since that will require approval of the NEMA Codes and Standards committee of the document presented to them. A separate NEMA document is currently being prepared (*3*), which contains more precise language on how to perform the tests and analyze the results for the same group of measurements. The close cooperation between the members of the SNM and NEMA committee members (who are represented by the author list) has led to a general agreement on all of the tests, and one unified proposal for PET performance standards.

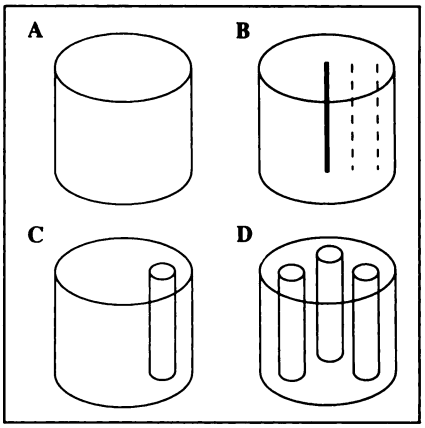

FIGURE 1. The PET performance phantom is 20 cm in diameter by 18.5 cm long and is shown in four configurations: (A) for sensitivity, count rate losses, uniformity, and count rate linearity correction, (B) for scatter fraction, (C) for scatter correction, and (D) for attenuation correction.

PERFORMANCE TESTS

The performance tests are divided into two groups. The first group includes the basic intrinsic measurements of: (1) spatial resolution (2) scatter fraction, (3) true sensitivity, and (4) count rate losses and randoms. The second group includes measurements of the accuracy of corrections for physical effects, specifically: (5) uniformity correction, (6) scatter correction (7) attenuation correction, and (8) count rate linearity correction. The only isotope that is used is ^{18}F, which of course requires a cyclotron. It is the experience of the authors that ^{68}Ga, which is available from a generator, is not as simple to calibrate for the sensitivity and count rate measurements and the longer positron range precludes its use for spatial resolution. Each test requires that the operating parameters of the scanner be adjusted as they would be for a typical patient study, including the energy window, axial acceptance angle, coincidence time window, and slice thickness. The data processing and reconstruction algorithms should also be the same as used for a typical patient study, although some tests require the use of a ramp reconstruction filter for standardization among systems. System International (SI) units are used for all reports of performance measurements, although customary units such as millicuries for activity can be included in parenthetical statements along with the SI unit of Becquerels.

The EEC group includes similar tests for the four intrinsic measurements as well as for uniformity correction. The other tests of accuracy of correction were discussed but not included in the final set of measurements.

Phantom

The PET performance phantom (Fig. 1) is a fillable cylinder of Lucite (polymethyl methacrylate) with an outside diameter of 20.3 ± 0.3 cm and a wall thickness of 0.3 ± 0.1 cm (which corresponds to a standard 8 inch tube with a wall thickness of 0.125 inch). The interior length of 18.5 cm was chosen since it is longer than the axial field-of-view (FOV) of all current machines. In the event that a machine has an axial FOV longer than the phantom, the phantom will need to be positioned more than once to measure slice sensitivity and scatter fraction for the whole axial length. The performance phantom is larger than a head but smaller than a body, and therefore serves as a compromise for both types of imaging. The phantom is shown in four configurations in Figure 1: (A) with no inserts for the sensitivity, count rate losses, uniformity, and count rate linearity correction measurements; (B) with line source inserts at 0, 4.0 and 8.0 cm radii for the scatter fraction measurement; (C) with a 5-cm diameter fillable Lucite cylinder (for water) at 6 cm radius for the scatter correction measurement; and (D) with three 5-cm diameter cylindrical inserts of different attenuation at 6 cm radius for the attenuation correction measurement; two are fillable (for air and water) and one is solid Teflon (for bone).

The outside dimensions of the source used for the spatial resolution and axial profile measurements, which are performed in air, should be no larger than 2 mm in diameter for the line, or in all directions for the point. The source should be made with ^{18}F in either a needle or capillary tube.

The EEC tests utilize the same cylindrical phantom, as well as an additional body-shaped phantom, together referred to as the EEC Emission phantom. There are three configurations of this phantom to simulate brain (although, as noted, the performance phantom is larger than a head), heart, and thorax imaging (*3*). The phantom allows for line source and point source inserts for the spatial resolution tests, which are performed in water.

Intrinsic Measurements

Spatial Resolution. The purpose of the measurement of spatial resolution is to characterize the width of the image point spread function (PSF), which is reported as the full-width at half-maximum (FWHM). A line source is positioned along the axial direction for the measurement of the transverse resolution in two directions; along the radial direction (radially) and perpendicular to the radial direction (tangentially) (Fig. 2). Most multi-ring systems with slice-defining septa (two-dimensional slice imaging) do not have sufficient axial sampling to measure a meaningful axial resolution, so the axial slice profile width is measured by moving a point source in fine steps along the axial direction. Since this axial scanning technique does not correspond to the axial resolution in a clinical situation, we prefer to call this latter measurement the axial slice

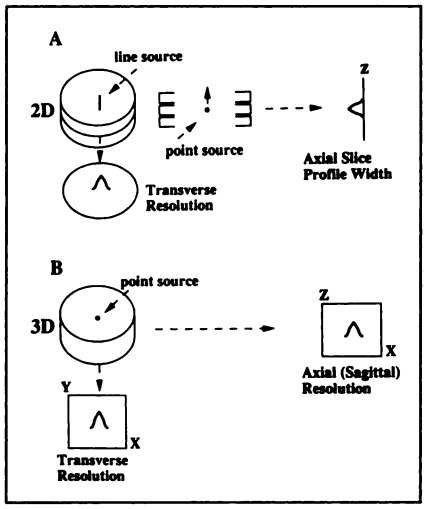

FIGURE 2. (A) Experiment setup to measure transverse spatial resolution for a two-dimensional multi-ring system with a line source parallel to the axial direction and a point source moved incrementally in axial direction to measure the axial slice profile width. (B) For a three-dimensional volume imaging system, a point source can be used to simultaneously measure both transverse and axial resolution.

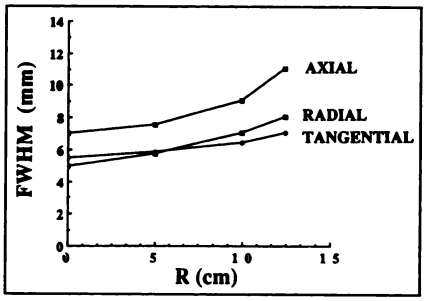

FIGURE 3. Sample data for spatial resolution as a function of radius shows tangential, radial, and axial resolutions at the FWHM levels.

profile width, rather than axial resolution as it has traditionally been referred to. For systems that have fine axial sampling (three-dimensional volume imaging), the axial resolution can also be measured, using a stationary point source. We consider a system to have fine spatial sampling in the axial direction if the sampling is less than one-third of the axial slice profile width.

Transverse resolution is determined by imaging the source *in air* at several locations within the FOV, at R = 0, 5, 10, 15, 20 cm (or as far as the transverse FOV permits). It is recommended that at each radial distance, four measurements be taken at different positions in the FOV in order to average the result and thus minimize the experimental uncertainty. For the axial measurements, the point source is also radially positioned in 5-cm increments (starting from the center). The axial slice profile width measurement requires the source to be moved in fine steps along the entire axial FOV, while for the resolution measurement, it is sufficient to collect data for several axial positions, Z = 0, 2, 4, 6 cm (for the extent of the axial FOV). The data for the resolution measurements are reconstructed with a ramp filter for filtered backprojection. The intrinsic spatial resolution is measured only if the pixel dimension in the image matrix is less than one-tenth the FWHM. However, with a typical matrix size of 128 × 128 and transverse FOV of 26 cm, a spatial resolution of 5 mm (FWHM) implies a pixel dimension of four-tenths the FWHM; thus the measured spatial resolution will be slightly worse than the true value. If zooming can be used in the reconstruction to image a smaller FOV, which also decreases the pixel dimension for a given matrix size, then the spatial resolution should be measured under these circumstances, as well.

The FWHM is calculated by linear interpolation between nearest channels. The data should be tabulated or plotted for transverse resolution, both tangential and radial, and axial resolution (if applicable), as a function of the radius, averaged over all slices (Fig. 3). For the situation where the results depend on the type of slice, such as "direct" and "cross" slices, then separate plots should reflect the average of each type.

The SNM/NEMA measurement of spatial resolution accomplishes several goals. It characterizes the shape of the reconstructed PSF, at the FWHM level, in the three orthogonal directions; tangential, radial, and axial. This allows a best-case evaluation of scanners, taking into account the variation over the FOV. The data are taken only at low count rates, so that potential high count rate problems are not encountered. These problems may occur with systems that use a high degree of analog encoding between the crystals and photomultiplier tubes (PMTs) because of pileup of events.

Since the images are reconstructed with sharp filters, the results do not directly indicate the spatial resolution expected in tissue with limited statistics. In this case, the data are usually reconstructed with a smooth filter, the choice of which is very subjective in choice, and, therefore, difficult to standardize. Assuming all other performance parameters are equal, a scanner with better intrinsic spatial resolution in air will have better spatial resolution in a clinical situation, as well. The intrinsic spatial resolution does not directly provide information that can be used to perform quantitative partial volume corrections. A test is included in the sensitivity measurement (the axial position sensitivity profile test) to assess the severity of the partial volume error due to coarse axial sampling.

The EEC procedure for measuring the spatial resolution recommends taking the spatial resolution measurements *in water*, which effectively includes scatter as part of this measurement, and using a smooth (i.e., clinical) reconstruction filter. Therefore, these results will indicate a poorer spatial resolution, but one which is closer to that seen with patient studies. In addition, an attempt is made to measure the recovery coefficients using spheres of different sizes. In practice, the corrections for finite spatial resolution are very difficult to implement accurately, since they depend on the unknown size and shape of the object.

Scatter Fraction. The scatter fraction is a measure of relative system sensitivity to scattered radiation. For a given source distribution, a lower scatter fraction is more desirable, regardless of how accurate the method for scatter correction, since all correction techniques add noise to the

image. Here, we define the scatter fraction to be the ratio of scattered events to total events, which are measured by the scanner at a low enough count rate so that random coincidences are negligible. Total events therefore are the sum of unscattered events (trues) and scattered events.

Data are acquired with a line source of activity in the performance phantom filled with water, and sorted into a sinogram, which is the two-dimensional projection space representation (projection ray versus angle) of a transverse plane. The line source is placed sequentially at three radii, 0, 4.0, and 8.0 cm, in order to determine the average scatter over the whole field-of-view. Since the physical FOV of different scanners vary, a fixed diameter of 24 cm (4 cm larger than the phantom) is used for the calculation of scatter, as well as some of the subsequent tests. The sinogram profile is used to calculate the number of scatter events within the FOV, and the number of trues within a 2-cm radius of the source (Fig. 4). The scatter within the peak is estimated by assuming a constant background, the level of which is determined by the average of the pixel intensities near the edge of the peak (at ±2 cm). For the source in the center, the sinogram profile is independent of angle, and so can first be summed over all angles before it is analyzed. For the off-center positions, the sinogram profile needs to be analyzed as a function of angle, and averaged. The average scatter fraction over the whole area of the phantom is calculated by weighting each position of the source by the area of the annulus at that radius, since the scatter for the cylinder can be assumed to be symmetric about the center. The scatter fraction is given by:

$$SF = \frac{[R_s(0) + 8 \cdot R_s(4.0) + 16 \cdot R_s(8.0)]}{[R_{tot}(0) + 8 \cdot R_{tot}(4.0) + 16 \cdot R_{tot}(8.0)]},$$

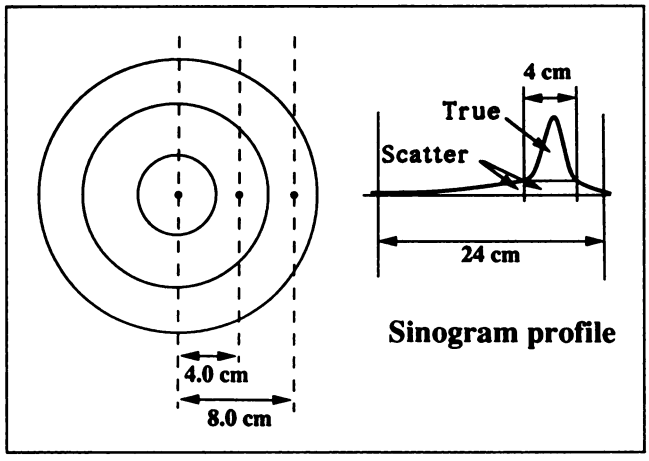

FIGURE 4. Scatter fraction measurement requires three acquisitions of a line source in the PET performance phantom filled with water. The fraction of scattered events is determined from the sinogram line spread function profile, averaged over angles. The scattered events under the peak are estimated using linear interpolation between the points at ±2 cm from the source center.

where R_s is the count rate per unit activity of scattered events (at 0, 4.0 and 8.0 cm) and R_{tot} is the total count rate per unit activity. For long acquisitions, one needs to calculate the mid-time of the scan (for ^{18}F, $t_{1/2}$ = 109.8 min) to determine count rate per unit of activity. Typically, the scatter fraction is highest in the center, but the relative contribution to the average over the FOV is small. The scatter fraction should be reported for each slice, as well as the average for all slices.

The scatter fraction, as determined from the sinogram data, depends greatly on the intrinsic geometry and shielding design of the system, as well as the energy window, which is dictated by the choice of detector material. The scatter fraction, however, does not directly indicate the contrast and noise in the image. Another important consideration is that the scatter fraction, as measured in the performance phantom, is not representative of a realistic head or body size. It is useful for standard evaluations among scanners of different configurations. However, in the effort to use one phantom, the diameter (20 cm) is larger than a brain and smaller than a body, and the axial length (18.5 cm) was chosen to be larger by several centimeters than all existing scanners. This makes it difficult to predict the scatter fraction of a realistic distribution, such as a head, which is typically 18 × 10 cm.

A more direct measurement of the scatter fraction in the reconstructed image has been proposed by the EEC group. The measurement procedure is the same; however, the analysis is performed on the reconstructed image instead of on the sinogram data. This, in fact, is actually a measure of the contrast, rather than the intrinsic scatter, in the image, since the contribution from the reconstruction algorithm is included. The results from this procedure typically indicate a lower scatter fraction than that from the raw, unprocessed sinogram data, because of the effect of the reconstruction filter.

Sensitivity. The volume sensitivity measurement is designed to measure the counting efficiency of the system with a known amount and distribution of activity. A low specific activity concentration (where the random fraction and %deadtime are each less than 1%) is loaded into the performance phantom, and the count rate is recorded from the sinogram data from within a 24-cm transverse FOV (Fig. 5A). This allows one to determine the true sensitivity, by subtracting the fraction of scattered events from the total, since the scatter fraction was measured for the same water-filled phantom. The volume sensitivity is recorded as cps/given activity concentration (μCi/ml or MBq/ml). The volume sensitivity is the sum of the slice sensitivities, as long as each event is recorded once. In other words, sensitivity should be independent of processing such as axial smoothing. For a multi-ring system, the direct slices will typically have a lower average sensitivity than the cross slices which accept data from adjacent slices (Fig. 6). Some new systems without septa, such as the UGM PENN-PET 240H and Siemens/CTI-953 (septa retracted),

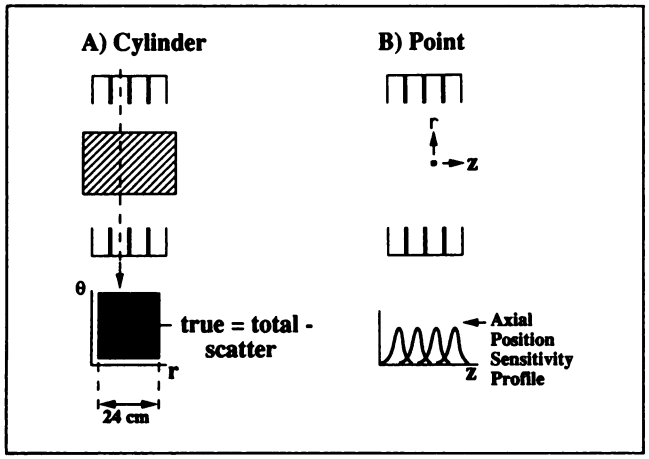

FIGURE 5. (A) System sensitivity measurement requires acquisition with the standard phantom filled with a uniform distribution of activity. (B) Axial position sensitivity profile is determined from the combination of the normalized axial slice profiles.

have a large axial acceptance angle, and the concept of direct and cross slices is not valid, although there may still be a variation in sensitivity as a function of axial position. For these systems, the sensitivity should be measured for each choice of axial acceptance angle that may be used for patient studies.

The volume sensitivity measurement reflects the true sensitivity in the absence of randoms and deadtime, with the estimated scatter subtracted, for a known, low-level distribution of activity. The trues, however, do not typically increase linearly with activity, because of deadtime, and therefore the sensitivity value is valid only in the limit of low activity. The performance under high count rate conditions, which is described below, is often necessary to predict the behavior of the system for realistic scanning conditions. In addition, the sensitivity cannot be used directly to estimate the noise in the image, because it does not take into consideration the amplification of noise from

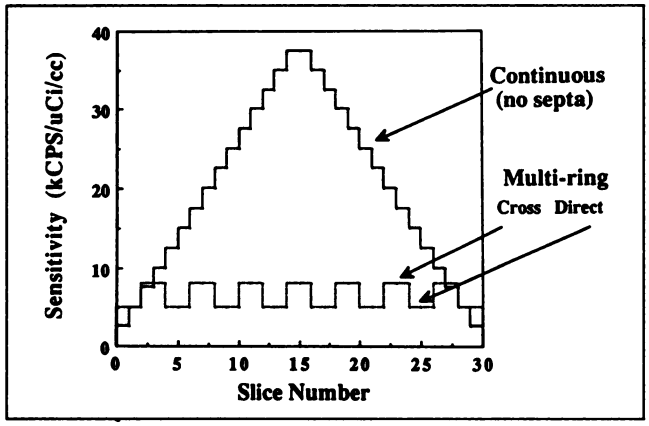

FIGURE 6. Sample data for sensitivity as a function of slice for multi-ring system with direct and cross slices (dotted line) and volume imaging system without septa (solid line). The sum of the slice sensitivities is the system sensitivity.

the reconstruction algorithm. While it is always true that more counts are better, it is difficult to know how *many* more counts are needed to make a noticeable difference in image quality, particularly with the tradeoffs in other parameters that are often necessary to increase sensitivity.

It is also worth noting that with this measurement, as well as with the count rate losses measurement, there is no independent check of the activity calibration. The accuracy of the dose calibrator is assumed, and every effort should be made to ensure that it is operating properly by checking it regularly with a standard calibration source.

An axial position sensitivity profile test is included to assist in assessing the severity of partial volume effects (Fig. 5B). The data from the axial slice profile width measurement are used, normalizing the response function for each slice to the slice sensitivity calculated above. The axial position sensitivity profile is the combination of all the normalized axial response functions. The profile is analyzed to determine the peaks (when the source is closely aligned with the center of each slice) and valleys (when the source is between two slices), and an average peak-to-valley ratio is reported. A large ratio implies a large partial volume error, depending on the axial location of the source. A volume imaging system with fine axial sampling is expected to have a smaller partial volume error than a multi-ring system with coarse axial sampling.

The EEC measurement for slice and volume sensitivity are essentially the same, although they do not have a test for the axial position sensitivity profile. As mentioned before, a test for partial volume error is included in their spatial resolution measurement.

Count Rate Losses and Randoms. Not all patient PET studies are carried out under conditions of low deadtime and randoms rates. It is necessary to measure losses of events due to deadtime and randoms as a function of activity level in order to understand the count rate behavior of the system for a wide variety of scanning conditions. The same performance phantom is used with an initial high specific activity loaded. Although ^{18}F ($t_{½}$ = 109.8 min) is recommended, ^{11}C ($t_{½}$ = 20 min) can also be used, since its shorter decay time makes the measurement time considerably shorter. Data should be taken until the randoms and deadtime are completely negligible. The total count rate from the sinogram data, as well as the randoms within a 24-cm transverse FOV, are measured as the activity decays (Fig. 7). The randoms need to be measured independently, if one is not able to determine them from either the singles rate or the delayed coincidence technique. The true events (T) are determined by subtracting the scatter (attained from the scatter fraction measurement) and randoms from the total. The percent deadtime (% DT), as a function of activity, is given by

$$\%DT = 1 - T/T_{extrap},$$

where the T_{extrap} is a linear function of the activity and is extrapolated from the low count rate data.

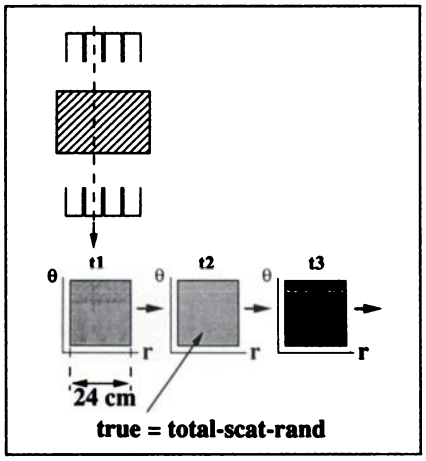

FIGURE 7. Count rate measurement requires dynamic acquisition (t1, t2, t3, . . .) of the PET performance phantom filled with a uniform distribution of activity. Trues and randoms are determined for a 24-cm transverse FOV, for each slice, from the sinogram data.

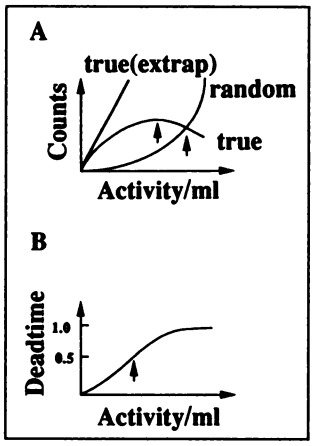

FIGURE 8. Sample data for (A) true and random count rates as a function of activity concentration and (B) %deadtime as a function of activity concentration. The three arrows indicate the points at which: (1) the randoms equal the trues, (2) the true count rate reaches its peak, or saturates, and (3) the %deadtime is 50%.

The trues and randoms should be plotted as a function of activity concentration (Fig. 8A). Traditionally, the point at which the trues equal the randoms has been used to rate the capability of the system. With most modern systems, deadtime is more of a problem because of the larger number of crystals coupled to each PMT, and so the point at which the deadtime is 50% is an alternative measure of the count rate capability of the system (Fig. 8B). A third limit is when the true count rate saturates or reaches its peak. We suggest reporting the point at which at least two of these three conditions are met, since it may not be possible to measure all three. It is important to know the count rate at which this occurs, as well as the activity level. A highly sensitive system may saturate at a relatively low activity concentration; but the true count rate at that level will likely be higher than that of a system with lower sensitivity at a higher activity concentration.

The count rate measurement defined above does not directly indicate the signal-to-noise ratio (SNR) in the image, which could be used to assess the image quality at different activity levels. It is possible, however, to calculate the noise equivalent count rate (NEC) directly from the true (T), scatter (S), and random (R) rates:

$$NEC = T/(1 + S/T + R/T).$$

This quantity defines an effective true count rate by accounting for the additional noise from the scatter and randoms, and assuming a noiseless estimate of the scatter and randoms corrections. The square root of the NEC is proportional to the SNR, although the NEC gives only a global measure which is not sensitive to local variations of particular source distributions.

This measurement, as well as the other measurements with the performance phantom, does not provide data for a realistic distribution, such as a brain or a heart. The EEC group suggests collecting count rate data for three different configurations, to simulate the head (which is identical to the one described here), the heart, and the thorax to attain realistic distributions. The difficulty with this procedure is that three times as much data must be taken and processed. On the other hand, the results with cardiac and thorax configurations are significantly different than for the performance phantom configuration, and do provide additional useful information for these imaging protocols. In addition, a test is included to investigate the misplacement of events, axially, as the count rate increases. This is most pronounced in systems which incorporate position encoding, since event pile-up results in miscalculation of axial position, as well as transverse position.

Measurement of Accuracy of Corrections

Uniformity. The purpose of the uniformity measurement is to measure deviations in the reconstructed image from a uniform response.

The performance phantom is filled with a moderate amount of activity (such that the random fraction and deadtime are each less than 20%). The phantom is placed 2.5 cm vertically off-center in order to reduce the symmetry of the scanning geometry, but still allow the test to be performed in a brain scanner. The phantom is scanned for a long enough period so that an average 20 million counts/slice are collected. The data are reconstructed with a ramp filter and with corrections for detector normalization, deadtime, randoms, scatter, and attenuation. Attenuation correction is done analytically, with the known attenuation coefficient of water, to eliminate potential effects of the transmission method of correction.

The analysis involves using a region of interest (ROI) map consists of contiguous 1 × 1-cm square regions, fully inscribed into a circle of 18-cm diameter and centered on the image of the cylindrical phantom (Fig. 9). The plus and minus nonuniformities for each slice are defined as follows:

$$NU(+) = +(C_{max} - C_{avg})/C_{avg},$$

$$NU(-) = -(C_{avg} - C_{min})/C_{avg},$$

where C_{max} is the maximum number of counts in any square region within the slice, C_{min} is the minimum number of counts in any square region within the slice, and

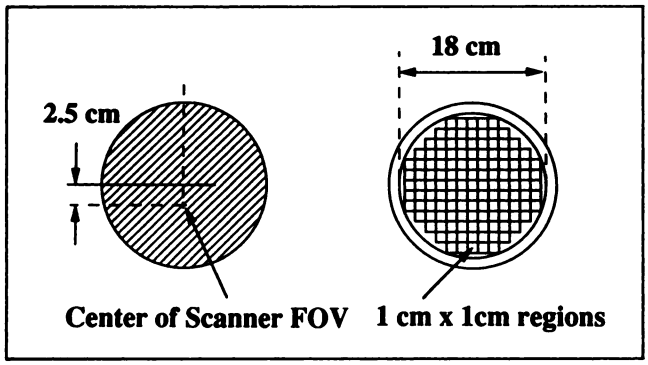

FIGURE 9. Uniformity measurement requires a long acquisition of the PET performance phantom filled with a uniform distribution of activity, offset by 2.5 cm from the center of the scanner, as shown on the left. The percent nonuniformity is determined from the counts within 1-cm square ROIs in the image, as shown on the right.

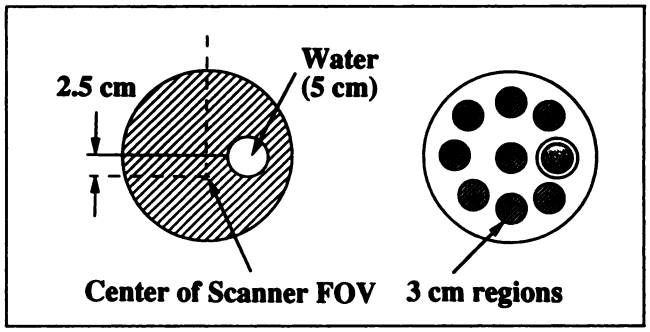

FIGURE 10. Scatter correction measurement requires acquisition of the PET performance phantom filled with a uniform distribution of low-level activity with a cold-water insert. The scatter correction error is determined from the ratio of the counts in the insert region to the counts in the background regions.

C_{avg} is the average number of counts of all the regions within the slice. In addition, the standard deviation and coefficient of variation are calculated for the regions. Volume nonuniformity is calculated for the whole system, as well, by searching for the C_{max}, C_{min}, and C_{avg} values over all slices.

Although it is time consuming to collect as many as 20 million counts/slice, and very difficult for systems with many thin slices, the statistical variation in the data is significant at count densities below that level. The idea of the uniformity measurement is to minimize the deviations due to statistical fluctuations, and measure only systematic variations due to system imperfections, including those due to software corrections and the reconstruction algorithm. The EEC measurement is almost identical to this measurement, except for the fact that they recommend only 2 million counts/slice. While this is indeed more representative of the count density encountered with patient studies, it is more a measure of the random fluctuations, which are independent of scanner performance, than the systematic variations, which depend on scanner performance.

Scatter Correction. The scatter correction measurement assesses the accuracy of the scatter correction technique in the image.

The performance phantom is filled with a low amount of activity (random fraction and %deadtime each less than 1%) and a 5-cm diameter cold insert (water) is placed in the phantom. The phantom is again placed 2.5 cm vertically off-center. The data are collected (at least 2 million counts per slice) and reconstructed with all corrections and a ramp filter. Again, an analytical attenuation correction is used. Eight ROIs are defined in the background area (warm activity) and one ROI (3-cm diameter) in the cold area (Fig. 10). The scatter correction error is given by:

$$\Delta S_{corr} = C_{insert}/C_{back(avg)},$$

where C_{insert} is the number of counts in the cold ROI and $C_{back(avg)}$ is the average of the number of counts in the warm ROIs. The scatter correction error, which should be close to zero, is determined for each slice, as well as the average for the whole system. A positive error implies an underestimation of the scatter, while a negative error implies overestimation. The error does not indicate the degree to which the scatter correction increases the noise in the image.

As previously mentioned, the EEC group does not include a measurement of the scatter correction error nor the attenuation and count rate linearity corrections described in the next two sections.

Attenuation Correction. This measurement is designed to assess the accuracy of the transmission method of attenuation correction. The performance phantom (offset by 2.5 cm) with three inserts are used, each with different attenuation coefficients; air (lungs), water (tissue), and Teflon (bone). It is recommended to collect the emission data first, with low activity (random fraction and %deadtime each less than 1%) to minimize count rate loss corrections, but with a yield of at least 2 million counts per slice. The transmission study should be performed after the activity has decayed (e.g., the next morning), to minimize the repositioning error. The transmission scan and blank scan are measured by the same method used for patient studies. Each image is reconstructed with all corrections applied. Eight ROIs per image are defined over the background (warm area) and one over each insert in order to calculate remnant errors after attenuation correction for each insert (Fig. 11). The residual attenuation error for each insert is given by:

$$\Delta A(I) = C_{insert(I)}/C_{back(avg)}.$$

As defined, this value includes effects of attenuation correction and scatter corrections, although with the scatter correction measurement, one attempts to differentiate between these errors by using an analytical attenuation correction.

Count Rate Linearity Correction. The accuracy of cor-

rections for deadtime losses and randoms can be determined from the count rate data taken earlier. The data need to be reconstructed with all corrections using a ramp filter and analytical attenuation correction. One large region is defined (18 cm) in the center of the image and the residual error as a function of count rate is given by

$$\Delta R = 1 - R/R_{extrap},$$

where the values R_{extrap} are determined from the low count rate data (using the last three acquisitions) and assumed to be a linear function of activity. The errors are calculated for each slice as a function of activity and reported at two of the three activity levels discussed previously: (1) the randoms equal the trues, (2) the %deadtime is 50%, and (3) the trues are saturated or have reached their peak.

SUMMARY AND DISCUSSION

The goal of the SNM and NEMA committees is to standardize the performance measurements for PET scanners so that both manufacturers and users will benefit from a common set of testing procedures. For both groups, it is important that the measurements be simple, yet physically meaningful. While the measurements do not require any special equipment, other than the performance phantom, it may be necessary to write some programs for data analysis. This will likely be provided by the manufacturers, since they will perform the same tests in the factory. It is also important that the measurements can be performed in a reasonable amount of time. The experience of the committee members suggests that the measurements and analysis can be performed in about two weeks, although this assumes that one has experience with the measurement procedures and the experiments do not need to be repeated.

Considerable effort has been made to keep the American and European measurements similar to each other. In certain cases they are identical, but in others they differ with respect to either the activity distribution or the method of data analysis. The overriding concern of the SNM and NEMA groups is to establish a standard to characterize the physical performance of PET scanners, in order to evaluate performance and understand the abilities and limitations of the scanner for use in patient studies. The overriding concern of the EEC group is that the measurements would approximate a clinical situation as closely as possible, so that the measurements can be used to predict and interpret patient studies. For example, should spatial resolution be measured in air or in water, and should the data be reconstructed with a ramp or with a smooth "clinical" filter? Both measurements are useful and complimentary. The SNM/NEMA measurement represents the intrinsic performance of the scanner and is useful for comparison. The EEC measurement gives one a better idea what the resolution will be for a study with limited statistics. The difference between the results is not so much a question of which is more relevant for patient studies, but which is a better standard to use to measure the performance of the scanner. That is a matter of opinion, and a difficult point to resolve. As another example, should the scatter fraction be measured in the sinogram or in the image? The SNM/NEMA measurement represents the intrinsic scatter, while the EEC measurement represents the contrast (rather than the scatter) in the image.

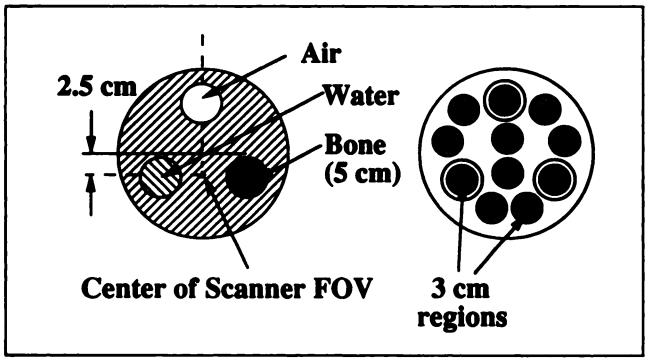

FIGURE 11. Attenuation correction measurement requires a transmission scan and an emission scan of the PET performance phantom with three inserts of different attenuation coefficients. The attenuation correction error in the emission scan is determined from the ratio of the counts in the insert regions to the counts in background regions.

The SNM/NEMA performance phantom is identical to the EEC head insert, but there is no comparable SNM/NEMA phantom to the EEC body phantom. The 20-cm diameter by 18.5-cm long cylinder is not really representative of the head, and actually offers a standard configuration which is between the size of the head and the size of the body. In an effort to minimize the time required to do the tests and the number of phantoms required, the SNM and NEMA groups decided to use only one phantom. The EEC body phantom allows one to study configurations representing cardiac and thorax imaging, and therefore offers additional information to the performance tests presented here and in the NEMA document (*2*). It is arguable, however, that each group of investigators will want to design tests even more specific to their imaging protocol, which clearly falls outside the boundaries of a standard set of performance measurements.

It is hoped that the set of measurements put forth by the NEMA and SNM committees is general enough so as not to be outdated in the near future due to scanner design evolution. The groups considered all types of scanners currently available, including multi-ring BGO systems, time-of-flight systems, and volume imaging systems. Major design changes were anticipated to ensure that the tests will remain valid for a number of years. It is possible, however, that even with the current state-of-the-art scanners, the present set of measurements may need to be revised as the general community becomes familiar with the tests. To date, the set of measurements has been performed on only a small number of scanners, by a small

number of investigators. It is hoped that the proposed set of standards presented in this paper will be used by many PET investigators, and that as the database of measurements increases, the tests can be reevaluated, and perhaps modified.

Finally, there are several measurements which were discussed by the groups, but not included because of the difficulty in developing standards. An example of such a measurement is a signal-to-noise measurement, or some other form of image quality measurement. The SNM/NEMA measurements are an attempt to standardize a set of measurements which are already accepted by the general PET community, rather than to research new ideas. Each aspect of performance is measured separately, particularly for the intrinsic measurements. The uniformity measurement, on the other hand, is in some ways a general image quality measurement which includes many effects together. It is hoped that the set of measurements finds acceptance by the PET community and proves to be useful for evaluating and understanding the performance of PET scanners.

ACKNOWLEDGMENTS

The authors thank Robert Zimmerman, who organized the SNM committee, and Richard Eaton from NEMA, who organized and coordinated the NEMA meetings. We also thank Dr. R. Guzzardi, who is responsible for the cooperation between these groups and the EEC Instrumentation Task Group.

REFERENCES

1. Muehllehner G, Wake RH, Sano R. Standards for performance of scintillation cameras. *J Nucl Med* 1981;22:72–77.
2. EEC Concerted Action (PET Investigation of Cellular Regeneration and Degeneration). Performance evaluation of positron emission tomographs, 1990.
3. NEMA. NEMA Standards Publication. Performance measurements of positron emission tomographs. (in preparation).