Published in final edited form as:

IEEE Trans Nucl Sci. 2012 October ; 59(5): 1940–1947. doi:10.1109/TNS.2012.2197414.

Effect of Scan Time on Oncologic Lesion Detection in Whole-Body PET

Dan J. Kadrmas [Member, IEEE],

Utah Center for Advanced Imaging Research (UCAIR), Department of Radiology, University of Utah, Salt Lake City, UT 84108-1218 USA

M. Bugrahan Oktay [Member, IEEE],

Utah Center for Advanced Imaging Research (UCAIR), Department of Radiology, University of Utah, Salt Lake City, UT 84108-1218 USA

Michael E. Casey [Senior Member, IEEE], and

Siemens Medical Solutions USA, Knoxville, TN 37932 USA

James J. Hamill [Member, IEEE]

Siemens Medical Solutions USA, Knoxville, TN 37932 USA

Dan J. Kadrmas: kadrmas@ucair.med.utah.edu

Abstract

Lesion-detection performance in oncologic PET depends in part upon count statistics, with shorter scans having higher noise and reduced lesion detectability. However, advanced techniques such as time-of-flight (TOF) and point spread function (PSF) modeling can improve lesion detection. This work investigates the relationship between reducing count levels (as a surrogate for scan time) and reconstructing with PSF model and TOF. A series of twenty-four whole-body phantom scans was acquired on a Biograph mCT TOF PET/CT scanner using the experimental methodology prescribed for the Utah PET Lesion Detection Database. Six scans were acquired each day over four days, with up to 23 ^{68}Ge shell-less lesions (diam. 6, 8, 10, 12, 16 mm) distributed throughout the phantom thorax and pelvis. Each scan acquired 6 bed positions at 240 s/bed in listmode format. The listmode files were then statistically pruned, preserving Poisson statistics, to equivalent count levels for scan times of 180 s, 120 s, 90 s, 60 s, 45 s, 30 s, and 15 s per bed field-of-view, corresponding to whole-body scan times of 1.5–24 min. Each dataset was reconstructed using ordinary Poisson line-of-response (LOR) OSEM, with PSF model, with TOF, and with PSF +TOF. Localization receiver operating characteristics (LROC) analysis was then performed using the channelized non-prewhitened (CNPW) observer. The results were analyzed to delineate the relationship between scan time, reconstruction method, and strength of post-reconstruction filter. Lesion-detection performance degraded as scan time was reduced, and progressively stronger filters were required to maximize performance for the shorter scans. PSF modeling and TOF were found to improve detection performance, but the degree of improvement for TOF was much larger than for PSF for the large phantom used in this study. Notably, the images using TOF provided equivalent lesion-detection performance to the images without TOF for scan durations 40% shorter, suggesting that TOF may offset, at least in part, the need for longer scan times in larger patients.

© 2012 IEEE

The content is solely the responsibility of the authors and does not necessarily represent the views of the NIBIB or NCI.

A preliminary version of this paper was presented at the 2010 IEEE Nuclear Science Symposium and Medical Imaging Conference, and made available in the published proceedings of that conference.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Index Terms

Image quality; positron emission tomography

I. Introduction

POSITRON emission tomography has emerged as the lead modality for imaging cancer, and PET with ^{18}F -fluorodeoxyglucose (FDG) is routinely used to facilitate initial diagnosis of suspicious lesions, to stage, grade, monitor response to therapy, and to identify residual or recurrent disease following therapy. Many of these clinical applications involve identifying and/or localizing focal lesions of elevated tracer uptake within structured and noisy backgrounds—*i.e.*, lesion-detection tasks. Research and development of new PET instrumentation, reconstruction algorithms, and image processing methods require objective evaluation of the image quality offered by these advances. A powerful and accepted approach is to evaluate image quality in terms of performance for clinically-relevant tasks, *e.g.*, the detection of focal warm lesions for general oncologic PET imaging.

A significant body of research has been performed in the development of human and numerical observer techniques for evaluating images in terms of lesion-detection performance [1]–[8], and these methods have been used to assess several aspects of PET instrumentation and algorithms [9]–[22]. Our group has developed a phantom-based methodology for evaluating focal lesion-detection performance with high statistical power using measured experimental data in sophisticated whole-body phantoms. The experimental methods provide numerous lesion-present and lesion-absent test images representative of general oncologic PET detection tasks, with known truth suitable for receiver operating characteristic (ROC) studies [23]–[26] with both human and numerical [7], [8], [14], [16], [27]–[31] observers, and the datasets have been compiled into an inter-institutional collaborative resource for assessing PET lesion-detection performance called the Utah PET Lesion-Detection Database [34]. This approach was recently used to evaluate the lesion-detection performance of several landmark algorithms for iterative fully-3D PET reconstruction, plus modeling of the spatially-variant point spread function (PSF) and prototype time-of-flight (TOF) imaging [32], [33]. Related studies of lesion-detection in TOF PET have been performed by other groups using simpler phantoms [20] and human data with synthesized lesions [21], [22], and our studies complement this work.

As patient volume for oncologic PET/CT studies increases, financial and logistical factors favor increasing throughput by using scans of shorter duration. However, image quality is strongly affected by count statistics, with shorter scan times producing noisier images. On the other hand, advanced reconstruction algorithms, PSF modeling, and TOF techniques can improve image quality. The primary objective of this work is to characterize the relationship between reduced count levels (as a surrogate for reduced scan times) and general oncologic PET lesion-detection performance when using these techniques. This complements related work studying the effects of TOF PET on lesion-detection using the Hotelling observer [22] and human localization ROC (LROC) [21] in human subjects with manually inserted lesions. Four reconstruction methods are evaluated and compared at eight count levels, corresponding to scan times ranging from 1.5–24 minutes whole-body (“eyes-to-thighs”). This work also adds a new experimental series to the Utah PET Lesion-Detection Database, consisting of 24 whole-body scans acquired on a modern TOF PET/CT scanner. These experiments used a custom-built large thorax phantom approximately 50% larger than the previously-used “medium” thorax, with the same brain and pelvis compartments. Since the benefits of TOF PET are closely dependent on object size [35]–[37], this larger phantom provides an important extension to evaluation of time-of-flight PET for larger patient sizes.

The following sections describe the experimental acquisitions, data processing methods, and LROC studies using the channelized non-prewhitened (CNPW) numerical observer. The effect of reducing scan time for each of four reconstruction techniques is then analyzed, and conclusions are drawn based on these data.

II. Methods

A. Large Whole-Body Lesion-Detection Phantom

The whole-body lesion-detection phantom, pictured in Fig. 1, is designed to model general oncologic FDG PET imaging. The phantom consists of a 3D Hoffman brain phantom (Data Spectrum Corp., Hillsborough, NC), a thorax component containing liver and lung compartments plus realistic rib cage and spine attenuating structures (Radiology Support Devices, Inc., Long Beach, CA), and a pelvis component consisting of a $31.8 \times 23.4 \times 20.0$ cm elliptical cylinder (Data Spectrum) with bladder compartment. New to this study, a custom large version of the thorax phantom was constructed and used. The fillable volume of this phantom (~ 20 liters) is 65% larger than that of the standard (“medium”) RSD thorax phantom. With the increased size of the body cavity, the large thorax has a removable spine/rib cage assembly as well as the lung and liver inserts. Accounting for the missing mass of arm and legs (based on the standard man model), the medium and large whole-body phantoms model patients of approximately 75 kg and 92 kg, respectively. The phantom also has a number of custom features, including the use of nylon mesh bags in the lungs and low water resistance open cell black foam in the tissue compartments, as described in our earlier work [18], [32]–[34].

This work used a number of shell-less ^{68}Ge ($T_{1/2} = 270.8\text{d}$) silicone gel lesions [38], with diameters of 6, 8, 10, 12, and 16 mm. The lesions were made with two activity concentrations: “warm” and “hot” at 0.35 and $0.50 \mu\text{Ci}/\text{cm}^3$, respectively, calibrated to the days of the experiment. These long-lived lesions provide challenging detection tasks with variable lesion contrasts as described in the next section.

B. Experimental Acquisitions

The phantom was scanned 24 times over four days on a Biograph mCT TOF PET/CT scanner with TrueV™ (axial field-of-view = 21.8 cm, time resolution 527.5 ± 4.9 ps [44], Siemens Medical Solutions), acquiring 6 sequential whole-body scans each day over a period of two ^{18}F half-lives as listed in Table I. The overall activity levels for the 6 scans broadly cover the full range of activity levels that would be present clinically for sites administering 8–15 mCi FDG with uptake times of 60–120 min. For example, the total activity present for Scan 1 mimics administration of 15 mCi FDG with uptake time of 60 min., whereas Scan 6 models administration of approximately 8 mCi FDG with uptake time of 2 hours. Each scan 1–6 provides differing count levels, deadtime, and levels of random coincidences corresponding to the activity levels just summarized. In order to model a narrower range of activity levels that might be encountered at a single institution with more limited variability in administered activity and uptake time, a subset of the data including only scans 2–4 (activity levels 5.3–8.4 mCi) was also processed separately from the full scan 1–6 datasets.

The phantom was prepared with ^{18}F concentrations in water according to the following relative activity distributions, based on 12 oncologic FDG PET studies performed at our institution: soft tissue 1.0:1; lungs 0.3–0.5:1; liver 1.8:1; brain(average) 6.0:1; and bladder 15.0:1. The total activity was calibrated to 10.6 mCi at the start of scanning, and was the same each day. One day of scanning had no lesions present, while the other 3 days had 21–23 lesions distributed throughout the lungs, mediastinum, liver, abdomen, and pelvis compartments, with different distributions each day. Each whole-body scan acquired

listmode data for 240 s per bed field-of-view (FOV) over 6 bed positions. Since the ^{18}F background decayed by a factor of 4.0 while the ^{68}Ge lesions remained essentially constant, this scanning protocol gave six whole-body scans with both varying count levels and increasing lesion target:background ratios. This is an important feature of our scanning procedure, as it produces a range of noise levels and lesion contrasts representative of those encountered clinically, while also ensuring a high number of lesions that are close to the verge of detectability (necessary to obtain high statistical power for the observer studies).

C. Statistical Pruning to Shorter Scan Times

The original scan was acquired at 240 seconds per bed position. Each listmode file was then sequentially pruned, preserving Poisson statistics, down to count levels for equivalent scan times of 180 s, 120 s, 90 s, 60 s, 45 s, 30 s, and 15 s (corresponding to 6-bed whole-body scans of 24, 18, 12, 9, 6, 4.5, 3, and 1.5 min., respectively). These scan times were selected to characterize the full range of current clinical scan times down to extremely short scans, and should *not* be misinterpreted as recommending very short scans. Also note that the objective of this study was to broadly characterize the effect of reducing scan time upon lesion-detection performance, not to directly link count level to detectability. The original 240 s per bed scan data include the full range of activity levels, and hence count levels, that would be encountered for clinical sites administering from 10 to 15 mCi FDG and waiting 60–120 min. before scanning. Also note that the count levels also vary slice-by-slice according to the phantom structures. Pruning of these data down to shorter scan times broadly models reducing scan times across the typical clinical range of administered activities and uptake times.

The statistical pruning procedure considered each coincidence event individually, passing it through an imaginary attenuator. Pseudo-random numbers evenly distributed across [0.0,1.0) were computed for each listmode event and compared to a retention threshold (*e.g.*, 0.75 for pruning from 240 s to 180 s). The event was either retained or discarded if the number was less than or greater than this threshold, respectively. Overall, this procedure reduced the count level in each listmode file to equivalent reduced scan times *without* changing the measurement activity distributions or lesion target:background ratios.

D. Image Reconstruction and Processing

Each listmode file was binned and reconstructed using four different reconstruction algorithms with the manufacturer-supplied scanner software. Each reconstruction included normalization and deadtime corrections, randoms correction using delayed-event subtraction with randoms smoothing, and scatter correction using the single scatter estimation technique extended for time-of-flight data. The baseline reconstruction algorithm was Ordinary Poisson line-of-response ordered-subsets expectation-maximization (LOR-OSEM) [39]–[42] with 14 subsets, where all corrections including arc-correction were applied within the system matrix of the iterative reconstruction. Here, all TOF bins were combined into a single bin to ignore the TOF measurements (*i.e.*, non-TOF). The in-plane reconstructed pixel size and slice thickness were 4.073 mm and 2.027 mm, respectively. The effects of PSF modeling (TrueXTM) and TOF were then each studied by reconstructing with spatially-variant PSF model [43], and then with 13 bins TOF data [44]. Finally, the combined effect of PSF modeling + TOF was studied by reconstructing with both together. We refer to these as LOR, PSF, TOF, and PSF+TOF, respectively. Each reconstruction was run to 10 iterations with 14 subsets, storing the intermediate images from each iteration for subsequent analysis. Finally, the effects of changing filter strength were studied by post-filtering each image (at each iteration) by 21 three dimensional Gaussian filters with standard deviations ranging from 0.0 (no filter) to 2.0 voxels in 0.1-voxel increments. Overall, this resulted in

6720 sets of reconstructed images (8 scan times $\times$ 4 algorithms $\times$ 10 iterations $\times$ 21 filters), with each set including 18 lesion-present and 6 lesion-absent whole-body scans.

For each set of images, the true location of each lesion was determined using both co-registered CT images and late “lesion-only” scans acquired > 12 hours after the initial scans of each day to allow for the ^{18}F background to decay to near zero. A total of 402 lesion-present cases (21–23 lesions $\times$ 6 scans/day $\times$ 3 days with lesions present) were available, along with 402 corresponding lesion-absent cases taken from the scans acquired without lesions. All lesions (and corresponding lesion-absent slices) were used for the observer studies described in the next section.

E. LROC Studies With Numerical CNPW Observer

The large number of images generated by this study precluded the use of human observers; however, model numerical observers have been found to correlate with human observer performance in many cases [7], [8], [14], [16], [27]–[31]. The experimental protocols used in this work provide the necessary data for training and applying the channelized non-prewhitened (CNPW) observer as developed by Gifford *et al.* [7], [8]. Briefly, the CNPW observer computes a perception measurement, z_n , at each image voxel according to:

$$z_n = \mathbf{w}_n^T [\hat{\mathbf{f}} - \mathbf{b}], \quad (1)$$

where $\mathbf{w}_n^T$ is the CNPW template image at voxel n , $\hat{\mathbf{f}}$ is the image to be tested, and $\mathbf{b}$ is the mean lesion absent image. The CNPW template $\mathbf{w}_n^T$ is the mean 2D lesion profile over a set of training images, mathematically projected onto a set of channel responses. The ten difference-of-Gaussian channels described in [8] were used, and the interested reader is referred to that paper for additional details. Training and application of the CNPW observer to our experimental lesion-detection phantom data is further described in [32], [34]. This work used a two-dimensional version of the CNPW observer applied to transaxial slices centered on each lesion (for the lesion-present cases). The training method labeled Method I in [32] was used, where all available data were used to compute $\mathbf{w}_n^T$ and $\mathbf{b}$. This training method produces the lowest noise in the observer setup, but also gives the observer some prior knowledge of the test cases as they are used for both training and testing. The approach was validated in [32] against more complex training methods that completely separate test and training data, and found to produce identical results to the precision level used.

It is well known that iterative convergence rates and optimal filtering parameters are dependent upon many factors, including the object size, activity distribution, count level, and reconstruction algorithm. In order to ensure fair characterization of the relationship between reduced scan times and lesion-detection performance, we first optimized the number of iterations and filter strength for each reconstruction algorithm and scan time. For each case, 210 LROC studies were performed, one each for iterations 1–10 at 21 different filter strengths. A radius of correct localization of 2.5 voxels was found to correctly identify ‘hits’ while minimizing random localizations, and was used throughout this study. Figures-of-merit including the probability of correct localization (P_{LOC}) and area under the LROC curve (A_{LROC}) were computed using the Mann-Whitney-Wilcoxon statistic and nonparametric estimation method of [26]. Error estimates for each metric were estimated as the standard deviation over 10,000 bootstrap trials. The number of iterations and filter strength maximizing A_{LROC} for each algorithm and scan time were identified, and these parameters were then used to characterize the scan time vs. lesion-detection performance tradeoff for each algorithm.

III. Results

Fig. 2 shows example maximum intensity projection (MIP) images for six sequential scans from one day of experiment reconstructed with 4 iterations of the baseline LOR-OSEM algorithm and no post-filter. These images provide an overview of the image features, varying count levels, and varying lesion contrasts provided by the experimental protocol at a single scan time (240 s/bed position). The following sections present the results for optimizing the number of iterations and filter strength for each algorithm and scan time, followed by analysis of the impact of reduced scan time on lesion detection performance.

A. Selection of Reconstruction Parameters

Objective comparison of lesion-detection performance for the different scan times and algorithms requires that each reconstruction be evaluated using near-optimal parameters. Optimization of the number of iterations and filter strength for each scan time and reconstruction algorithm is a multi-dimensional optimization problem. The very large number of test images provided by our experimental protocol, coupled with the computational power of modern workstations, enables this optimization to be robustly performed through fine empirical sampling of the parameter space for each algorithm and scan time. This work required 6,720 separate LROC studies, reading a total of 5,402,880 2D test images. Run on a Linux workstation with four 6-core Intel Xeon X5660 CPUs @ 2.80 GHz, all numerical observer studies were completed in 4 days running time. As such, this methodological approach permits rapid repeat LROC studies on a very large scale.

Fig. 3 summarizes all study results for optimizing the number of iterations and filter strengths for the full scan1–6 datasets. Overall, lesion-detection performance was weakly dependent on the number of iterations provided that sufficient iterations were performed to achieve near-peak performance (at least ~ 4 iterations for all cases). Optimal filter strength, on the other hand, showed a stronger dependence on both reconstruction algorithm and scan time. Reconstructions without PSF modeling required moderate-to-strong filtering in all cases to obtain near-peak detection performance, whereas the reconstructions with PSF model reached peak performance with only moderate filtering. Overall, the reconstructions with PSF model were less sensitive to filter strength, which may be considered an advantage of PSF modeling in practice as case-by-case filter optimization is not feasible in routine use. In addition, a strong relationship between the best filter strength and scan time was identified, where significantly stronger post-filtering was required to obtain peak detection performance for the noisier images resulting from the shorter scan times. While not surprising, these relationships clearly demonstrate important scan time-related effects in selecting the best reconstruction parameters for oncologic PET imaging.

B. Impact of Reduced Scan Time on Lesion-Detection

Fig. 4 shows example images for each algorithm at each of the eight scan times studied. Visualization of the warm 10 mm diameter lesion in the right lung of these image slices varied dramatically with imaging time, and the effects varied for the four reconstruction algorithms. This was representative of the overall detection study results. Fig. 5 presents LROC curves for each algorithm and scan duration, where all lesion sizes, contrasts and locations have been pooled for each algorithm and scan time. The overall results are summarized in Fig. 6, where A_{LROC} is plotted versus scan time for each case. Lesion-detection performance dropped markedly with the shortest scan times, reflecting the degradation of image quality as noise levels increased. All four cases were well-fit empirically by a logarithmic relationship between A_{LROC} and scan duration, with coefficients of determination R^2 from 0.98–0.99 for each case. This approximate parameterization may have some value in roughly estimating scan time effects, but requires

further investigation and should not be considered broadly predictive of lesion-detection performance versus scan time.

The data in Fig. 6 show different overall levels of lesion-detection performance for each of the four algorithms, with PSF modeling offering some improvement over the baseline LOR-OSEM algorithm, and TOF offering a dramatic improvement. As compared to our earlier work with the medium-sized phantom [32], [33], the relative improvement due to TOF vs. PSF was much larger in this work using the new large-sized phantom—while the PSF cases consistently outperformed the non-PSF cases by a small amount, the degree of improvement offered by TOF was much larger than that offered by PSF. This is consistent with previous work indicating that TOF PET offers more benefit for larger patients [35]–[37]. Interpolating between the datapoints along the curves in Fig. 6, TOF at 145 s scan time provided equivalent lesion-detection performance as the non-TOF LOR reconstruction at 240 s, effectively offering a 40% reduction in scan time for the same detectability. Note that this result is specific to the particular lesion sizes, contrasts, locations, and body size used in this study; similar performance differences can be expected clinically, but the exact degree of improvement would vary depending on the specifics of the imaging situation. Since lesion-detection is hampered in larger subjects due to increased attenuation, scatter, and randoms, we do not recommend reducing scan times in such subjects. Rather, the results of this study suggest that it may not be necessary to *increase* the scan time for larger subjects when TOF is used.

IV. Discussion

This study has performed a detailed characterization of the relationship between lesion-detection performance and scan time using experimental datasets that broadly encompass the range of typical activity and count levels that would be encountered by sites using a variety of scan protocols. The datasets include cases representative of administering from 8–15 mCi FDG with uptake times ranging from 1 to 2 hours. The data are broadly representative of oncologic PET imaging as a whole, offering results that are applicable to a wide range of imaging protocols; however, the results are also limited in that they encompass many different count rates and count levels, and thus offer less precise information regarding specific protocols than would be obtained from a study designed to investigate a single imaging protocol (but, similarly, would be less generally applicable).

Fig. 6 also includes results from a repeat analysis of the data including only scans 2–4 from each day, covering the more limited activity range of 5.3–8.4 mCi. These data are more indicative of what might be encountered for a single imaging site with a fixed target activity level and uptake period. The results for scans 2–4 closely match those for all scans 1–6. Slightly different values of A_{LROC} were obtained, but the relative performance of each algorithm as well as the dependence upon scan time were very similar. The biggest difference in these datasets was an increased uncertainty in the A_{LROC} estimates (error bars in Fig. 6 plots), due to the inclusion of fewer datapoints in the scan 2–4 analysis and hence greater statistical uncertainty in the results. These results suggest that the characterized tradeoff between scan time and detection performance observed here applies to a range of activity levels.

The generalized experiment was designed to complement previous work investigating the relationships between TOF, scan time, and oncologic lesion detection performance [20]–[22]. In [20], Surti and Karp studied detection of 10 mm diameter lesions in cylindrical phantoms with uniform backgrounds. Detectability was estimated using the non-prewhitening matched filter signal-to-noise ratio, and studied for scan times 2, 3, 4, and 5 min. The results showed improved detection performance for TOF vs. non-TOF at each scan

time for the cylindrical phantoms with uniform backgrounds. In [21] and [22], the work was extended to much more realistic datasets by manually inserting 10 mm diameter lesions into 100 patient studies with normal FDG biodistribution and no evidence of abnormal lesions. Two scan times were studied, 1 m and 3 m per bed position, and lesions were limited to a single size and contrast in the liver (contrast 3.5:1) and lungs (contrast 3.0:1). A full LROC study was performed with human observers in [21], and the channelized Hotelling numerical observer was used in [22]. These studies demonstrated improved performance for TOF vs. non-TOF images, improved lesion detectability for the longer scan time, and a greater improvement from TOF in larger patients (body mass index > 26) as compared to smaller patients.

This paper complements the previous work just summarized by investigating the relationship between scan time and lesion detectability in greater detail, by studying a range of lesion sizes (6–16 mm diam.), contrasts, and locations throughout the thorax, and by studying datasets that broadly cover the range of count levels that would be encountered for differing clinical protocols. The results of the current study are in agreement with the previous work, confirming that the improvements in detection performance offered by TOF and longer scan times are applicable to a broader range of imaging situations. The relationship between detection performance and scan time is elucidated in detail, providing broad guidance for sites considering reducing scan times when using PSF modeling or TOF systems. The study also provides new information on the relative value of TOF PET vs. PSF modeling for larger subjects. The degree of improvement for TOF was significantly higher than for PSF in this study, which contrasts with previous results in a smaller phantom where PSF and TOF were found to provide similar degrees of improvement.

Care must be taken in comparing the results of differing observer studies, especially with respect to the absolute value of the A_{LROC} and similar measures. For example, both the current study in the large phantom and our previous study using the smaller phantom [33] found the same rank performance of LOR, PSF, TOF, and PSF+TOF (listed in order of increasing performance). However, one should not compare the *absolute value* of A_{LROC} between these studies, as the lesion distributions (sizes, contrasts, locations) were not controlled between the studies. For example, one could arbitrarily shift A_{LROC} down or up by including or excluding a number of undetectable lesions. While this would shift the absolute value of A_{LROC} , it would not affect the rank-order results of the cases studied (unless the “undetectable” lesions were detectable by one or more of the cases). As such, one should not compare the absolute value of A_{LROC} from this study versus, say, [33], though the rank-order results of both studies can be meaningfully compared.

V. Summary and Conclusions

This paper has performed a systematic analysis of the relationship between reduced count levels (as a surrogate for reduced scan times) and focal lesion-detection performance using an extensive series of anthropomorphic phantom experiments modeling whole-body oncologic FDG PET imaging. The paper presents first results from a custom-built large thorax phantom approximately 65% larger than the previously-used phantom, facilitating assessment of the importance of TOF PET for a body size more representative of typical oncology patients. An extensive series of LROC studies was performed with the CNPW numerical observer to study and optimize selection of reconstruction parameters for four leading reconstruction approaches at eight scan times.

While the number of iterations maximizing detection performance was not sensitive to scan time, stronger post-reconstruction filters were required for the shorter scan times in order to manage the higher levels of noise. Lesion-detection performance was found to drop

significantly with shorter scan times for all four algorithms, the dependence of which was strongly modeled by an empirical logarithmic relationship. The relative performance of the four algorithms was consistent with previous results [32], [33] and related work studying the effects of TOF on PET lesion detection [21], [22], while also demonstrating a larger relative improvement of from TOF vs. PSF modeling for the larger phantom size in this study. The reconstructions with PSF modeling were less sensitive to filter strength, offering a practical advantage for routine use where it is not feasible to objectively optimize filter strength on a case-by-case basis.

For these data, the use of TOF provided equivalent lesion-detection performance with scans 40% shorter as compared to scanning without TOF. These results do not take into account other effects, such as reduced incidence of patient motion artifacts, that would also likely arise when reducing scan times in clinical practice. The characterization presented here may provide useful guidance for individual centers considering reducing whole-body PET/CT scan times for oncologic applications. Using modern tomographs with advanced reconstruction techniques, image quality can be improved while concurrently increasing patient throughput to some degree.

Acknowledgments

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering under Grant R03EB014454, in part by the National Cancer Institute under Grant R01CA107353, in part by Siemens Medical Solutions, and in part by the Ben B. and Iris M. Margolis Foundation.

The custom designed large thorax phantom was constructed by Radiology Support Devices, and the authors gratefully acknowledge their special efforts in constructing this unique phantom. The authors thank X. Liu for manufacturing the shell-less ^{68}Ge lesions used in this work, and C. Seegmiller and T. Bradshaw, whose meticulous efforts in crafting foam, filling the phantom, and assisting with the experiments were invaluable to this work. Finally, the authors thank the anonymous reviewers, whose comments and suggestions greatly improved the paper.

References

1. Metz CE. Basic principles of ROC analysis. *Sem Nucl Med.* 1978; 8:283–98.
2. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiol.* 1982; 143:29–36.
3. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol.* 1986; 21:720–33. [PubMed: 3095258]
4. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys.* 1996; 23:1709–25. [PubMed: 8946368]
5. Barrett HH, Abbey CK, Clarkson E. Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions. *J Opt Soc Amer A.* 1998; 15:1520–35.
6. Barrett, HH.; Myers, KJ. *Foundations of Image Science.* Hoboken, NJ: Wiley; 2004.
7. Gifford HC, King MA, Pretorius PH, Wells RG. A comparison of human and model observers in multislice LROC studies. *IEEE Trans Med Imag.* Feb; 2005 24(2):160–9.
8. Gifford HC, Kinahan PE, Lartizien C, King MA. Evaluation of multiclass model observers in PET LROC studies. *IEEE Trans Nucl Sci.* Jan; 2007 54(1):116–23. [PubMed: 18392119]
9. Farquhar TH, Llacer J, Hoh CK, Czernin J, Gambhir SS, Seltzer MA, Silverman DH, Qi J, Hsu C, Hoffman EJ. 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; 40:2043–52. [PubMed: 10616885]
10. Farquhar TH, Llacer J, Sayre J, Tai YC, Hoffman EJ. ROC and LROC analyses of the effects of lesion contrast, size, and signal-to-noise ratio on detectability in PET images. *J Nucl Med.* 2000; 41:745–54. [PubMed: 10768578]
11. Qi J, Huesman RH. Theoretical study of lesion detectability of MAP reconstruction using computer observers. *IEEE Trans Med Imag.* Aug; 2001 20(8):815–22.

12. Kadmas DJ, Christian PE. Comparative evaluation of lesion detectability for 6 PET imaging platforms using a highly reproducible whole-body phantom with (22)Na lesions and localization ROC analysis. *J Nucl Med.* 2002; 43:1545–54. [PubMed: 12411558]
13. Lartizen C, Kinahan PE, Swensson R, Comtat C, Lin M, Ville-magne V, Trebussen R. Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging. *J Nucl Med.* 2003; 44:276–90. [PubMed: 12571221]
14. Kim J-S, Kinahan PE, Lartizen C, Comtat C, Lewellen TK. A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging. *IEEE Trans Nucl Sci.* Feb; 2004 51(1):34–40.
15. Lartizen C, Kinahan PE, Comtat C. A lesion detection observer study comparing 2-dimensional versus fully 3-dimensional whole-body PET imaging protocols. *J Nucl Med.* 2004; 45:714–23. [PubMed: 15073270]
16. Lartizen C, Kinahan PE, Comtat C. Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography. *Acad Radiol.* 2004; 11:637–48. [PubMed: 15172366]
17. van Velden FH, Kloet RW, van Berckel BN, Wolfensberger SP, Lammertsma AA, Boellaard R. Comparison of 3D-OP-OSEM and 3D-FBP reconstruction algorithms for high-resolution research tomograph studies: Effects of randoms estimation methods. *Phys Med Biol.* 2008; 53:3217–30. [PubMed: 18506070]
18. Kadmas, DJ. Effect of reconstruction kernel width on optimal regularization for focal lesion detection in PET. *Proc. Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imag. Conf;* 2009. p. 2469–71.
19. Nuyts J, Michel C, Brepoels L, De Ceuninck L, Deroose C, Goffin K, Mottaghy FM, Stroobants S, Riet JV, Verscuren R. Performance of MAP reconstruction for hot lesion detection in whole-body PET/CT: An evaluation with human and numerical observers. *IEEE Trans Med Imag.* Jan; 2009 28(1):67–73.
20. Surti S, Karp JS. Experimental evaluation of a simple lesion detection task with time-of-flight PET. *Phys Med Biol.* 2009; 54:373–84. [PubMed: 19098351]
21. Surti S, Scheuermann J, El Fakhri G, Daube-Witherspoon ME, Lim R, Abi-Hatem N, Moussallem E, Benard F, Mankoff D, Karp JS. Impact of time-of-flight PET on whole-body oncologic studies: A human observer lesion detection and localization study. *J Nucl Med.* 2011; 52:712–9. [PubMed: 21498523]
22. El Fakhri G, Surti S, Trott CM, Scheuermann J, Karp JS. Improvement in lesion detection with whole-body oncologic time-of-flight PET. *J Nucl Med.* 2011; 52:347–53. [PubMed: 21321265]
23. Swensson RG. Using localization data from image interpretations to improve estimates of performance accuracy. *Med Decis Making.* 2000; 20:170–85. [PubMed: 10772355]
24. Khurd P, Gindi G. Fast LROC analysis of bayesian reconstructed emission tomographic images using model observers. *Phys Med Biol.* 2005; 50:1519–32. [PubMed: 15798341]
25. Khurd P, Gindi G. Rapid computation of LROC figures of merit using numerical observers (for SPECT/PET reconstruction). *IEEE Trans Nucl Sci.* Jun; 2005 52(3):618–26.
26. Popescu LM. Nonparametric ROC and LROC analysis. *Med Phys.* 2007; 34:1556–64. [PubMed: 17555237]
27. Fiete RD, Barrett HH, Smith WE, Myers KJ. Hotelling trace criterion and its correlation with human-observer performance. *J Opt Soc Amer A.* 1987; 4:945–53. [PubMed: 3598746]
28. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Amer A.* 1987; 4:2447–57. [PubMed: 3430229]
29. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Nat Acad Sci.* 1993; 90:9758–65. [PubMed: 8234311]
30. Wollenweber SD, Tsui BMW, Lalush DS, Frey EC, LaCroix KJ, Gullberg GT. Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging. *IEEE Trans Nucl Sci.* Dec; 1999 46(6):2098–2103.
31. Gifford HC, King MA, de Vries DJ, Soares EJ. Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. *J Nucl Med.* 2000; 41:514–21. [PubMed: 10716327]

32. Kadrmas DJ, Casey ME, Black NF, Hamill JJ, Panin VY, Conti M. Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes. *IEEE Trans Med Imag.* 2009; 28:523–34.
33. Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. *J Nucl Med.* 2009; 50:1315–23. [PubMed: 19617317]
34. Kadrmas, DJ. The Utah PET lesion detection database. *Proc. Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*; 2009. p. 3693-99.
35. Budinger TF. Time-of-flight positron emission tomography: Status relative to conventional PET. *J Nucl Med.* 1983; 24:73–8. [PubMed: 6336778]
36. Tomitani T. Image reconstruction and noise evaluation in photon time-of-flight assisted positron emission tomography. *IEEE Trans Nucl Sci.* Dec; 1981 28(6):4581–9.
37. Wong WH, Mullani NA, Philippe EA, Hartz R, Gould KL. Image improvement and design optimization of the time-of-flight PET. *J Nucl Med.* 1983; 24:52–60. [PubMed: 6600276]
38. Hamill, JJ.; Arnsdorff, E.; Casey, ME.; Liu, X.; Raulstron, WJ. A Ge-68 PET hot sphere phantom with no cold shells. *Proc. Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imag. Conf*; 2005. p. 1606-10.
39. Reader AJ, Erlandsson K, Flower MA, Ott RJ. Fast accurate iterative reconstruction for low-statistics positron volume imaging. *Phys Med Biol.* 1998; 43:835–46. [PubMed: 9572508]
40. Huesman RH, Klein GJ, Moses WW, Qi J, Reutter BW, Virador PR. List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling. *IEEE Trans Med Imag.* May; 2000 19(5):532–7.
41. Vandenberghe, S.; 'Asseler, YD.; Van de Walle, R.; Lemahieu, I. Correction for external LOR effects in listmode reconstruction for PET. *Proc. Conf. Rec. IEEE Int. Symp. Biomed. Imag*; 2002. p. 537-40.
42. Kadrmas DJ. LOR-OSEM: Statistical PET reconstruction from raw line-of-response histograms. *Phys Med Biol.* 2004; 49:4731–44. [PubMed: 15566171]
43. Panin VY, Kehren F, Michel C, Casey M. Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE Trans Med Imag.* Jul; 2006 25(7):907–21.
44. Jakoby BW, Bercier Y, Conti M, Casey ME, Bendriem B, Townsend DW. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. *Phys Med Biol.* 2011; 56:2375–89. [PubMed: 21427485]

Fig. 1.

The whole-body phantoms include a brain compartment; thorax with liver, lungs, and rib cage/spine; and pelvis with bladder. The upper-left panel shows the large and medium thorax phantoms together; note that while the rib cage assembly is the same size in both phantoms, it is embedded in the wall of the medium phantom but fits within the fillable cavity of the large phantom (*e.g.*, bottom-left panel). The example CT images at bottom-right reveal the heterogeneity of the lungs and soft tissue compartments, arising from the use of nylon mesh bags and low water resistance open cell foam in these compartments, respectively. These inhomogeneous structures provide a more realistic and challenging model for oncologic FDG PET lesion detections studies.

Fig. 2.

Maximum intensity projection images from one day of experiment with 6 sequential scans acquired over a period of two ^{18}F half-lives. Note the increase in statistical noise (*e.g.*, in the liver) in moving from the first scan (left) to the sixth scan (right), due to reduced count levels as the ^{18}F background decays. The images also clearly show the increase in lesion contrast for each successive scan, arising because the ^{68}Ge lesions remain essentially constant for all scans as the ^{18}F background decays—providing multiple lesion contrasts for increased statistical power in the observer studies.

Fig. 3.

Overview of results for optimizing number of iterations (top row) and filter strength (bottom row) for each reconstruction algorithm and scan time. The plots on the top row provide the results as a function of iteration, where the optimal filter for each iteration was used; similarly, the plots on the bottom row show results as a function of filter, where the iteration that maximized A_{LROC} for that filter was used. Lesion-detection performance was only moderately dependent upon number of iterations, provided that sufficient iterations were performed to reach near-peak performance (requiring approximately 4 iterations minimum for each case). However, performance was more strongly dependent on filter strength, with the optimal filter varying with both scan time and reconstruction algorithm.

Fig. 4.

Example reconstructed images for each algorithm and scan time (per bed position), showing a slice in the mediastinum with a 10 mm diameter warm lesion in the right lung. Each image is shown for the best number of iterations and filter strength for each case based on the results shown in Fig. 3. Visualization of the lesion in the right lung varies dramatically across this range of images, and is representative the lesion-detection study results obtained for all algorithms and scan times.

Fig. 5.

LROC curves for each algorithm and all scan times. The vertical axis is the correctly-localized true positive fraction (TPF), and the horizontal axis is the false positive fraction (FPF). The intersection with $\text{FPF} = 1.0$ equals the probability of correct localization for each case; *i.e.*, fraction of lesions that were correctly localized by the scanning CNPW observer. The curves were computed using the Mann-Whitney-Wilcoxon empirical approach, rather than a parametric fitting method, so that the effective sampling of the curves is revealed in the plots—each discontinuous transition or “bump” represents the curve’s crossing of the CNPW rating for an individual lesion.

Fig. 6.

Lesion-detection performance, as measured by A_{LROC} , plotted as a function of scan time for the four reconstruction cases studied. The left panel shows data from all scans 1–6, covering a broad range of activity levels and count densities, whereas the right panel includes only scans 2–4 from each day, representing a narrower range of activity levels that might be encountered in a single institution. Error bars show the standard deviation over 10,000 bootstrap resampling trials for each datapoint. The dashed lines show logarithmic fits of the data, revealing an empirical logarithmic relationship between detection performance and scan time in this study. The marked improvement in A_{LROC} for the reconstructions with TOF, as compared to those without TOF, reflects the value of TOF PET in larger subjects such as the large phantom used for this study. The horizontal dashed line marks the performance of LOR-OSEM at 240 s, and crosses the curve for TOF at 145 s. This demonstrates that TOF offered equivalent lesion-detection performance as the non-TOF reconstruction with a 40% reduction in scan time.

\$watermark-text

\$watermark-text

\$watermark-text

TABLE I

Activity Levels and Lesion Contrast by Scan.

Scan	Start Time (h:mm:ss) ¹	Decay Factor	Total Act. (mCi) ²	Soft Tissue Act. Conc. (μCi/cc)	³ Soft Tissue Lesion T:B
1	0:00:00	1.000	10.6	0.265	1.9:1
2	0:36:35	0.794	8.4	0.210	2.4:1
3	1:13:11	0.630	6.7	0.168	3.0:1
4	1:49:46 ²	0.500	5.3	0.133	3.8:1
5	2:26:22	0.397	4.2	0.105	4.8:1
6	3:02:57	0.315	3.4	0.085	5.9:1

¹ ^{18}F T_{1/2} = 109.77m = 1:49:46

²Total Activity present in phantom at start of the scan.

³Target:background ratio for hot lesion in soft-tissue compartment.