

Simultaneous Estimation of Bias and Resolution in PET Images With a Long-Lived “Pocket” Phantom System

Paul E. Kinahan¹, Darrin W. Byrd¹, Brian Helba², Kristen A. Wangerin¹, Xiaoxiao Liu², Joshua R. Levy³, Keith C. Allberg⁴, Karthik Krishnan², and Ricardo S. Avila⁵

¹Imaging Research Laboratory, University of Washington, Seattle, WA; ²Kitware, Inc., Clifton Park, NY; ³The Phantom Laboratory, Salem, NY; ⁴RadQual, LLC, Weare, NH; and ⁵Accumetra, LLC, Clifton Park, NY

Corresponding Author:

Paul Kinahan, PhD

Box 357987, University of Washington, Seattle, WA 98105;

E-mail: kinahan@uw.edu

Key Words: quantitative positron emission tomography imaging, calibration, standardized uptake value, clinical trials

Abbreviations: Positron emission tomography (PET), computed tomography (CT), standardized uptake values (SUVs), full-width half-maximum (FWHM), point-spread function (PSF), ordered-subsets expectation-maximization (OSEM), 3D filtered-backprojection (FBP), regions of interest (ROIs)

ABSTRACT

A challenge in multicenter trials that use quantitative positron emission tomography (PET) imaging is the often unknown variability in PET image values, typically measured as standardized uptake values, introduced by intersite differences in global and resolution-dependent biases. We present a method for the simultaneous monitoring of scanner calibration and reconstructed image resolution on a per-scan basis using a PET/computed tomography (CT) “pocket” phantom. We use simulation and phantom studies to optimize the design and construction of the PET/CT pocket phantom (120 × 30 × 30 mm). We then evaluate the performance of the PET/CT pocket phantom and accompanying software used alongside an anthropomorphic phantom when known variations in global bias ($\pm 20\%$, $\pm 40\%$) and resolution (3-, 6-, and 12-mm postreconstruction filters) are introduced. The resulting prototype PET/CT pocket phantom design uses 3 long-lived sources (15-mm diameter) containing germanium-68 and a CT contrast agent in an epoxy matrix. Activity concentrations varied from 30 to 190 kBq/mL. The pocket phantom software can accurately estimate global bias and can detect changes in resolution in measured phantom images. The pocket phantom is small enough to be scanned with patients and can potentially be used on a per-scan basis for quality assurance for clinical trials and quantitative PET imaging in general. Further studies are being performed to evaluate its performance under variations in clinical conditions that occur in practice.

INTRODUCTION

In oncology clinical trials and clinical practice, estimation of standardized uptake values (SUVs) of malignant lesions in positron emission tomography (PET) images can be used to assess response to therapy (1–6). Evaluation of response on a per-patient basis is central to the concept of precision medicine, which is prevention and treatment strategies that take individual variability into account (7). However, measured SUVs have a large degree of variability owing to physical and biological sources of error, as well as variations in image acquisition, processing, and analysis (8–10).

Important sources of variability are global shifts in SUVs due to scanner calibrations, operator error, or other reasons. During calibration, scanner sensitivity is typically measured by computing the number that scales PET images in arbitrary scanner units to match the known radiotracer concentration. SUV bias due to this scale factor, or calibration bias, is unstable even when measurements are repeated at a single site (11, 12). Fur-

ther, biases of key factors in the computation of SUVs, from PET scanners and dose calibrators, are not correlated and thus do not cancel out (11, 13).

A second important source of variability in PET SUVs is a size-dependent bias caused by resolution loss, often called the partial volume error (or effect) (14, 15). This is due to a combination of the intrinsic resolution of the PET acquisition (typically leading to 5-mm full-width half-maximum [FWHM] image resolution) and smoothing applied during image reconstruction to suppress noise. In addition, this bias increases as object size decreases, leading to the well-known recovery coefficient curves (14).

Many methods have been proposed for correction of partial volume effects (15), but attempts to recover signal lost in the imaging process are often constrained either by noise amplification (if they aim to restore high spatial frequencies) or the requirement that the exact lesion geometry and the scanner's resolution be known, so that the fraction of the lost signal can be

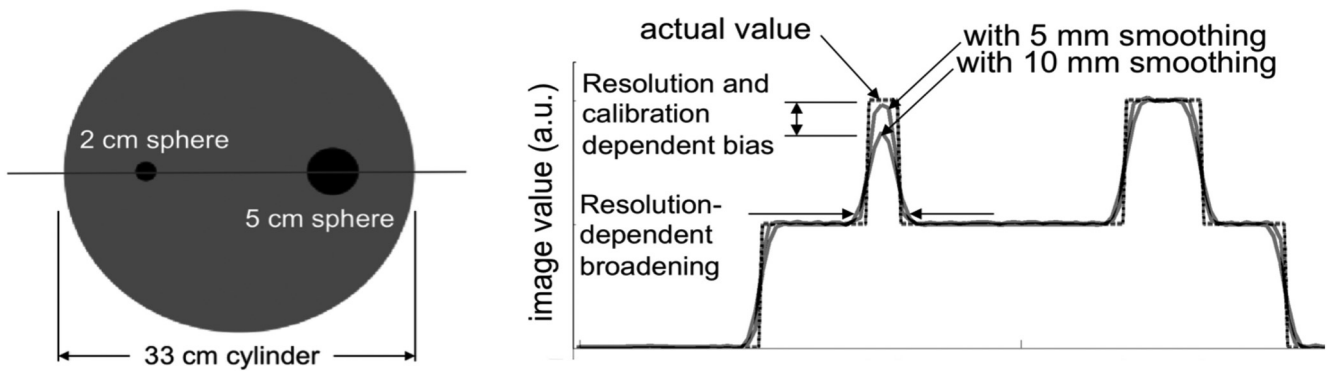


Figure 1. Noise-free simulation study of positron emission tomography (PET) acquisition and image reconstruction illustrating the effects of resolution on reconstructed signal. Left: True object indicating profile location. Right: Profile values for different image reconstruction smoothing parameters. The effect of intrinsic PET scanner resolution is also included.

determined. In practice, resolution is often unknown because of its complicated dependence on both user-selected parameters, which vary widely in practice (16-18), and variations in the image reconstruction methods, which are both proprietary and scanner-specific.

Although best-case PET image resolution is on the order of 5-mm FWHM, the final image resolution in practice is typically on the order of 10-mm FWHM or more. This means that a homogeneous spherical lesion would need to be larger than roughly 30 mm in diameter to avoid SUV bias at the lesion’s center. For objects <30 mm in diameter, it is not possible to tell if a measured bias is caused by resolution effects, global calibration effects, or a combination of the 2. These effects are illustrated in Figure 1 for the 20-mm sphere. This confounding mix of biases has likely hindered the use of small calibration sources in PET scanning, even though the idea has been proposed anecdotally for several decades.

These biases are important, as both scanner bias and image resolution are prone to vary, particularly in multicenter studies. Both will contribute to increased SUV variance if they are not carefully monitored. This can reduce study power in clinical trials that use SUVs as biomarkers (19).

In this study, we develop and evaluate a “pocket phantom” system using a source small enough to be imaged with a patient, which provides simultaneous estimation of the global bias and the final resolution of the image. “Pocket” connotes the compactness of the phantom—small enough to fit in one’s pocket—compared with current quality control phantoms. First we describe the algorithm used to estimate global bias and resolution. We then use simulation and phantom studies to optimize the design and construction of the PET/computed tomography (CT) pocket phantom. We then evaluate the performance of the PET/CT pocket phantom in practice when imaged alongside an anthropomorphic phantom and propose a method for the correction of SUVs in biased images.

METHODOLOGY

Pocket Phantom Estimator Method

The pocket phantom estimation process is summarized in Figure 2. The prototype, shown in Figure 3, borrows some features, including overall geometry, from a CT-specific phantom that also used spherical inclusions to estimate image properties (20). The phantom contains spherical radioactive regions of known size and activity. The phantom active regions contain solid epoxy

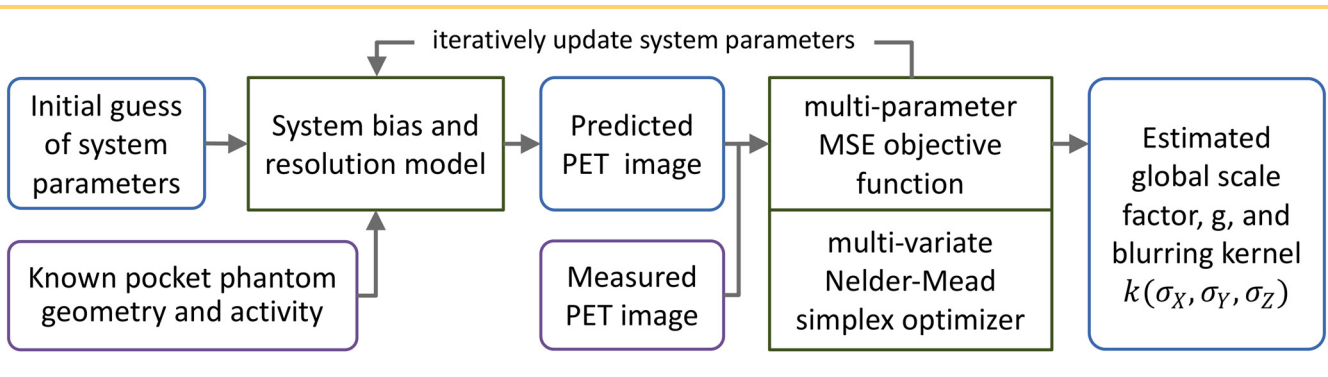


Figure 2. The pocket phantom estimation algorithm based on an iterative updating of model parameters comparing measured phantom images with predicted PET images generated from the known geometry and activity of the pocket phantom.

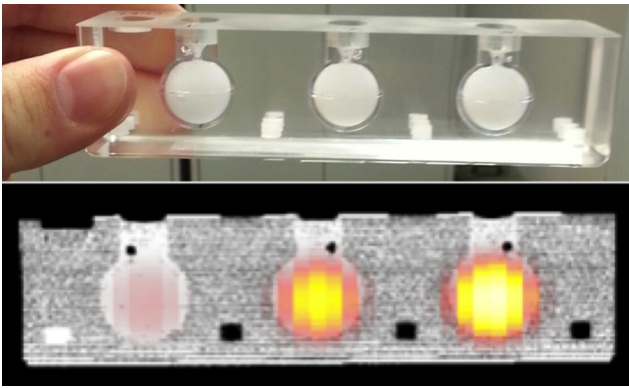


Figure 3. Top: Photograph of the PET/CT phantom prototype. Bottom: PET/CT image of prototype showing the 3 activity concentrations used.

infused with $^{68}\text{Germanium}/^{68}\text{Gallium}$ ($^{68}\text{Ge}/^{68}\text{Ga}$). This provides 2 advantages. First, the half-life of ^{68}Ge , which decays to ^{68}Ga , is 271 days. In turn, the ^{68}Ga decays by positron emission with a half-life of 68 minutes. This decay scheme makes $^{68}\text{Ge}/^{68}\text{Ga}$ a useful long-lived reference source, with replacement typically needed every 1–2 years. The phantom was manufactured with accurately specified radiotracer quantities that are National Institute of Standards and Technology (NIST)-traceable (21). Second, although a small phantom could be readily filled with ^{18}F in solution, there would be an additional variance added by difficulties of accurate calibration and operator variability in filling the phantom.

The overall algorithm models bias and resolution effects to produce a synthetic PET image from the known phantom geometry. The parameters of this model are then adjusted iteratively to match measured images of the pocket phantom. For the present study, images were converted from scanner-generated Digital Imaging and Communications in Medicine (DICOM) files or variables in MATLAB (MathWorks Inc., Natick, MA) to the Meta-Image format (22). Analysis was performed in VolView (23) and MATLAB.

Imaging Model. The imaging system model used here is expressed in equation (1):

$$I(x, y, z) = g \cdot p(x, y, z) * k(\sigma_x, \sigma_y, \sigma_z) + n(x, y, z), \quad (1)$$

where $I(x, y, z)$ is the 3-dimensional (3D) image generated by the PET scanner, g is a multiplicative global scale factor (ie, $g = 1$ means there is no global bias), and $p(x, y, z)$ is the true distribution of the PET signal source (ie, the physical concentration of radiotracer in the field of view). The function $k(\sigma_x, \sigma_y, \sigma_z)$ approximates the point-spread function (PSF) in the image with resolutions in the (x, y, z) directions given by standard deviations $(\sigma_x, \sigma_y, \sigma_z)$. Here the PSF is assumed to be both Gaussian and spatially-invariant. The 3D convolution operation is denoted by “*”, and $n(x, y, z)$ is an additive noise vector.

Estimator Algorithm. In words, the system produces an image that is a blurred and scaled version of the true PET tracer distribution in the phantom. If we can estimate the scale factor and the PSF, we can check for consistency between different imaging centers and in test-retest studies.

The estimator algorithm uses a synthetic sphere image generator function $s(\mathbf{x}_i, r_i, \rho_i)$, where i is the index for a specific sphere. For example, if 1 pocket phantom is used, then there are 3 spheres. The location, radius, and activity for the i -th sphere are given by $(\mathbf{x}_i, r_i, \rho_i)$. The radius and activity of each sphere are known a priori, and an initial estimate of the location of each sphere is obtained by segmenting the spheres from the CT image. Using the sphere image generator function, the predicted PET image is generated using the following equation:

$$\tilde{I}(x, y, z) = k(\sigma_x, \sigma_y, \sigma_z) * [\sum_{i=1}^N s(\mathbf{x}_i, r_i, g_i \cdot \rho_i)] \quad (2)$$

where $\tilde{I}(x, y, z)$ is the predicted noise-free sphere PET image, N is the number of spheres, and g_i is a scale factor for the activity of each sphere. A Nelder–Mead downhill simplex optimizer is used to estimate the standard deviations $(\sigma_x, \sigma_y, \sigma_z)$ and (\mathbf{x}_i, g_i) for $i = 1 \dots N$ by minimizing the mean squared difference objective function $\Phi(\sigma_x, \sigma_y, \sigma_z, \mathbf{x}_i, g_i) = |I(x, y, z) - \tilde{I}(x, y, z)|^2$ by using voxels in the neighborhood of the spheres. The algorithm terminates after a fixed number of iterations. The multiplicative global scale factor g in equation (1) is then estimated as the average of the individual sphere scaling factors g_i .

Pocket Phantom Design Study

Simulated and measured PET data were used to evaluate the performance of the algorithm. For a range of activity levels and sphere sizes, real and simulated phantom images were multiplied by scalars and smoothed with different filters to simulate variable scanner calibration and reconstruction settings. These tests led to the selection of design parameters for prototype pocket phantoms.

Design Study Using Simulated Data. As a first test of the estimator algorithm, a synthetic test object containing 2 spheres 15 mm in diameter having an activity concentration of 5 kBq/mL was simulated. Noise-free emission data sets (sinograms) for this object were generated using the University of Washington’s ASIM package (24). The detector configuration was modeled after a General Electric Discovery STE PET/CT scanner (General Electric Healthcare, Waukesha, Wisconsin). In MATLAB, the effects of detector parallax (25) and Poisson noise were added. Total detected coincidences were 4.8×10^5 (high noise) and 8.7×10^6 (low noise). Images were reconstructed with a fully 3D ordered-subsets expectation-maximization (OSEM) algorithm (26) or 3D filtered-backprojection (FBP) (27). For all OSEM reconstructions in this work, 4 iterations with 28 subsets were used. Voxel dimensions were $2.73 \times 2.73 \times 3.27$ mm for both OSEM and FBP. Further, 3 Gaussian postreconstruction smoothing filters (transaxial FWHM of 4, 8, and 12 mm) were also applied to the OSEM images. The axial filter FWHM for all OSEM images was 4.6 mm.

The resulting images were then rescaled such that the maximum signal was the same in each, creating images in which a calibration bias and resolution effects were mixed in ways unknown to the algorithm. The algorithm was then used to determine the image resolution parameters $(\sigma_x, \sigma_y, \sigma_z)$. As a check of the algorithm’s accuracy, the width of the user-specified postreconstruction filter was calculated by comparison with the PSF from an unfiltered image. This was done by assuming that the intrinsic PSF and filter width added in quadrature, such that $\sigma_{\text{additional}}^2 = \sigma_{\text{filtered}}^2 - \sigma_{\text{unfiltered}}^2$.

Fillable Testbed Phantom. To estimate the effect of sphere diameter, a cast urethane disc with fillable spheres was con-

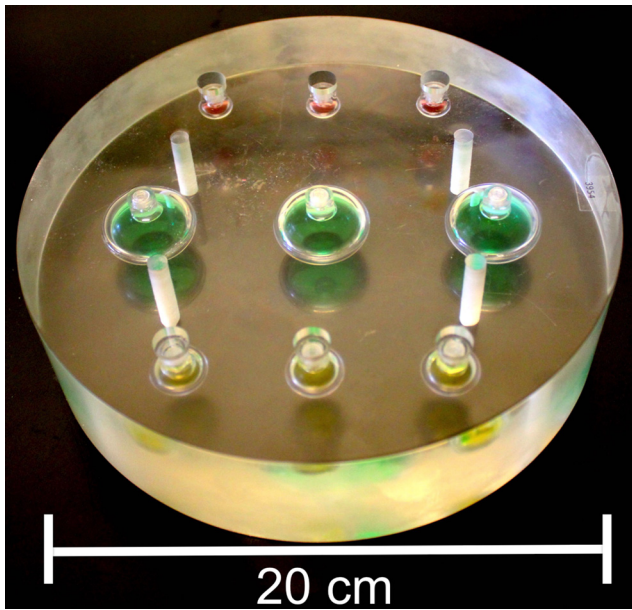


Figure 4. Fillable ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) phantom with 10-mm (top, red), 15-mm (bottom, yellow), and 30-mm (middle, green) fillable spheres.

constructed (Figure 4). The disc contained 3 spheres at each of 3 diameters (10, 15, and 30 mm) and was scanned on a General Electric Discovery STE PET/CT scanner. ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) was used as the radiotracer.

A single solution of ¹⁸F-FDG was used to fill all spheres, and the phantom was scanned with all sphere centers in a single transaxial plane. CT-based attenuation correction was performed using a 120-kV CT scan. Acquisitions and reconstructions varied as shown in Table 1. OSEM images were not filtered, while the FBP reconstruction used an 8.2-mm Hanning window. The axial voxel dimension, or slice width, was 3.27 mm for all images. The estimated scale factors g_i for all 9 spheres were recorded without averaging, and the bias and variance as a function of size across all 24 parameter sets were evaluated.

Table 1. Imaging Parameters for the Fillable Testbed Phantom of Figure 4

Imaging Parameter	Variation
Reconstruction algorithm	OSEM, FBP
Transaxial voxel dimension (mm)	2.73, 5.56
Detected events (millions)	0.5, 0.8, 1.6
Activity concentrations (kBq/mL)	6.0, 32.0
Sphere diameter (mm)	10, 15, 30

Abbreviations: OSEM, ordered-subsets expectation-maximization; FBP, 3D filtered-backprojection. Twenty four reconstructed images were generated, with each image having 3 spheres at each of 3 diameters.

Table 2. Imaging Parameters for Scanning of Prototype ⁶⁸Ge/⁶⁸Ga Phantoms With ¹⁸F-FDG Phantom (Figure 7)

Imaging Parameter	Variation
Reconstruction method	OSEM
Postreconstruction transaxial smoothing FWHM (mm)	3, 6, 12
Postreconstruction axial smoothing FWHM (mm)	4.6
Simulated global scale factor g	0.6, 0.8, 1.0, 1.2, 1.4

Abbreviations: OSEM, ordered-subsets expectation-maximization; ¹⁸F-FDG, ¹⁸F-fluorodeoxyglucose; FWHM, full-width half-maximum.

Testing of Pocket Phantom Prototypes

Based on the results from the simulated data and fillable phantom, 2 long-lived prototype pocket phantoms were constructed using epoxy infused with ⁶⁸Ge/⁶⁸Ga. Each phantom had 3 spheres of 15 mm diameter (Figure 3) in a rectangular 3- × 3- × 12-cm cast urethane block. The activity concentrations of the 3 spheres in the first phantom were 30, 74, and 118 kBq/mL. For the second phantom, the concentrations were 47, 109, and 190 kBq/mL.

Phantom Measurements. The prototype long-lived pocket phantoms were measured alongside an anthropomorphic phantom that contained 3 different concentrations of ¹⁸F-FDG radiotracer in 3 regions corresponding to liver, lung, and background. Scan parameters are shown in Table 2. The duration was 5 min and the voxel size was 2.73 × 2.73 × 3.27 mm. The mean signal intensity was measured in regions of interest (ROIs) in the anthropomorphic phantom.

Addition of Known Bias and Smoothing. To simulate multi-center clinical variability of scanner calibration and image resolution, we systematically varied the global scalar bias and postreconstruction filtering of our measured prototype phantom images. As shown in Table 2, the images had 3 levels of smoothing and 5 scale factors applied. These scale factors were applied after the PET/CT scanner had applied all physical corrections to the data to generate correctly calibrated images. We denote the image for the scan with the j -th applied scale factor and k -th filter width as $I^{jk}(x, y, z)$, and the estimated scale factor, after averaging over spheres, as g^{jk} .

For each image in our test-space of reconstructions, we generated a bias-corrected image, I_c^{jk} , according to equation (3).

$$I_c^{jk} = \frac{I^{jk}}{g^{jk}} \quad (3)$$

As a test of the pocket phantom system's ability to correct scanner calibration errors, we report ROI values from the corrected and uncorrected images.

Pocket Phantom Data Rescaling. It is known that scatter and attenuation correction can lead to bias in some solid phantoms (28). Our calculation of the scale factor g was therefore modified to use premeasured pocket phantom image data as a reference. As a test case, the reconstruction with a scale factor of 1.0 and a 12-mm post filter was used as a reference image. Scale factors g_i from the spheres in this scan were used as normalization factors

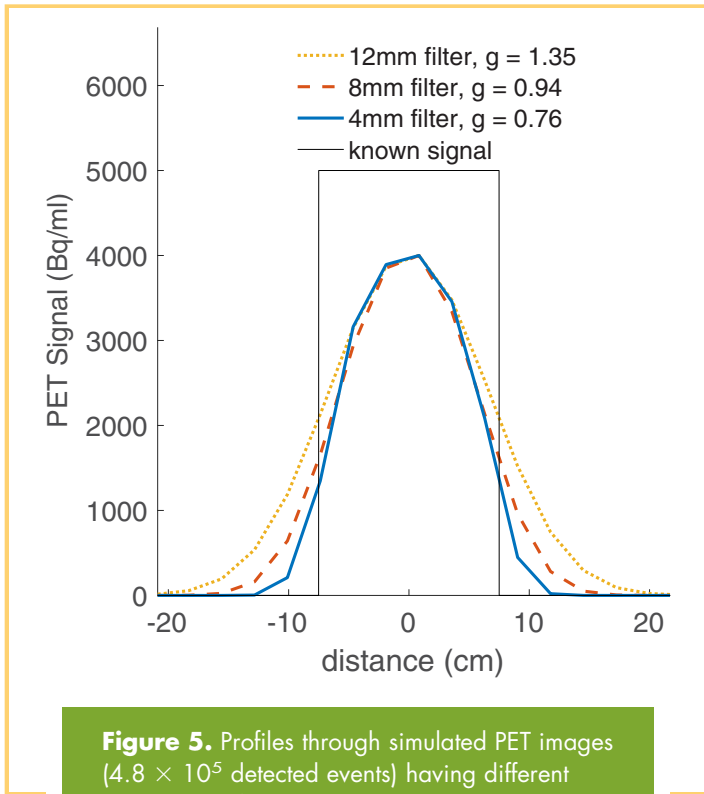


Figure 5. Profiles through simulated PET images (4.8×10^5 detected events) having different global scale factors and resolution losses that lead to the same maximum signal.

to calculate rescaled estimates of the scale factors for the corresponding spheres in all images.

RESULTS
Design Study Results

Simulation Results. Profiles through a subset of simulated phantom spheres are shown in Figure 5. The profiles confirm that bias from either resolution losses or global scaling are not unique. In other words, the same recovery coefficient can result from different combinations of global bias and resolution bias.

Table 3 shows the true and estimated values of the applied scale factor and applied transaxial filter width [equation (1)]. Estimates of the filter width in the axial direction, which was 4.6 mm, had a distribution of 4.57 (0.16) mm over all simulated

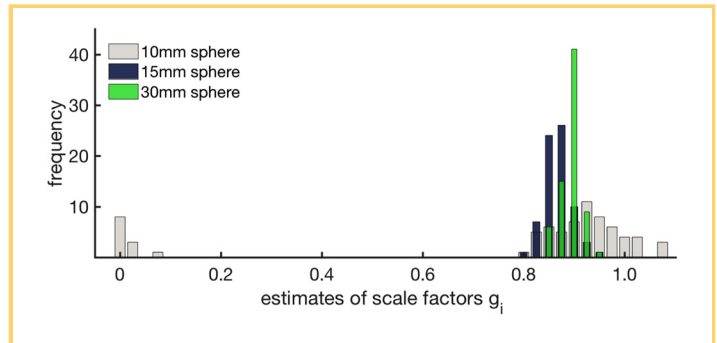


Figure 6. Estimated g_i for all spheres in the fillable testbed phantom over the parameter variations listed in Table 1.

images. The performance of the pocket phantom system was similar over all simulated parameters, including variations in sphere size (data not shown). In other words, the estimator algorithm accurately predicted the applied global scale factor and image smoothing.

Fillable Testbed Phantom. Figure 6 shows the distribution of g_i scale factor estimates for all spheres in the reconstructions listed in Table 1. In some cases, the algorithm returned anomalously low g_i values for the 10-mm sphere, indicating algorithm failure for this sphere size. For the 15-mm spheres, g_i had a mean of 0.868 (0.025) across all reconstructions. Performance of the algorithm with 30-mm spheres was similar. Bias estimates were stable as the reconstruction method changed. For the 15-mm sphere, g_i values were 0.872 (0.036) for OSEM images and 0.869 (0.014) for FBP.

For the 15- and 30-mm spheres, the distribution of resolution estimates are shown in Table 4. Here, reported statistics are over variations in image noise and activity concentration (rows 3 and 4 of Table 1). Changing the transaxial voxel sizes in OSEM images led to changes in transaxial resolution estimates. In the axial direction, for which voxel dimensions were the same for all reconstructions (3.27 mm), the agreement was better, with average estimates from OSEM images agreeing to within 0.8 mm as sphere size and voxel size varied. Resolution estimates from FBP images showed better agreement than OSEM.

Our testing indicated that the 15-mm sphere size was optimal based on its acceptable performance in simulated and physical

Table 3. Applied and Estimated Image Parameters of Simulations Having the Same Maximum Signal

Noise Level	Applied Global Scale Factor	Estimated Global Scale Factor	Applied Transaxial Filter (mm)	Estimated Transaxial Filter (mm)
High	0.76	0.79	4	5
High	0.94	0.95	8	8.5
High	1.35	1.36	12	12.4
Low	0.81	0.83	4	4.8
Low	0.97	0.98	8	8.5
Low	1.39	1.40	12	12.4

The “high noise” data correspond to the profile in Figure 5.

Table 4. Estimates of Final Image Resolution (in millimeters) From the 15- and 30-mm Spheres in the Fillable Testbed Phantom Reconstructions of Table 1

	15-mm Spheres		30-mm Spheres	
	2.73-mm Voxels	5.46-mm Voxels	2.73-mm Voxels	5.46-mm Voxels
OSEM				
Transaxial	3.69 (0.343)	1.42 (0.215)	3.98 (0.341)	2.24 (1.668)
Axial	3.70 (0.340)	4.47 (0.450)	4.23 (0.172)	4.37 (0.250)
FBP				
Transaxial	9.12 (0.163)	9.26 (0.261)	9.50 (0.191)	9.65 (0.147)
Axial	6.06 (0.098)	6.15 (0.096)	5.97 (0.093)	6.05 (0.138)

Abbreviations: OSEM, ordered-subsets expectation-maximization; FBP, 3D filtered-backprojection. Columns are labeled with transaxial voxel size. Axial voxel width (slice width) was 3.27mm for all images.

testing and the ease of manufacturing versus 30-mm spheres in the final phantom.

Pocket Phantom Results

Measured Data. Figure 7 shows the scan configuration and representative data from the pocket phantom prototype measurements acquired with the anthropomorphic chest phantom. This scan roughly represents the intended clinical scan configuration with the pocket phantoms below the patient. The PET images and profile show that the pocket phantom images have excellent signal-to-noise properties and match the magnitude of signal in the anthropomorphic phantom.

Table 5 shows the PET signal measured in images created with the parameters of Table 2 before and after correction by equation (3). Expressed as a percentage of the range midpoint, ranges of mean ROI signal were reduced from 80% in uncorrected images to <5% for corrected ones, indicating that the pocket phantom system successfully compensated for the simulated scanner miscalibration in our test image set.

Figure 8 shows the measured ROI values (A_{ROI}) for the pocket phantom spheres after division by known activity concentration. The differing slopes for A_{ROI} show the dependence of partial volume effects on the variable image resolution. The square A_{Cal} markers represent the ratio of the applied scale factor (Table 2) to the estimated scale factor g . A value of 1 for A_{Cal} therefore

corresponds to the accurate estimation of bias. After averaging over the 6 pocket phantom spheres in the images, A_{Cal} values ranged from 0.95 to 1.06, indicating that the bias-corrected images were accurate to within 6% regardless of the changes in image filtering or global image bias.

As the reconstruction postfilter width varied between 3, 6, and 12 mm, estimates of final transaxial, or transverse, resolution varied as in Table 6. These estimates of final image resolution include effects of both the postfilter and intrinsic PSF. The small standard deviations demonstrate that transverse resolution estimates are stable as global scaling varies. In addition, axial resolution estimates are stable as transverse resolution and global scaling vary.

DISCUSSION

We have tested and evaluated design parameters for small phantoms that allow the simultaneous estimation of scanner global calibration bias and reconstructed image resolution. We have constructed and tested a prototype phantom on the basis of these results, and have demonstrated the ability of the phantom and software to detect changes in the bias and resolution of measured images. For the prototype phantom, the 15-mm spheres were chosen based on their providing similar performance to the 30mm spheres while allowing the phantom itself to be smaller.

The algorithm succeeded in estimating global bias independently of resolution. In particular, Table 3 shows that the variations of parameters shown in Figure 5 have been successfully separated. Table 5 shows that the range of signal biases in our set of test images was reduced to <5% using the pocket phantom correction factors regardless of changes in the applied postreconstruction smoothing. Further, bias estimates did not show any dependence on the image reconstruction method. The global scale factor for the 15-mm sphere had a coefficient of variation of <3% over all instances of parameter variations shown in Table 1. The agreement of bias estimates for these very different reconstructions suggests that the Gaussian model used by the estimator algorithm can accommodate a range of resolutions and reconstruction methods.

The absolute accuracy of bias estimates is more difficult to evaluate. In the simulated data, for which bias was known, the pocket phantom system found the global scale factor to within 3% of the true value for all resolutions tested (Table 3). In PET/CT measurements of epoxy-based solid phantoms, the PET

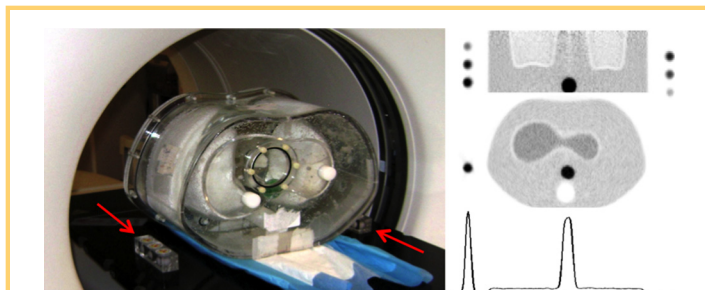


Figure 7. Left: Pocket phantoms, identified by red arrows, in scan configuration with an anthropomorphic phantom. Right: PET images and a profile through the transaxial image. The hot spot in the center of the phantom is a 30-mm sphere used as a separate test object.

Table 5. Ranges of Measured Signal (kBq/ml) in Biased and Corrected Images

Applied Smoothing	Background Region		Liver Region	
	Original ROI Values	Corrected ROI Values	Original ROI Values	Corrected ROI Values
3 mm	1.27–2.97	2.07–2.14	4.96–11.58	8.05–8.33
6 mm	1.28–2.98	2.10–2.20	4.97–11.60	8.20–8.60
12 mm	1.28–2.98	2.09–2.15	4.99–11.63	8.15–8.38

Abbreviations: ROI, region of interest. ROI mean values are shown.

image value is known to be biased owing to attenuation correction that is not correct for synthetic materials (28). Although our scanner was carefully calibrated, Figure 6 shows global scale factor estimates were generally less than one. ROI measurements of activity in the centers of the largest spheres in the urethane fillable testbed phantom, which were not subject to partial volume effects, showed that this bias was real and not a failure of the algorithm. This prevents us from computing scanner calibration bias directly from the known radiotracer concentration. To correct this problem in our solid prototypes and future work, we have proposed and tested the use of a calibration prescan (see Section *Pocket Phantom Data Rescaling*.) where the algorithm is precalibrated to compensate for biases in the pocket phantom signal from physical effects such as attenuation and scatter correction. With this method, the impact of scatter and attenuation correction on the pocket phantom is assumed to be constant for a given scanner. The A_{Cal} data in Figure 8 show that for our initial tests, the precalibration led to accurate correction of our simulated global image bias.

Unlike calibration bias, resolution effects cannot be easily corrected. Partial volume correction methods have been proposed, but these have been shown to add bias and variance (15, 29). However, if changes in resolution can be detected, this information can help with quality control either for clinical practice or clinical trials in which the quantitative accuracy of PET images is relied upon. For example, in clinical trials, the removal of data with uncontrolled biases, including those due to resolution, can increase the study power even if the sample size decreases (19). In our measured data (Table 6), the pocket phantom system returned estimates that were well separated when resolution was varied, with standard deviations of 0.01 and 0.09 mm for the 3- and 6-mm postreconstruction filtering, respectively. Importantly, these results were stable even when global scaling was varied by up to $\pm 40\%$ (Figure 8).

Currently, efforts to reduce variability in PET mainly consist of accreditation procedures (30) and consensus documents on best practices (31–33). Scanner accreditation often involves “cross calibration,” in which dose calibrator and scanner measurements are required to concur, but this process may not ensure biases are stable over time (13).

Resolution may be addressed by specifying a range of acceptable signal bias for a range of lesion sizes (34) or by requiring visibility of specific features of a given size (30). Methods for quantifying resolution in the literature vary and may involve profiles through FBP images of point sources near the scanner’s center (35), ROI signal from multiple sphere sizes in a large calibration phantom (36), or solving for the radial PSF in Fourier space (37). However, we note that none of these methods is compatible with a clinical scan with a patient in the field of view.

With its unique combination of software and manufacturing, the pocket phantom system aims to provide new capabilities

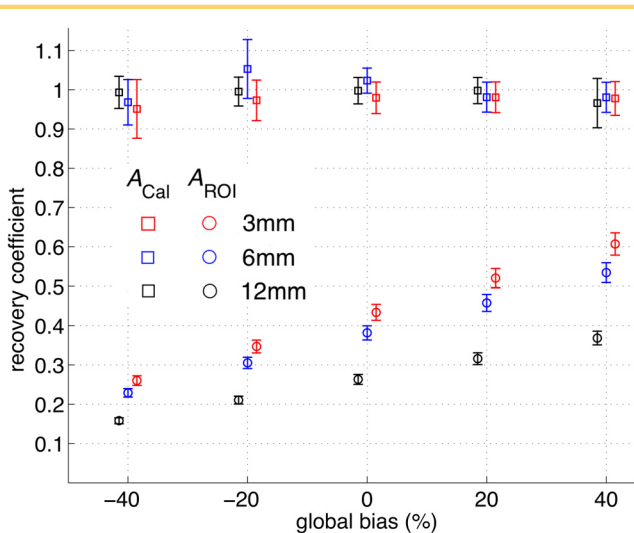


Figure 8. For varying applied bias and smoothness, recovery of signal (measured/known) for a region of interest (ROI) on reconstructed images (A_{ROI}) and theoretical residual calibration bias in images after applying corrective factors from the pocket phantoms (A_{Cal}).

Table 6. Estimated Image Resolution (in millimeters) for Images Generated Using Parameters of Table 2

Reconstruction Smoothing (mm)	Estimated Transverse Resolution (mm)	Estimated Axial Resolution (mm)
3	7.46 (0.01)	7.06 (0.01)
6	9.43 (0.09)	7.18 (0.18)
12	14.16 (0.07)	7.09 (0.04)

in PET quality control. The long-lived phantoms provide a more stable signal than the manually-filled phantoms used in cross calibration. The spherical symmetry of the active regions allows estimates of resolution along 3 independent directions, regardless of the phantom orientation. In particular, the spherical design offers an advantage over line sources, from which axial resolution cannot be estimated. In addition, the software modeling allows the phantoms to be small enough to be scanned with patients, enabling quality control during patient scans.

Future work will address the practical requirements for translating our initial results into a more widely useable quality control system. We have already published the preliminary results on our user-facing software that will make the algorithm available to off-site imagers (38). In addition, a more detailed subsequent analysis of the phantom performance, including the dependence on scan configuration and radiotracer concentrations, will allow us to optimize the protocol for phantom scanning and finalize the manufacturing parameters.

Our study has some limitations. The global bias due to CT-based attenuation correction of the epoxy-based phantom, and the

precalibration workaround, have already been discussed. The dependence of resolution estimates on voxel size seen in Table 4 is likely due to the way the model images are downsampled before the smoothing of equation (3). In cases where voxel dimensions approach the resolution, the effect of downsampling may become significant and lead to unreliable resolution estimates. We note that for the more heavily smoothed FBP images, this problem did not occur. Our initial evaluation of the pocket phantom system was limited to a single scanner. Future work will include repeated measurements on different makes and models of scanners.

The pocket phantom system can estimate and correct changes in calibration bias in measured PET images, and it can simultaneously detect changes in the reconstructed image resolution. Over the imaging scenarios tested, the system returned stable estimates of both bias and resolution, as long as voxel size was not too large. This suggests that the pocket phantom system is a viable method for quality assurance in PET, particularly in clinical trials. However, the robustness of the imaging model should be further investigated for multiple imaging systems.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants U01CA148131, R01CA042593, R41CA167907, and R42CA167907.

Disclosure: Dr. Kinahan reports grants from GE Healthcare. Dr. Wangerin is an employee of GE Healthcare. Mr. Avila reports work with the Radiological Society of North

America as well as pending patents on a table top image calibration phantom and an automated scan quality monitoring system.

Conflict of Interest: None Reported.

REFERENCES

- Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov.* 2003;2:566–580.
- Mankoff DA, Pryma DA, Clark AS. Molecular imaging biomarkers for oncology clinical trials. *J Nucl Med.* 2014;55:525–528.
- O'Connor JP, Abaogye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Fairv-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh DM, Lacombe D, Lambin P, Lassau N, Leach MO, Lee TY, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJ, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Strobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol.* 2017;14:169–186.
- Takahashi R, Hirata H, Tachibana I, Shimosegawa E, Inoue A, Nagatomo I, Takeda Y, Kida H, Goya S, Kijima T, Yoshida M, Kumagai T, Kumanogoh A, Okumura M, Hatazawa J, Kawase I. Early 18F]fluorodeoxyglucose positron emission tomography at two days of gefitinib treatment predicts clinical outcome in patients with adenocarcinoma of the lung. *Clin Cancer Res.* 2012;18:220–228.
- Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 Suppl 1:122S–50S.
- Weber WA, Petersen V, Schmidt B, Tyndale-Hines L, Link T, Peschel C, Schwaiger M. Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol.* 2003;21:2651–2657.
- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372:793–795.
- Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol.* 2010;195:310–320.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med.* 2009;50 Suppl 1:11S–20S.
- Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR.* 2010;31:496–505.
- Doot RK, Pierce LA, Byrd D, Elston B, Allberg KC, Kinahan PE. Biases in multicenter longitudinal PET standardized uptake value measurements. *Transl Oncol.* 2014;7:48–54.
- Lockhart CM, MacDonald LR, Alessio AM, McDougald WA, Doot RK, Kinahan PE. Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. *J Nucl Med.* 2011;52:218–224.
- Byrd D, Christopfel R, Arabasz G, Catana C, Karp J, Lodge MA, Laymon C, Moros EG, Budzevich M, Nehmeh S, Scheuermann J, Sunderland J, Zhang J, Kinahan P. Measuring temporal stability of positron emission tomography standardized uptake value bias using long-lived sources in a multicenter network. *J Med Imaging (Bellingham).* 2018;5:011016.
- Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr.* 1979;3:299–308.
- Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932–945.
- Beyer T, Czernin J, Freudenberger LS. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. *J Nucl Med.* 2011;52:303–310.
- Byrd D, Christopfel R, Buatti J, Moros E, Nehmeh S, Opanowski A, Kinahan P. Multicenter survey of PET/CT protocol parameters that affect standardized uptake values. *J Med Imaging (Bellingham).* 2018;5:011012.
- Graham MM, Badawi RD, Wahl RL. Variations in PET/CT methodology for oncologic imaging at U.S. academic medical centers: an imaging response assessment team survey. *J Nucl Med.* 2011;52:311–317.
- Kurland BF, Doot RK, Linden HM, Mankoff DA, Kinahan PE. Multicenter trials using (11)8F-fluorodeoxyglucose (FDG) PET to predict chemotherapy response: effects of differential measurement error and bias on power calculations for unselected and enrichment designs. *Clin Trials.* 2013;10:886–895.
- Henschke CI, Yankelevitz DF, Yip R, Archer V, Zahlmann G, Krishnan K, Helba B, Avila R. Tumor volume measurement error using computed tomography imaging in a phase II clinical trial in lung cancer. *J Med Imaging (Bellingham).* 2016;3:035505.
- Zimmerman BE, Cessna JT. Development of a traceable calibration methodology for solid (68)Ge/(68)Ga sources used as a calibration surrogate for (18)F in radionuclide activity calibrators. *J Nucl Med.* 2010;51:448–453.

22. MetalO. <https://itk.org/Wiki/ITK/MetalO/Documentation> (Accessed 2018-01-21).
23. VolView. <https://www.kitware.com/volview/> (Accessed :2018-01-21).
24. Comtat C, Kinahan PE, Defrise M, Michel C, Lartizien C, Townsend DW. Simulating whole-body PET scanning with rapid analytical methods. *Proceedings: 1999 IEEE Nuclear Science Symposium and Medical Imaging Conference*. 1999;3:1260–1264.
25. Alessio AM, Stearns CW, Tong S, Ross SG, Kohlmyer S, Ganin A, Kinahan PE. Application and evaluation of a measured spatially variant system model for PET image reconstruction. *IEEE Trans Med Imaging*. 2010;29:938–949.
26. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging*. 1994;13:601–609.
27. Kinahan PE, Rogers JG. Analytic 3D image-reconstruction using all detected events. *IEEE Transactions on Nuclear Science*. 1989;36:964–968.
28. Byrd D, Sunderland J, Allberg K, Kinahan P. Attenuation effects in solid phantoms used for PET/CT scanner calibration. *J Nucl Med*. 2014;55:2095.
29. Byrd D, Alessio A, Stearns C, Ross S, Ganin A, Kinahan P. Characterization of partial volume errors in PET through analytic simulation. *J Nucl Med*. 2013;54:2127.
30. MacFarlane CR; American College of Radiologists. ACR accreditation of nuclear medicine and PET imaging departments. *J Nucl Med Technol*. 2006;34:18–24.
31. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, Verzijlbergen FJ, Barrington SF, Pike LC, Weber WA, Stroobants S, Delbeke D, Donohoe KJ, Holbrook S, Graham MM, Testanera G, Hoekstra OS, Zijlstra J, Visser E, Hoekstra CJ, Pruim J, Willemsen A, Arends B, Kotzerke J, Bockisch A, Beyer T, Chiti A, Krause BJ; European Association of Nuclear Medicine (EANM). FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–354.
32. Graham MM, Wahl RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, Perlman ES, Kinahan PE, Christian PE, Hoekstra OS, Dorfman GS. Summary of the UPICT protocol for FDG PET/CT imaging in oncology clinical trials. *J Nucl Med*. 2015.
33. (QIBA) QIBA. FDG-PET/CT Biomarker Committee: FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy Profile. Technically Confirmed Version. Version 1.13. Available from: [RSNA.ORG/QIBA](https://www.rsna.org/qiba). 2016.
34. Boellaard R, Oyen WJ, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, Arends B, Verzijlbergen FJ, Zijlstra J, Paans AM, Comans EF, Pruim J. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging*. 2008;35:2320–2333.
35. Daube-Witherspoon ME, Karp JS, Casey ME, DiFilippo FP, Hines H, Muehlethner G, Simcic V, Stearns CW, Adam IE, Kohlmyer S, Sossi V. PET performance measurements using the NEMA NU 2-2001 standard. *J Nucl Med*. 2002;43:1398–1409.
36. Prieto E, Martí-Climent JM, Arbizu J, Garrastachu P, Domínguez I, Quincoces G, García-Velloso MJ, Lecumberri P, Gómez-Fernández M, Richter JA. Evaluation of spatial resolution of a PET scanner through the simulation and experimental measurement of the recovery coefficient. *Comput Biol Med*. 2010;40:75–80.
37. Lodge MA, Rahmim A, Wahl RL. A practical, automated quality assurance method for measuring spatial resolution in PET. *J Nucl Med*. 2009;50:1307–1314.
38. Zukic D, Mullen Z, Byrd D, Kinahan P, Enquobahrie A. A web-based platform for a high throughput calibration of PET scans. *Computer Assisted Radiology and Surgery*. 2016;11:S29–S30.