Translational Oncology
www.transonc.com
Volume 7 Number 1 February 2014 pp. 48–54 48

Biases in Multicenter Longitudinal PET Standardized Uptake Value Measurements[1]

Robert K. Doot*, Larry A. Pierce II†, Darrin Byrd†, Brian Elston†, Keith C. Allberg‡ and Paul E. Kinahan†

*Department of Radiology, University of Pennsylvania, Philadelphia, PA; †Department of Radiology, University of Washington, Seattle, WA; ‡RadQual, LLC, Weare, NH

Abstract

This study investigates measurement biases in longitudinal positron-emission tomography/computed tomography (PET/CT) studies that are due to instrumentation variability including human error. Improved estimation of variability between patient scans is of particular importance for assessing response to therapy and multicenter trials. We used National Institute of Standards and Technology–traceable calibration methodology for solid germanium-68/gallium-68 (^{68}Ge/^{68}Ga) sources used as surrogates for fluorine-18 (^{18}F) in radionuclide activity calibrators. One cross-calibration kit was constructed for both dose calibrators and PET scanners using the same 9-month half-life batch of ^{68}Ge/^{68}Ga in epoxy. Repeat measurements occurred in a local network of PET imaging sites to assess standardized uptake value (SUV) errors over time for six dose calibrators from two major manufacturers and for six PET/CT scanners from three major manufacturers. Bias in activity measures by dose calibrators ranged from −50% to 9% and was relatively stable over time except at one site that modified settings between measurements. Bias in activity concentration measures by PET scanners ranged from −27% to 13% with a median of 174 days between the six repeat scans (range, 29 to 226 days). Corresponding errors in SUV measurements ranged from −20% to 47%. SUV biases were not stable over time with longitudinal differences for individual scanners ranging from −11% to 59%. Bias in SUV measurements varied over time and between scanner sites. These results suggest that attention should be paid to PET scanner calibration for longitudinal studies and use of dose calibrator and scanner cross-calibration kits could be helpful for quality assurance and control.

Translational Oncology (2014) 7, 48–54

Introduction

Molecular imaging with combined positron-emission tomography (PET) and X-ray computed tomography (CT) scanners has become a standard component of diagnosis and staging in oncology [1,2]. In addition to cancer detection and staging, PET/CT imaging is becoming more important as a quantitative biomarker for monitoring response to therapy and an evaluation tool for new drug therapies [3–5]. The most common radiotracer is fluorine-18 (^{18}F)-radiolabeled fluorodeoxyglucose (FDG) [6].

PET/CT scanners are designed to measure *in vivo* radioactivity concentration (kBq/ml) with users particularly interested in the relative tissue uptake of radiotracer in targeted areas *versus* uptake in normal tissue in the same patient. The two most significant sources of variation that occur in practice are amount of injected tracer and patient size. As a first-order compensation for these variations, the standardized uptake value (SUV) is commonly used as a *relative* measure of tracer uptake [7]. The basic expression for a SUV (g/ml) is:

$$SUV = \frac{A}{(\mathrm{I}/\mathrm{W})} \quad (1)$$

Address all correspondence to: Robert K. Doot, PhD, University of Pennsylvania, 415 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021.
E-mail: robdoot@upenn.edu
[1]This work was supported by a National Cancer Institute Cancer Imaging Program Science Applications International Corporation (SAIC) contract 24XS036-004 and U01 grant CA148131 from the National Institutes of Health.
Received 15 December 2013; Revised 27 January 2014; Accepted 27 January 2014

DOI 10.1593/tlo.13850

where A is the decay-corrected radioactivity activity concentration (kBq/ml) measured by the PET scanner within a region of interest (ROI), I is the decay-corrected amount of injected radiotracer (MBq), and W is the patient weight (kg), which is used as a surrogate for a distribution volume of tracer, with some sites preferring lean body mass or body surface area instead of weight. If all the injected radiotracer is retained and uniformly distributed throughout the body, the SUV everywhere will be 1 g/ml regardless of the amount of radiotracer injected or patient size. Ideally, the use of SUVs removes variability introduced by differences in patient size and the amount of injected radiotracer to facilitate comparison of the relative tracer uptake between tissues in the same patient or between lesions in different patients. An SUV from whole-body imaging is analogous to a brain tissue uptake ratio with a critical difference of a tissue uptake ratio using a second PET scanner measure of activity concentration in normal background tissue in place of the ratio of injected dose to weight used in SUV. The same instrument makes the two ROI activity measures in brain tissue uptake measures, so it could be reasonable to assume that any instrument calibration errors will cancel each other out. Three instruments consisting of a PET scanner, dose calibrator, and weight scale collect measurements for the weight-based SUV calculation, so any changing individual measurement biases in the different instruments due to measurement error or longitudinal drifts between calibrations could result in different SUV measurement biases for each PET scanner. In practice, there are several sources of bias and variance that are introduced in the measurement process. For example, in estimating SUVs for FDG uptake in tumors, there are biological factors as well as instrumentation factors that affect the SUV measurement [8–10].

The dose calibrator, which is required for quantitative PET imaging, is one of the potential sources of error in global scaling of PET image values and SUV calculations. A National Institute of Standards and Technology (NIST)–traceable mock dose standard for ^{18}F was developed to create a means of monitoring accuracy and precision in dose calibrator measurements of ^{18}F-labeled PET tracers [11]. As much as a 10% shift in calibration was observed when comparing dose calibrator models from different manufacturers of dose calibrators [12]. Thus, it is important to understand the longitudinal variability of an individual dose calibrator used for both scanner calibration and patient SUV measurements and any potential differences in longitudinal drift of measurement biases between the dose calibrator and PET scanner.

In principle, regular calibration of PET scanners should generate a global scanner calibration factor to correct for global sensitivity variations and any longitudinal drifts in individual sensor measurements. However, scanner calibrations typically occur monthly, quarterly, semiannually, and annually, (or not at all) and so cannot compensate for longitudinal drifts in PET measurement variations in the global scanner calibration factor that occur on a time scale shorter than months. There is little information available on the time-varying behavior of the global scanner calibration factor. Studies by Doot et al. show that for a period of minutes to a few hours, the global efficiency variation is approximately 0.3% for the General Electric (GE) Discovery STE (DSTE) PET scanner used in this study [13]. For longer periods (months to years), a recent study by Lockhart et al. indicated that, with careful calibration procedures, there is approximately 4% variability introduced by scanner calibration [14]. In this study, we use the NIST-traceable ^{18}F dose calibrator standard to construct an *implicitly* traceable ^{18}F PET scanner standard to analyze the long-term variations in PET scanner and dose calibrator instrumentation factors that affect the accuracy of SUVs. This manuscript expands an abstract presented at the 57th Annual Society of Nuclear Medicine (SNM) meeting [15] and summary results based on preliminary local site analyses [16] using more robust central analysis of Digital Imaging and Communications in Medicine (DICOM) files from all sites.

Materials and Methods

Phantom Design

The cross-calibration system consisted of two components as shown in Figure 1 [17]. The first is a NIST-traceable dose calibrator ^{18}F mock dose standard [11] constructed from germanium-68/gallium-68 (^{68}Ge/^{68}Ga) in epoxy (RadQual, LLC, Weare, NH). From the same batch of ^{68}Ge/^{68}Ga in epoxy, an implicitly traceable ^{18}F surrogate source for PET scanners was constructed (Figure 1). The PET scanner surrogate source was a 6 × 6-cm cylinder. The size was chosen on the basis of opposing constraints; avoiding resolution loss (also called partial volume errors) in a central ROI, *versus* avoiding unrealistic count-rate effects due to excessive total activity in the scanner field of view. The cylindrical ^{68}Ge/^{68}Ga source was attached to the bottom of a modified American College of Radiology (ACR) PET phantom in place of the standard cold rod component as shown in Figure 1. A PET cross-calibration kit (PET F-18 X-Cal System, RadQual, LLC) is commercially available. It is similar to the prototype described in this manuscript except the surrogate cylinder size was reduced to 4.5 × 4.5 cm and a third ^{68}Ge/^{68}Ga source was added for cross-calibration of well counters [16].

Scanning Protocol

Weight-normalized SUVs were used, with the patient weight of the phantom entered as 70 kg into PET scanners for all imaging sessions because variation in PET values due to body habitus measures was not of interest in this study. The same PET cross-calibration kit was sent to all participating sites twice to allow longitudinal measurements. Some of the community PET imaging centers required that we did not change any of their dose calibrator or PET scanner parameters from their normal ^{18}F settings used for their whole-body oncology ^{18}F-FDG PET scans. Therefore, we measured ^{18}F activities, activity concentrations, and weight-normalized SUV of the kit's two ^{68}Ge/^{68}Ga sources at all six sites using each site's whole-body clinical PET settings for ^{18}F-FDG scans and corrected to equivalent ^{68}Ge/^{68}Ga values. The ^{68}Ge/^{68}Ga scanner source was rescanned using ^{68}Ge/^{68}Ga settings during 8 of the 12 imaging sessions to allow comparison between ^{68}Ge/^{68}Ga concentrations calculated from ^{18}F and ^{68}Ge/^{68}Ga PET images.

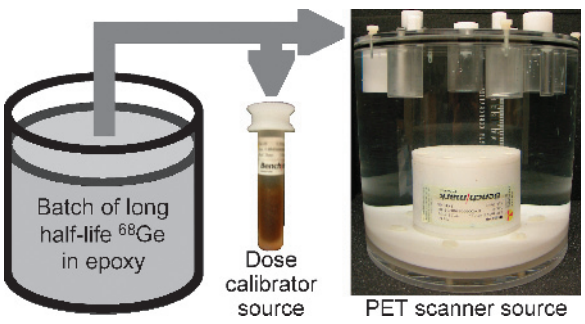

Figure 1. PET cross-calibration kit. One batch of epoxy containing ^{68}Ge/^{68}Ga was used to construct solid ^{68}Ge/^{68}Ga mock dose calibrator and surrogate scanner sources with identical activity concentrations [17].

PET Imaging Centers' Characteristics and Equipment

Activity in the same ^{68}Ge/^{68}Ga dose calibrator mock dose source and activity concentration in the corresponding ^{68}Ge/^{68}Ga PET scanner surrogate source were measured twice at six sites using the equipment listed in Table 1. The PET scanner measurements were made between May 2009 and April 2010. The repeat measurement occurred after a regularly scheduled PET scanner calibration (range of 29 to 226 days between measurements). Measurements were assessed using PET/CT scanners from three major manufacturers (GE Healthcare Technologies, Milwaukee, WI; Philips Healthcare, Eindhoven, The Netherlands; and Siemens Medical Solutions USA, Knoxville, TN) and dose calibrators from two major manufacturers (Biodex Medical Systems, Shirley, NY and Capintec, Inc, Ramsey, NJ).

Common "True" ^{68}Ge/^{68}Ga Activity Concentration Reference

A common "true" reference ^{68}Ge/^{68}Ga activity concentration, A_X, on the reference date (T_R) of 3 October 2008 was calculated by dividing the 1.469 MBq activity (a_x) of the batch of ^{68}Ge/^{68}Ga in epoxy measured by RadQual by the ratio of the RadQual-reported epoxy weight of 3 g (W_E) to epoxy density of 1.10 g/ml (ρ).

Activity Measurements of ^{68}Ge/^{68}Ga Mock Syringe in Dose Calibrators

The dose calibrator ^{18}F setting was used to measure the activity of the RadQual mock dose (a_D) on the experiment date (T_E) to enable detection of changes in dose calibrator settings between measurements. The corresponding ^{68}Ge activity (a_g) on T_E was then calculated by multiplying a_D by the relative response factor (F) of 1.054 between ^{68}Ge/^{68}Ga and ^{18}F dose calibrator measurements of the RadQual mock syringe as determined by the NIST [11]. The activity concentration of the ^{68}Ge/^{68}Ga epoxy in the mock dose (A_D) was calculated by the equation

$$A_D = \frac{a_D \cdot F \cdot \rho}{W_E} \tag{2}$$

by also accounting for the ^{68}Ge epoxy density and weight. The reference date (T_R) of ^{68}Ge activity concentration (A_C) was then calculated by the equation

$$A_C = A_D \cdot e^{(\lambda_G \cdot (T_E - T_R))} \tag{3}$$

where λ_G is the ^{68}Ge/^{68}Ga decay constant, which equals natural logarithm of 2 divided by RadQual-recommended half-life of ^{68}Ge of 270.8 days [18]. The reference date (T_R) dose calibrator measured ^{68}Ge activity (a_G) was calculated by substituting a_G and a_g for A_C and A_D in Equation 3.

Activity Concentrations of ^{68}Ge/^{68}Ga Sources from ^{18}F and ^{68}Ga Images

Mean activity concentration measurements in SUV units were extracted automatically from PET images using a cubical ROI with a volume of 3.375 cm^3 and sides of 1.5 cm centered in the middle of the ^{68}Ge/^{68}Ga scanner source using a plug-in application written in Objective-C for OsiriX open-source imaging software version 5.0 (Pixmeo SARL, Bernex, Switzerland).

For DICOM slices acquired during the first bed position, we convert measured SUV (S_M) into Bq/ml units decay corrected to the start of acquisition, A_{1S} (activity at time T_{1S}), using the equation

$$A_{1S} = S_M \cdot \left[\frac{A_I \cdot e^{-(\lambda_F \cdot (T_{1S} - T_i))}}{W \cdot 10^3}\right] \tag{4}$$

where the bracketed factor is from Quantitative Imaging Biomarkers Alliance SUV pseudocode (http://qibawiki.rsna.org/index.php?title=Standardized_Uptake_Value_%28SUV%29) and W is patient weight in the DICOM header field (0010,1030) resulting from a value of 70 kg entered by the PET scanner operator. We entered 70 kg for patient weight in PET scanners to avoid errors associated with scanner manufacturers' various methods of rounding entered weight values.

The decay factor d for ^{18}F radionuclides was computed as

$$d = e^{(\lambda_F \cdot (T_m - T_{1S}))} \tag{5}$$

where λ_F is the ^{18}F decay constant, $T_m = T_{1M}$ for slices in the first bed position and $T_m = T_{2M}$ for slices in the second bed position. For the Philips PET scanners, we define T_m equal to the average of the two acquisitions' mid time points of T_{1M} and T_{2M}. The resulting $d$values were similar to DICOM DecayFactor field (0054,1321) for GE and Siemens PET scanners (average difference was –0.4 ± 0.6%, ± SD; n = 10) but not similar for Philips scanners, which populated this field with a nominal value of 1. We used decay factor d calculated by Equation 5 instead of DICOM DecayFactor field (0054,1321) to uncorrect the amount of ^{18}F decay correction used

Table 1. PET/CT Scanner and Dose Calibrator Descriptions.

Trial Site	PET/CT Manufacturer and Model	PET Reconstruction Method	Dose Calibrator Manufacturer and Model
MMC*	Siemens Biograph TruePoint	3D-OSEM, 4i8s†	Biodex Atomlab 100
PIC‡	General Electric DST	2D-OSEM, 2i30s†	Capintec CRC-15R
SCCA§	General Electric DSTE	2D-FBP¶	Capintec CRC-15R
SVH#	Siemens Biograph 6	3D-OSEM, 2i8s†	Biodex Atomlab 100
UWMC**	General Electric DSTE	2D-FBP¶	Capintec CRC-127R
UWMC** Annex	Philips Gemini TF TOF	3D-BLOB-TF††	Capintec CRC-15W

*Multicare Medical Center, Tacoma, WA 98405.
†3-Dimensional space (3D) or 2-D dimensional space (2D) subsets expectation maximization (OSEM), iterations (i), and subsets (s).
‡Providence Imaging Center, Anchorage, AK 99508.
§Seattle Cancer Care Alliance, Seattle, WA 98109.
¶2-Dimensional space (2D), filtered back projection (FBP).
#Skagit Valley Hospital's Regional Cancer Care Center, Mount Vernon, WA 98273.
**University of Washington Medical Center, Seattle, WA 98195.
††This 3D lines of response–based iterative row-action maximum-likelihood algorithm incorporates time-of-flight information.

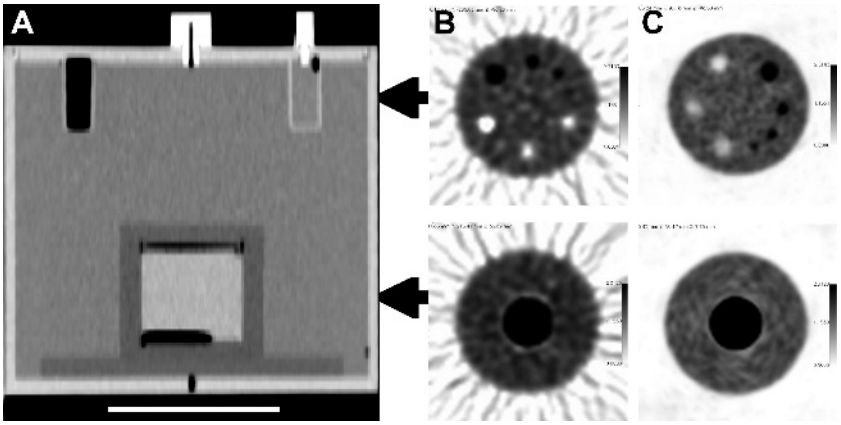

Figure 2. PET cross-calibration ^{68}Ge/^{68}Ga source for the scanner assembled in a modified ACR phantom in a CT image with a 10-cm white scale bar (A) and sample axial PET slices with ^{18}F activity in ACR PET lid cylinders and in the ACR phantom background from one site using analytical PET reconstruction (B) and another site employing iterative reconstruction (C). Air voids in the ^{68}Ge epoxy are apparent at the top and bottom of the cylinder in the CT image.

by scanners when imaging ^{68}Ge/^{68}Ga sources in ^{18}F acquisition mode to enable consistent quantitative correction of PET images from all three PET scanner manufacturers.

We then compensate for the scanners' ^{18}F decay correction of the ^{68}Ge/^{68}Ga activity concentration from the acquisition start, A_{1S}, to the middle of the acquisition, T_{1M}, corrected for the difference between the ^{68}Ge/^{68}Ga and ^{18}F branching ratios (positron fractions), and finally decay corrected to the reference time T_R. The corrected activity concentration is named A_F and defined by the equation

$$A_F = \frac{A_{1S}}{d} \cdot \frac{f_F}{f_G} \cdot e^{(\lambda_G(T_E - T_R))} \quad (6)$$

where f_F and f_G are the positron fractions for ^{18}F and ^{68}Ge/^{68}Ga radionuclide decays. We used the positron fraction values provided in each manufacturer's DICOM field (0018,1076) from the ^{18}F and ^{68}Ge/^{68}Ga scans. Activity concentrations of the ^{68}Ge/^{68}Ga source from ^{68}Ge/^{68}Ga PET images were decay corrected to reference time T_R by the equation

$$A_G = A_g \cdot e^{(\lambda_G(T_E - T_R))} \quad (7)$$

where A_g is activity concentrations of the ^{68}Ge source from ^{68}Ge/^{68}Ga images on the experiment date T_E.

Error Analysis

Biases in the dose calibrator measurements (E_D) were calculated by subtracting assumed true activity, a_x, from a_G and dividing the result by a_x. Biases in the PET scanner measurements (E_A) were similarly estimated by subtracting A_X from A_G and dividing the result by A_X. Biases in SUV measures (E_S) were then calculated by the equation

$$E_S = \frac{(1 + E_A)}{(1 + E_D)} - 1 \quad (8)$$

where the error in the measurement of body habitus (i.e., weight, lean body mass, or body surface area) in the SUV calculation is assumed to be negligible relative to the body habitus value. Biases between ^{68}Ge/^{68}Ga activities calculated from ^{18}F and ^{68}Ge/^{68}Ga PET images were determined by subtracting A_G from A_F and dividing the result by A_G.

Percentage biases in activity measures by dose calibrators were related to equipment manufacturers using the unpaired rank score t test.

Results

Figure 1 shows the two ^{68}Ge/^{68}Ga sources in the PET calibration kit [17] including an image of the cylindrical ^{68}Ge/^{68}Ga surrogate source inside a modified ACR phantom for PET scanner measurements. Figure 2 exhibits a sample CT cross section of the modified ACR phantom

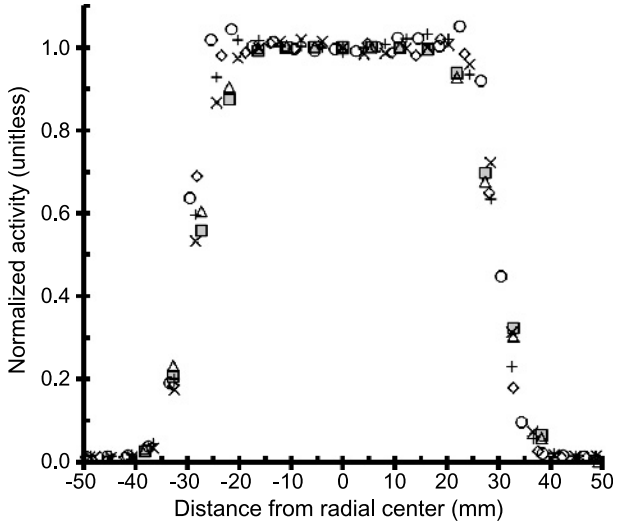

Figure 3. Transaxial linear activity profiles from one PET scan of same ^{68}Ge/^{68}Ga surrogate source from the first visit to all sites (n = 6). Transaxial linear activity profiles from three adjacent central axial slices were averaged and normalized by the average profile center activity values. A ringing artifact is apparent in profile (open circles) from the Philips Gemini TF TOF PET scan.

Table 2. Longitudinal Errors in PET Dose Calibrator and Scanner Measurements (Unitless).

Site*	DC† Error Scan 1	DC† Error Scan 2	Change in DC† Error	PET‡ Error Scan 1	PET‡ Error Scan 2	Change in PET‡ Error
1	9.0%	−50.3%§	−59.3%	−3.4%	−26.8%	−23.4%
2	8.4%	8.7%	0.3%	13.2%	1.1%	−12.0%
3	8.1%	8.8%	0.7%	6.0%	6.4%	0.4%
4	−0.4%	−2.5%	−2.1%	−7.6%	−14.2%	−6.6%
5	−0.3%	−3.8%§	−3.5%	−6.6%	−15.1%	−8.5%
6	1.8%	2.1%	0.3%	−18.2%	−7.1%	11.1%

*Site numbers are same in Tables 2 and 3 but in different order than in Table 1.
†Dose calibrator activity measurement.
‡PET scanner activity concentration measurement.
§Dose calibrator settings different for second measure.

and sample images from one site using analytical PET image reconstruction and another site using iterative PET reconstruction.

The patient weight in the DICOM header field (0010,1030) was 69.95 ± 0.06 kg (± SD, n = 12) due to each PET scanner manufacturer storing the manually entered 70 kg as a different value (70.000 for GE, 69.916 for Siemens, and 69.850 for Philips PET scanners). The assumption of negligible error in the body habitus value in Equation 8 is reasonable because the SD in the patient weight values was only 0.09% of the operator-entered reference value of 70 kg in our phantom study.

The average difference between ^{68}Ge/^{68}Ga activity concentrations estimated from ^{18}F and ^{68}Ge/^{68}Ga PET images was 0.1 ± 1.6% (range, −1.9%-3.7%; n = 8). The small average difference in calculated PET bias from ^{18}F and ^{68}Ge/^{68}Ga PET scans indicates that it is feasible to calculate bias from ^{18}F scans of ^{68}Ge/^{68}Ga sources. Subsequently reported values are based on ^{68}Ge/^{68}Ga activity concentrations extracted from ^{18}F PET images.

The plot of transaxial linear activity profiles from all sites in Figure 3 indicates that the employed cubical ROI with a volume of 3.375 cm^3 and sides of 1.5 cm will not experience partial volume measurement errors when centered in the source.

Table 2 shows longitudinal changes in errors in individual PET dose calibrator and scanner measurements ranged from −59% to 1% for dose calibrators and from −23% to 11% for PET scanners. Table 3 displays the corresponding longitudinal change in SUV measurement bias ranged from −11% to 59%. The site order in Table 1 was randomly changed for use in column 1 of Tables 2 and 3 to protect the identity of any site with unexpected large measurement errors such as site 1 in Tables 2 and 3. The fourth and fifth sites in Tables 2 and 3 used Atomlab dose calibrators (Biodex Medical Systems), whereas the other four sites used Capintec dose calibrators. Individual site results were relayed to each site so corrective actions could be taken if warranted by local quality control or quality assurance protocols. The dose calibrator error measurements correlated to equipment manufacturer after censoring the outlying −50% bias data point at site 1 (P = .001, n = 11).

Table 3. Biases in PET SUV Measurements (Unitless).

Site*	SUV Error† Scan 1	SUV Error† Scan 2	Change in SUV Error‡
1	−11.3%	47.3%	58.6%
2	4.4%	−6.9%	−11.4%
3	−1.9%	−2.2%	−0.3%
4	−7.2%	−12.0%	−4.7%
5	−6.3%	−11.7%	−5.4%
6	−19.7%	−9.0%	10.7%

*Site numbers are same in Tables 2 and 3 but in different order than in Table 1.
†SUV biases calculated by Equation 8.
‡Change in SUV error does not equal column 3 with column 2 in all cases due to rounding errors.

Discussion

By using a NIST-traceable mock dose source for the dose calibrator measurement and an implicitly traceable surrogate source for PET scanner measurements, we were able to estimate the absolute measurement bias for both dose calibrators and PET scanners for ^{18}F sources (Tables 2 and 3) and ^{68}Ge/^{68}Ga sources. From these measurement biases, we can then calculate the error in SUV estimation assuming the sources of bias are the same for patients and the phantom and no operator or patient weighing errors occurred during a patient scan.

The scan protocol in our studies used a ^{68}Ge/^{68}Ga surrogate source in the PET scanner using an ^{18}F acquisition protocol. This was motivated by the desire to use the most common clinical scan protocol at each site, which was primarily an FDG multibed ^{18}F acquisition protocol. This approach simplified the acquisition process. However, the use of the ^{18}F half-life (110 minutes), instead of the actual ^{68}Ge half-life (271 days), meant that post-reconstruction correction to the image data was a requirement. The corrections were complicated, requiring removing the effect of the ^{18}F positron fraction and half-life per bed position and then correction for ^{68}Ge/^{68}Ga positron fraction and half-life. Although the corrections were straightforward in concept as shown in Equations 4 through 7, it is likely that similar bias calculations derived from ^{18}F scans of ^{68}Ge sources will be limited unless PET scanner manufacturers can be convinced to provide the bias calculation software. Due to the complexity of corrections required to determine PET biases from ^{18}F images of ^{68}Ge/^{68}Ga sources, we recommend PET biases be calculated from ^{68}Ge/^{68}Ga images of ^{68}Ge/^{68}Ga to simplify bias calculation unless PET scanner manufacturers are willing to provide the necessary software.

After necessary corrections to PET image data, the resulting derived values are listed in Tables 2 and 3. We decided *a priori* to censor site identities for individual results to facilitate participation of community PET imaging centers located in the states of Alaska and Washington. As a consequence, the site numbers in Tables 2 and 3 do not correspond to the order of sites in Table 1. These results showed the following two notable effects: There was variability between sites in both absolute measurement (kBq/ml) and SUV, and this variability changed in time. Closer inspection shows that dose calibrator measurement biases do not always match PET scanner measurement biases and combined dose calibrator and PET scanner measurement biases do not cancel out to reduce SUV biases, as is sometimes proposed [8–10].

The reasons for exceptionally large bias in dose calibrator and PET scanner measurement errors for the second scan at site 1 in Table 2

are not conclusively known. The second scan measurements at this site were conducted and reported by local staff, and the site declined our offers for subsequent investigation of the large measurement biases. The corresponding range of change in SUV bias in Table 3 from −11% to 59% was larger than expected on the basis of the reported coefficients of variation from repeat measures of ^{68}Ge/^{68}Ga phantoms that range from a low 0.2% (15-cm mean ROIs for 80-minute scans on same day, n = 20) [13] to 3.7% (15-cm mean ROIs for images captured months apart on same scanner, n = 13) [14] to 43% (1.2-cm mean ROIs on small source with 2.5-cm diameter captured months apart at nine different national imaging centers using local analyses) [19]. We speculate that the higher than expected measurement biases may be due to the 9 to 15 operator-dependent steps involved in the calibration process [10]. Another potential source of error is longitudinal drift in PET scanner sensitivity relative to the typically more stable dose calibrator sensitivity. At this same site, the dose calibrator measurement bias changed from +9% in scan 1 to −50% in scan 2. This change in error is suspected to be due to an incorrect change in the calibration setting for the dose calibrator that was applied sometime between scans 1 and 2, though their self-reported new calibration setting was consistent with the current recommendations of their dose calibrator's manufacturer.

This study only evaluated six sites at two time points for a total of 12 PET scanner and 12 dose calibrator measurements (not including the eight ^{68}Ge/^{68}Ga repeat scans of the scanner surrogate source). Thus, the measured biases should not be considered an estimate of the ensemble of all PET imaging sites. Because the potential error is unbounded, there may not be meaningful statistics from such an analysis. We are now conducting a multiyear study of a larger number of sites where the cross-calibration process is used before repeat FDG scans of patients to enable assessment of measurement error associated with instrumentation separately from error associated with imaging patients. Although a small number of sites were evaluated, we believe that this is sufficient to demonstrate the potential for error at all sites. Another limitation is that ^{68}Ga decay generates a small number (~3%) of prompt γ rays [20]. These prompt γ rays can lead to triple coincidences, which may be handled differently by each scanner thus leading to small variations between scanners for equivalent PET scanner surrogate sources.

The use of a dose calibrator and PET scanner cross-calibration process as described here has several considerations. The potential value was demonstrated by measurement of varying bias across sites and time. The procedures used in this study were complicated by the use of an ^{18}F scan protocol with a ^{68}Ge/^{68}Ga source in the PET scanner. This would be considerably simplified by the adoption of a routine ^{68}Ge/^{68}Ga scan protocol. This points out that there are two different modes of use for the cross-calibration process: 1) checking the scanner and dose calibrator measurements for consistency, as was done in this study, or 2) integration of the cross-calibration sources directly into the scanner calibration process. The latter approach depends on the specific PET scanner calibration process, which varies between scanner manufacturers.

Conclusions

The results indicate that bias in serial SUV measurements on the same scanner varies over time and between scanner sites. The conventional wisdom of SUVs having less error than other PET measurements due to any biases in PET scanner and dose calibrator measurements canceling each other out during SUV calculation is not supported by the large range of changes in SUV error from −11% to 59% in Table 3. The findings suggest that use of PET dose calibrator and scanner cross-calibration kits will be useful in multicenter imaging trials to both assess bias and enable correction of biases due to instrumentation factors in serial PET studies. Methods could be adopted to improve data quality either by prospective scanner calibration or retrospective *post hoc* corrections.

Acknowledgments

The authors thank Seattle Cancer Care Alliance Network staff and faculty, especially Tove Thompson, Cecilia Zapata, and Ben Greer, for supporting recruitment of Network sites to the study and thank Brian Pankow and Raymond Noble in the Radiation Safety Office at the University of Washington for facilitating transport of ^{68}Ge/^{68}Ga sources to participants. We are also grateful for the assistance in making local dose calibrator and PET scanner measurements from Roland Zheng at Multicare Medical Center, Tiffany Kooken and Melissa Bridges at Skagit Valley Hospital's Regional Cancer Care Center, Karen Hanrahan, Norm Lind, and Kathryn Collins at Providence Imaging Center, Wendy McDougald at the University of Washington Medical Center, and Lawrence MacDonald at the Seattle Cancer Care Alliance. The authors also appreciate discussions with Joshua Scheuermann and Joel Karp from the University of Pennsylvania, members of Radiological Society of North America, Quantitative Imaging Biomarkers Alliance FDG-PET Technical Committee, members of American Association of Physicists in Medicine (AAPM) Task Group No. 126 on PET/CT Acceptance Testing and Quality Assurance, and members of the National Cancer Institute's Quantitative Imaging Network. The ^{68}Ge/^{68}Ga PET dose calibrator and scanner cross-calibration kit used in this manuscript was a prototype for the "PET F-18 X-Cal System" now sold by RadQual, LLC and protected by United States patent 7,858,925. Authors P.K., K.A., and R.D. are also inventors of the aforementioned patent [17] and may receive compensation from the patent licensing agreement between the University of Washington and RadQual, LCC.

References

[1] Weber WA, Grosu AL, and Czernin J (2008). Technology insight: advances in molecular imaging and an appraisal of PET/CT scanning. *Nat Clin Pract Oncol* **5**, 160–170.

[2] Fletcher JW, Djulbegovic B, Soares HP, Siegel BA, Lowe VJ, Lyman GH, Coleman RE, Wahl R, Paschold JC, Avril N, et al. (2008). Recommendations on the use of ^{18}F-FDG PET in oncology. *J Nucl Med* **49**, 480–508.

[3] Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, Oyen WJ, Kotzerke J, Hoekstra OS, Pruim J, et al. (2010). FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* **37**, 181–200.

[4] Frank R and Hargreaves R (2003). Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov* **2**, 566–580.

[5] Wahl RL, Jacene H, Kasamon Y, and Lodge MA (2009). From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med* **50**(suppl 1), 122S–150S.

[6] Kelloff GJ, Hoffman JM, Johnson B, Scher HI, Siegel BA, Cheng EY, Cheson BD, O'Shaughnessy J, Guyton KZ, Mankoff DA, et al. (2005). Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res* **11**, 2785–2808.

[7] Thie JA (2004). Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med* **45**, 1431–1434.

[8] Boellaard R (2009). Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* **50**(suppl 1), 11S–20S.

[9] Adams MC, Turkington TG, Wilson JM, and Wong TZ (2010). A systematic review of the factors affecting accuracy of SUV measurements. *Am J Roentgenol* **195**, 310–320.

[10] Kinahan PE and Fletcher JW (2010). Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR* **31**, 496–505.

[11] Zimmerman BE and Cessna JT (2010). 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* **51**, 448–453.

[12] Zimmerman B, Kinahan P, Galbraith W, Allberg K, and Mawlawi O (2009). Multicenter comparison of dose calibrator accuracy for PET imaging using a standardized source. *J Nucl Med Meeting Abstracts* **50**, 472.

[13] Doot RK, Scheuermann JS, Christian PE, Karp JS, and Kinahan PE (2010). Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys* **37**, 6035–6046.

[14] Lockhart CM, MacDonald LR, Alessio AM, McDougald WA, Doot RK, and Kinahan PE (2011). Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. *J Nucl Med* **52**, 218–224.

[15] Doot R, Allberg K, and Kinahan P (2010). Errors in serial PET SUV measurements. *J Nucl Med Meeting Abstracts* **51**, 126.

[16] Doot RK, Thompson T, Greer BE, Allberg KC, Linden HM, Mankoff DA, and Kinahan PE (2012). Early experiences in establishing a regional quantitative imaging network for PET/CT clinical trials. *Magn Reson Imaging* **30**, 1291–1300.

[17] Kinahan PE, Allberg KC, Doot RK, Lockhart CM, McDougald WA, inventors; University of Washington, assignee. Calibration method and system for PET scanners. US patent 7858925. 2010.

[18] Firestone RB, Shirley VS, Chu SYF, Baglin CM, and Zipkin J (1996). *The Table of Isotopes - 8th Edition and Beyond.*. John Wiley and Sons, New York.

[19] Fahey FH, Kinahan PE, Doot RK, Kocak M, Thurston H, and Poussaint TY (2010). Variability in PET quantitation within a multicenter consortium. *Med Phys* **37**, 3660–3666.

[20] Cherry SR, Sorenson JA, and Phelps ME (2003). *Physics in Nuclear Medicine.* (3rd ed). Saunders, Philadelphia, PA.