Assessing and accounting for the impact of respiratory motion on FDG uptake and viable volume for liver lesions in free-breathing PET using respiration-suspended PET images as reference

Guang Li[a] and C. Ross Schmidtlein
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York 10065

Irene A. Burger
Department of Radiology, University Hospital of Zurich, CH-8091 Zurich, Switzerland

Carole A. Ridge[b] and Stephen B. Solomon
Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065

John L. Humm
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York 10065

(Received 1 December 2013; revised 5 July 2014; accepted for publication 28 July 2014; published 14 August 2014)

Purpose: To assess and account for the impact of respiratory motion on the variability of activity and volume determination of liver tumor in positron emission tomography (PET) through a comparison between free-breathing (FB) and respiration-suspended (RS) PET images.
Methods: As part of a PET/computed tomography (CT) guided percutaneous liver ablation procedure performed on a PET/CT scanner, a patient's breathing is suspended on a ventilator, allowing the acquisition of a near-motionless PET and CT reference images of the liver. In this study, baseline RS and FB PET/CT images of 20 patients undergoing thermal ablation were acquired. The RS PET provides near-motionless reference in a human study, and thereby allows a quantitative evaluation of the effect of respiratory motion on PET images obtained under FB conditions. Two methods were applied to calculate tumor activity and volume: (1) threshold-based segmentation (TBS), estimating the total lesion glycolysis (TLG) and the segmented volume and (2) histogram-based estimation (HBE), yielding the background-subtracted lesion (BSL) activity and associated volume. The TBS method employs 50% of the maximum standardized uptake value (SUV_{max}) as the threshold for tumors with $SUV_{max} \geq 2\times SUV_{liver\text{-}bkg}$, and tumor activity above this threshold yields $TLG_{50\%}$. The HBE method determines local PET background based on a Gaussian fit of the low SUV peak in a SUV-volume histogram, which is generated within a user-defined and optimized volume of interest containing both local background and lesion uptakes. Voxels with PET intensity above the fitted background were considered to have originated from the tumor and used to calculate the BSL activity and its associated lesion volume.
Results: Respiratory motion caused SUV_{max} to decrease from RS to FB by $-15\% \pm 11\%$ ($p = 0.01$). Using TBS method, there was also a decrease in SUV_{mean} ($-18\% \pm 9\%$, $p = 0.01$), but an increase in $TLG_{50\%}$ ($18\% \pm 36\%$) and in the segmented volume ($47\% \pm 52\%$, $p = 0.01$) from RS to FB PET images. The background uptake in normal liver was stable, $1\% \pm 9\%$. In contrast, using the HBE method, the differences in both BSL activity and BSL volume from RS to FB were $-8\% \pm 10\%$ ($p = 0.005$) and $0\% \pm 16\%$ ($p = 0.94$), respectively.
Conclusions: This is the first time that almost motion-free PET images of the human liver were acquired and compared to free-breathing PET. The BSL method's results are more consistent, for the calculation of both tumor activity and volume in RS and FB PET images, than those using conventional TBS. This suggests that the BSL method might be less sensitive to motion blurring and provides an improved estimation of tumor activity and volume in the presence of respiratory motion. © *2014 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4892602]

Key words: positron emission tomography, respiratory motion, lesion uptake and volume, liver cancer

1. INTRODUCTION

In both image-guided thermal ablation and stereotactic body radiation therapy, assessment of viable tumor volume before, during, and after treatment is a useful tool that can influence treatment outcome. ^{18}F-fluoro-deoxy-glucose (^{18}F-FDG) positron emission tomography (PET) and computed tomography (CT) are the clinical standard imaging modalities that provide useful three-dimensional (3D) information on tumor location, metabolic activity, and viable volume. In liver and lung cancer treatment, respiratory motion is a major factor adding to the uncertainty in quantitative tumor imaging in PET of the standard uptake value (SUV), PET and CT tumor contour, localization, and response assessment.[1,2] Due to a

lack of ground truth in patients for tumor size and shape, it is a major clinical challenge to quantify accurately tumor activity and viable volume based on PET/CT images. This problem is further accentuated by tumor motion. Tumor ground truth is not attainable in the clinical setting without surgical intervention.

Previous approaches to reducing motion artifacts have included four-dimensional PET (4D PET) with 4D CT. Studies using these techniques have reported inaccuracies associated with tumor motion and their impact on SUV, tumor delineation, and localization.[3–5] Although 4D PET improves image quality significantly, substantial residual motion artifacts remain in the retrospectively reconstructed 4D PET and 4D CT, primarily due to irregularities in patient breathing.[6–10] There are two main types of 4D PET scan and reconstruction, including respiratory-gated and motion-compensated 3D PET: (1) scanning throughout breathing cycles and sorting raw data into multiple bins of respiratory phases, or scanning only one particular phase with respiratory gating[11–13] and (2) scanning breathing cycles and applying a deformable model derived from 4DCT to reconstruct a single phase scan without suffering from low quantum or long acquisition.[14–16] The former requires considerably longer scan duration than a 3D PET scan, to achieve the same count statistics, while the latter may suffer from deformation uncertainty, which is propagated from residual motion artifacts in 4DCT (Ref. 6) and uncertainty in deformable image registration,[17] the uncertainties will propagate into the new motion-compensated 3D PET images. Therefore, the accuracy and reliability of 4D PET images are of concern. Breath-hold image acquisition is another approach to reduce the motion artifact in PET, including deep inspiration breath hold,[18–20] but the variation of intra- and interbreath hold (15–20 s per breath hold in repetitive holds) during long PET acquisition introduces uncertainty and some patients are incapable of breath holds. Therefore, it is a clinical challenge to achieve a motion-free ground state for patient PET/CT scan acquisitions.

Recently, a medical ventilation method has been applied to achieve temporary respiration arrest and create an essentially motion-free tumor for stereotactic radiotherapy and ablative treatment under institutional research board protocols. Lovelock *et al.*[21] reported using the respiration-suspended (RS) technique to treat liver cancer without motion in single-fraction stereotactic radiosurgery. Patient respiration is suspended on a ventilator under general anesthesia for ~5 min. Ryan *et al.*[22] applied the same RS technique in a PET/CT guided percutaneous liver ablation procedure. In this latter instance, the initial RS PET imaging was used for treatment guidance with first FDG injection and a posttreatment RS PET with a second, double-dose FDG injection (while the first FDG injection has substantially decayed) was used for instantaneous evaluation of the ablative treatment response.

Tumor delineation based on elevated FDG uptake in PET images is a clinical challenge,[23] because no single method meets all major clinical needs, including accuracy, reliability, and performance. Reported segmentation methods include manual drawing, use of a SUV threshold, edge extracting, stochastic modeling, and machine learning. Validation of these segmentation methods includes physician contouring and physical phantom testing. The most popular segmentation method is threshold-based, usually at 42% of maximum standardized uptake value (SUV_{max}), although Monte Carlo simulation can produce a lesion-specific threshold with clinically acceptable volume estimation (±10%).[24] Recently, Burger *et al.*[25,26] proposed a histogram-based method to reproducibly estimate tumor uptake. This method estimates tumor uptake based on a fit of the local nontumor background defined by a Gaussian in the activity-volume histograms. The tumor volume is then estimated by summing the remaining partial bin volume in the activity-volume histogram after background subtraction. This method is called background-subtracted lesion (BSL) method, and it provides a reproducible method for estimating the tumor activity within a volume that contains activity above background.

Therefore, the objective of the present study was first to measure the effect of respiratory motion on defining a target volume based on a fixed threshold method and second to propose a technique to achieve a consistent determination of the amount of radioactivity within a tumor volume that is less sensitive to respiratory motion. PET/CT scans were acquired for patients with liver lesions under the normal condition of free-breathing (FB) and on a ventilator with respiratory suspension (RS), which is assumed to be as close as possible to motion-free "gold standard" image data in a human subject.

2. METHODS AND MATERIALS

2.A. PET/CT image acquisition and analysis

In this study, we report on 20 patients with metastatic liver cancer who underwent tumor ablation by an interventional radiologist. Before the ablative procedure, a PET/CT was acquired with ^{18}F-FDG under FB, followed by a second PET/CT scan under anesthesia on a ventilator with RS. We subsequently refer to this PET scan as RS PET. During the respiratory suspension time, a CT and up to 1.5 min of PET emission data were acquired. The acquisition was performed, while the patients were immobilized in a supine position with body rotation along the longitudinal axis, facilitating probe placement. Patients ($n = 20$, age $= 59 \pm 8$ years) were first administered ^{18}F-FDG by IV injection (4.2 ± 0.2 mCi, or 155 ± 7 MBq) approximately 82 min on average prior to the procedure. The images were acquired on a GE D690 PET/CT scanner (GE Healthcare, Milwaukee, WI) that was installed in the operation room. All reconstructions were performed over a 70 cm field of view with a 128×128 matrix using 2-iterations and 16-subsets with Gaussian transaxial filter of 6.0 mm FWHM and the "heavy" axial filter (3-point smoothing [0.25 0.5 0.25]). Both time-of-flight and sharp iterative reconstruction were used in the reconstructions. Random corrections were done by singles and all other standard PET/CT corrections were performed.

The FB and RS PET/CT image data sets were analyzed in order to determine the magnitude of the error in SUV determination and FDG viable tumor volume estimation associated with free-breathing. Two methods used for activity and

volume assessments were: (1) a threshold-based segmentation (50% SUV_{max}) and (2) a histogram-based estimation (HBE) method with background subtraction, the BSL method. The differences between the FB PET scan and the RS PET scan (a surrogate for motion-free ground truth) were analyzed and the significance of their differences was assessed using both paired Student *t*-test and Wilcoxon signed-rank test, based on *p*-value (significance: $p \leq 0.05$) and z-score (significance: $z \geq 2.0$), respectively. When comparing data between FB and RS, RS is always used as the reference. The relative difference for a quantity (X) is defined as: $\%\text{Diff} = (X_{FB}–X_{RS})/X_{RS} \times 100\%$.

2.B. Conventional TBS method

The first method used was a simple 50% maximum signal threshold value to delineate the tumor edge. This threshold was selected, not because this group believes it best identifies the lesion boundary, but as the theoretical threshold, which defines a radioactive edge at a planar boundary with uniform FDG on one side and no activity on the other. Tumors within human subjects exhibit neither a uniform distribution of activity within the tumor nor a negligible background. Consequently, no single SUV threshold can be argued to accurately segment lesions and in the absence of a surgical pathology "gold standard" we adopted the 50% threshold as a surrogate volume with which the behavior of respiratory motion on the threshold-based methods to detect lesions could be studied. In this study, the threshold method was implemented within the GE software package called PET VCAR (Volume Computer Assisted Reading). For most liver lesions, a threshold of 50% of SUV_{max} was used. This was successful whenever SUV_{max} exceeded $SUV_{background}$ by two folds as was the case for most patients. For small lesions with low SUV uptake, a threshold up to 70% was used. This was necessary to prevent the threshold volume expanding in an uncontrolled manner into a large segment of normal liver. The background uptake was estimated by drawing a spherical volume containing normal liver tissue, distant from the lesion and the liver edge. The volume of segmented region was calculated automatically, together with the mean and standard deviation of the uptake within the volume.

2.C. BSL method

The second method used was the BSL approach, which is a HBE method. If a box (or any arbitrary shape) is placed around a lesion in a PET image and the frequency with which an activity within small ranges is plotted so as to generate an activity-volume histogram, then if the activity were uniformly distributed within the lesion and the local background, the result would be two Gaussian profiles separated by voxel activities corresponding to partial volumes associated with the lesion edges. Real clinical data are not uniform and therefore results in more complex activity volume histograms. For our purposes it was practical to implement the BSL method using a boxed volume of interest (VOI) drawn around the lesion with the inclusion local background activity. Using this volume a histogram was generated using the Freedman–Diaconis binning rule.[27] To provide additional robustness for the histogram of tumors with backgrounds, the initial VOI is broken up via a watershed transform and some of low uptake watersheds were removed prior to generating the histogram. This reduced the asymmetric spread of the low uptake region in the histogram representing background activity allowing a more consistent Gaussian fit. The Gaussian fit was then subtracted from the histogram and all positive values beyond a cutoff were summed to represent the total lesion uptake, the BSL activity. The cutoff was defined as the mean plus two standard deviations (mean $+ 2\sigma$) of the Gaussian. The sum of all signals provide the BSL uptake in SUV*ml.

The lesion volume was estimated by simply summing the volume of the voxels used to estimate the BSL volume. In addition, an equivalent threshold defined as the threshold necessary to compute the same total uptake (SUV*ml) as the BSL method was computed to provide a visual confirmation of the BSL estimate.

3. RESULTS

Table I shows patient general information, PET imaging conditions, and the maximum SUV value for both FB and RS images. The location of lesions is defined using Couinaud segmental anatomy of the liver. Among the 20 patients, the lesion locations were distributed over the entire liver. A *t*-test analysis is performed on maximum SUV between RS and FB (the SUV_{max} in FB is significantly smaller than that in RS by $-15\% \pm 11\%$, $p = 0.01$).

3.A. Lesion volume calculated from threshold-based segmentation

Table II shows the volumetric results from a TBS tool (PET VCAR, GE Healthcare, Milwaukee, WI). An arbitrary threshold (% of SUV_{max}) is used for each lesion. A 50% threshold was applied to 17/20 lesions and 70%/60%/20% for 3/20 lesions (the first two cases where the lesion to liver background ratios were less than 2 and the third case in which the lesion was close to the right kidney). Both segmented lesion volume ($p = 0.01$) and mean lesion SUV values ($p = 0.01$) are significantly different between RS and FB images, while the mean normal liver background SUV is not ($p = 0.15$). The RS lesion volume is smaller than FB volume by $47\% \pm 52\%$, while its mean SUV is higher by $18\% \pm 9\%$ on average. Figure 1 shows two adjacent lesions in RS. These lesions become smeared into a single lesion in FB. In this case, the FB lesion volume increases by 101% in comparison with RS lesion volume, while the mean SUV drops by 37%. Both the *p*-value in paired Student *t*-test and z-score in Wilcoxon signed-rank test indicate that the differences in mean SUV and tumor volume between FB and RS are of significance, while the difference in background uptakes is not. These results agree with previous observations,[1,2,28,29] and in addition quantify the difference between RS and FB (Fig. 3) in patients.

TABLE I. Patients ($n = 20$) with liver metastases were treated with local percutaneous thermal ablation. PET/CT images were acquired during both FB and RS conditions.

Patient	Age	Sex	Segmental location[a]	Injected activity(mCi)	Uptake time (min)	Time Interval from RS to FB (min)	Maximum SUV (g/ml)		
							RS	FB	%Diff
1	57	f	4b	4.4	112	− 7.7	9.8	10.1	4.0
2	53	m	2	4.4	58	15.3	9.5	9.1	− 3.8
3	70	f	6/7	4.4	95	61.2	24.4	19.4	− 20.2
4	45	f	8	4.0	66	10.5	7.2	5.5	− 23.1
5	47	m	5/6	4.2	80	9.5	8.7	7.1	− 19.0
6	52	f	…	4.4	85	10.5	21.9	16.2	− 26.1
7	37	f	7	4.0	97	6.5	7.2	6.2	− 14.6
8	53	f	8/4a	4.3	85	10.0	15.8	13.7	− 13.6
9	52	f	…	4.0	50	10.0	9.1	5.7	− 37.8
10	69	f	4	4.2	60	10.5	9.3	8.5	− 9.3
11	66	f	7	4.3	94	27.7	10.4	8.7	− 16.2
12	52	m	…	4.1	133	9.9	26.8	21.9	− 18.1
13	68	f	8	4.2	92	13.6	5.6	5.7	2.5
14	61	m	8	4.4	67	16.5	13.7	10.8	− 20.7
15	67	m	5	4.1	67	13.6	14.8	13.3	− 10.3
16	62	m	4a	4.1	64	7.6	10.1	8.1	− 19.7
17	65	m	6/7	4.4	78	10.0	6.7	6.3	− 5.5
18	65	m	8	4.0	110	13.5	4.4	4.8	7.4
19	52	f	7	4.0	60	65.4	19.7	24.5	24.2
20	52	m	2	4.0	83	28.2	10.9	11.7	7.3
Mean	59			4.2	82	17.1			− 15.4
St.dev	8			0.2	21	17.4			10.5
p-value								0.01 (≤ 0.05)	
z-score								2.8 (≥ 2.0)	

[a]The Couinaud segmental anatomy of the liver.

TABLE II. Lesion volume assessment using threshold segmentation method in PET images. The lesion volume, mean SUV for lesion and background are listed between RS and FB. Significant differences in segmented lesion volume ($p = 0.01$) and the mean SUV ($p = 0.01$) are between RS and FB PET images, while the averaged background is similar ($p = 0.15$).

Patient	Threshold(%)	Lesion volume (cm^3)			Mean SUV (g/ml)			TLG (g/ml cm^3)			Distant mean liver bkgd (g/ml)		
		RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff
1	50	1.3	2.0	54.3	7.5	6.1	− 18.9	9.5	11.9	24.5	2.4	2.4	− 2.1
2	50	3.4	6.2	80.1	6.9	5.2	− 23.0	23.3	32.2	38.5	2.0	2.2	7.4
3	50	12.0	14.2	17.9	13.9	10.4	− 24.6	166.6	148.0	− 11.2	2.6	3.3	25.0
4	70	2.2	6.5	200.0	5.7	4.4	− 22.8	12.2	28.2	131.5	2.7	2.7	0.0
5	50	2.7	3.6	32.1	5.0	4.4	− 12.1	13.5	15.7	16.2	3.1	2.8	− 7.8
6	50	25.7	51.7	101.1	10.4	6.6	− 36.7	267.0	339.9	27.3	3.0	2.9	− 1.3
7	50	3.2	5.8	78.6	4.8	3.7	− 21.9	15.5	21.6	39.6	2.4	2.4	0.0
8	50	5.6	5.0	− 10.4	11.4	9.7	− 14.9	63.5	48.4	− 23.8	2.8	2.8	1.4
9	50	1.5	2.5	72.8	6.5	4.3	− 34.7	9.6	10.8	13.2	2.8	2.8	1.4
10	50	6.9	7.7	12.8	6.5	6.0	− 8.3	44.5	46.0	3.4	2.5	2.5	− 0.4
11	50	11.1	11.7	6.2	7.6	6.3	− 17.9	84.1	73.4	− 12.7	2.2	2.4	5.8
12	20	27.6	35.8	29.8	10.7	9.2	− 14.2	295.5	329.4	11.5	3.1	3.2	2.3
13	60	6.0	7.1	19.6	4.0	4.0	0.3	23.8	28.6	19.9	2.7	2.8	2.9
14	50	21.3	30.8	44.5	8.9	7.3	− 18.3	190.3	224.8	18.1	3.3	3.1	− 6.3
15	50	2.8	3.5	23.9	10.4	9.4	− 9.3	29.4	33.1	12.7	3.1	3.0	− 1.6
16	50	6.2	5.8	− 6.3	6.4	5.3	− 17.1	39.3	30.5	− 22.4	3.5	2.9	− 17.1
17	50	2.4	3.8	56.1	4.6	4.0	− 13.1	11.3	15.4	35.7	2.0	2.1	3.9

TABLE II. *(Continued).*

		Lesion volume (cm^3)			Mean SUV (g/ml)			TLG (g/ml cm^3)			Distant mean liver bkgd (g/ml)		
Patient	Threshold(%)	RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff
18	50	2.2	8.3	286.5	3.1	2.9	−8.0	6.7	23.9	254.6	1.4	2.0	40.8
19	50	10.7	13.2	23.8	14.1	18.	28.2	150.4	238.7	58.8	2.2	4.3	94.5
20	50	13.4	14.7	9.5	7.3	8.5	16.3	98.1	124.8	27.2	2.6	3.6	41.4
Mean	50.0			47.3			−18.4			17.9			0.6
St.dev	9.7			52.1			9.3			36.0			8.7
p-value			0.01			0.01			0.04			0.15 (>0.05)	
z-score			3.8			2.8			2.2			1.4 (<2.0)	

3.B. Lesion volume calculated from BSL method

The BSL method is illustrated in Fig. 2, in which the histogram of the PET voxel activity in the VOI is shown. The large peak observed in the histogram represents background voxels that are distributed over fewer bins and are thus more numerous than those from the lesion. The mean plus two standard deviations (2σ) of Gaussian fitted to background SUV is used with this fit to separate the background from signal. For visualization purpose, a threshold was derived by finding the equivalent volume necessary to capture this signal. Table III shows the assessment of lesion volume using the BSL method, showing that the lesion volumes in RS and FB PET images are very similar ($0.1\% \pm 16\%$, $p = 0.94$). The BSL values in RS and FB are also shown a relevant difference ($-8\% \pm 10\%$, $p = 0.005$). That was within the range of relative change of averaged local background uptakes between the RS and FB PET images ($6\% \pm 12\%$, $p = 0.10$). What can be explained by the later time points (17 ± 17 min) of the RS scans compared to the FB scans leading to an increase of tumor activity and a decrease in background activity.

3.C. Tumor volume comparison between the two methods

Figures 3 and 4 show the results for the lesion activity and the volume corresponding to the region with activity greater than background for the TBS and BSL methods, respectively.

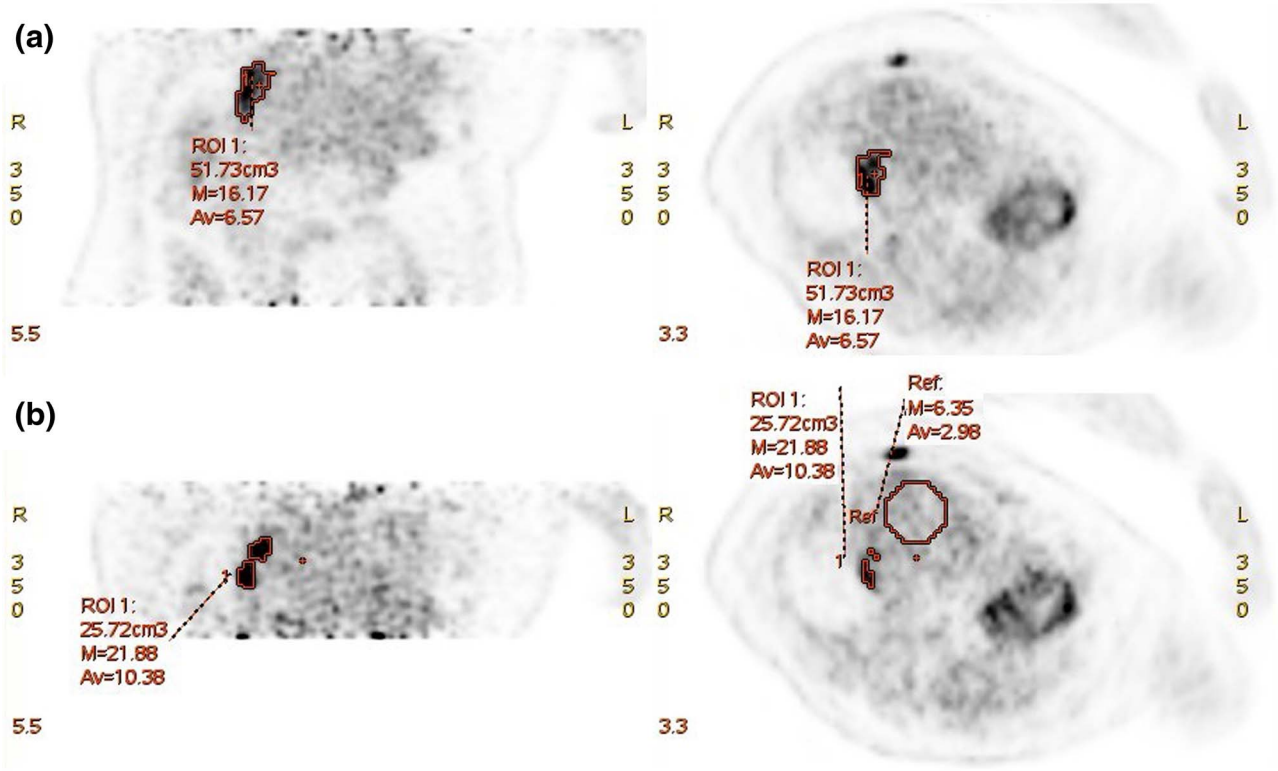

FIG. 1. PET images of patient #6 acquired under (a) FB and (b) RS. Segmentation of the tumor VOI was based on a threshold of 50% SUV_{max}. Two adjacent lesions shown in RS blurred into one bigger lesion in FB, due to respiratory motion. Since some normal liver tissue is blended into the tumor volume, as shown the lesion volume raised from 25.7 to 51.7 cm^3, not only the maximum SUV but also the averaged SUV are dramatically reduced from RS to FB.

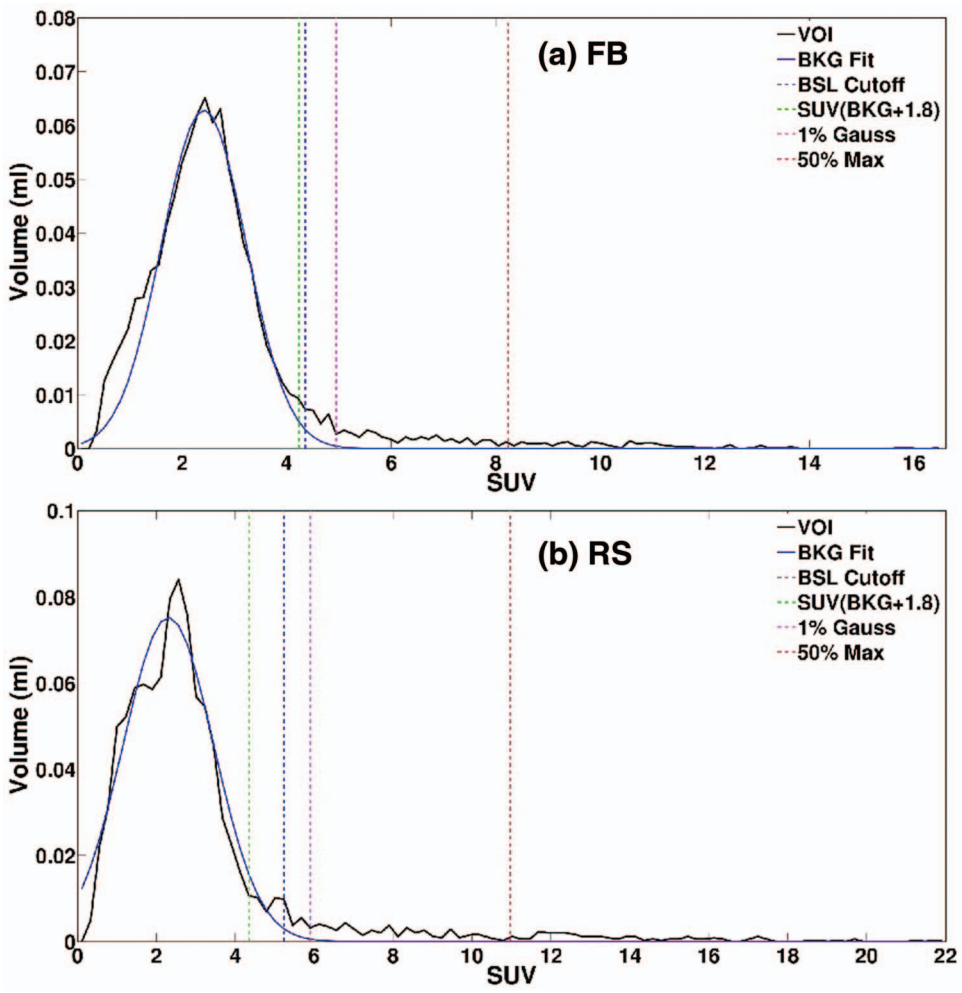

FIG. 2. Histograms of FDG PET volume-activity under (a): (FB) and (b): (RS) scanning conditions for patient #5. The VOI contains tumor and surrounding background (BKG), which fits with Gaussian distribution. Four threshold dotted-lines were drawn: BSL cutoff (mean + 2σ, blue), SUV (BKG + 1.8, green), 1% of the Gaussian distribution (pink), and 50% of SUV_{max} of the lesion (red). SUV_{max} for FB is lower than that for RS, due to motion smearing effect.

The volume values between RS and FB are very different when segmented using TBS method ($47\% \pm 52\%, p = 0.014$), while very similar when estimated with BSL method ($0\% \pm 16\%, p = 0.54$). The TLG difference between RS and FB from TBS method is $17.9\% \pm 36.0\%$ (Table II), while the BSL activity has $-7.7\% \pm 10.4\%$ difference (Table III). The similarity in estimated activity and volume between RS and FB suggests that this measurement of the lesion uptake conserves these quantities regardless of the presence of motion. In addition, the lesions in these 20 patients were spread throughout most segments of the liver (Table I), and were insensitive to patient-specific and location-specific motion. This is potentially useful to assess viable tumor volume in the presence of respiratory motion.

4. DISCUSSION

4.A. Motion-free ground truth for liver tumors

To establish a motion-free reference image of an irregular shaped lesion with a nonuniform radiotracer distribution in patients is not a trivial task. A major confounding factor in determining any radiotracer distribution within patients is involuntary motion. In PET/CT of thoracic or abdominal lesions, respiratory motion has a major impact on the integrity of lesion position, shape, volume, and SUV. To study the effects of respiratory motion on such parameters, we have obtained FDG PET/CT data conducted as part of a PET/CT guided percutaneous liver ablation procedure. In this procedure, the normal physiological breathing control mechanism is suspended under general anesthesia using a mechanical ventilator for up to 2 min, with continuous safety monitoring of patient oxygen and carbon dioxide levels. Within this limited time window, a PET scan can be acquired under motion-free conditions, allowing the establishment of unique patient PET "ground truth"—motion-free PET image data set. Such image data may provide the most stationary radiotracer distribution possible in a patient and the closest PET acquisition to a scan of a phantom. Such data are important because they allow a comparison of PET images of lesions to be studied under normal free-breathing relative to an almost complete respiratory stasis, and thus quantitatively assess the impact of lesion motion in a clinical setting. Using such motion-free PET images, lesion activity can be better quantified with higher confidence

TABLE III. Lesion volume assessment using BSL method. Three differences in PET images between RS and FB are listed. A *t*-test confirms similar estimated lesion volume ($p = 0.54$) between RS and FB, but significant difference background-adapted threshold ($p = 0.002$) between RS and FB.

	Lesion volume (cm^3)			BSL ($g/ml\,cm^3$)			Lesion-specific cutoff threshold			Local SUV bkgd (g/ml)		
Patient	RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff	RS	FB	%Diff
1	21.5	22.2	3.2	136.6	133.8	−2.0	0.23	0.23	0.9	2.4	2.2	−8.5
2	9.3	9.2	−0.8	43.6	41.4	−4.9	0.31	0.33	6.7	1.8	2.2	20.8
3	27.6	28.2	2.0	261.3	238.3	−8.8	0.18	0.24	32.8	2.6	3.3	28.0
4	4.0	3.7	−8.3	17.3	16.7	−3.9	0.47	0.64	35.8	2.7	2.8	3.9
5	11.4	8.4	−26.3	49.7	34.2	−31.2	0.64	0.55	−13.5	2.8	3.0	6.8
6	46.2	47.6	3.0	362.3	326.5	−9.9	0.19	0.26	34.7	2.7	2.5	−8.3
7	6.1	6.6	8.1	24.5	23.0	−5.9	0.43	0.50	16.2	2.5	2.4	−3.1
8	18.9	19.8	4.4	122.9	112.6	−8.4	0.22	0.25	13.0	2.8	3.0	5.3
9	4.0	4.6	13.6	18.6	17.5	−5.4	0.35	0.57	60.0	2.8	3.0	8.4
10	14.2	18.5	30.3	71.3	83.2	16.7	0.31	0.34	9.4	2.5	2.6	5.8
11	25.3	22.1	−13.0	134.9	117.6	−12.8	0.26	0.34	29.5	2.1	2.5	16.5
12	39.8	41.3	3.8	349.6	352.5	0.8	0.17	0.19	12.5	3.1	3.2	3.4
13	17.3	16.2	−6.6	92. 9	87.9	−5.4	0.25	0.28	11.6	3.1	3.2	6.1
14	62.0	62.2	0.4	384.6	383.6	−0.3	0.29	0.37	30.3	3.1	3.6	17.9
15	17.4	13.0	−25.2	95.3	79.8	−16.2	0.25	0.28	13.3	3.0	3.3	8.9
16	6.4	7.4	15.1	38.9	36.1	−7.2	0.51	0.49	−4.0	4.0	3.0	−23.9
17	7.3	8.8	20.4	29.8	30.7	3.2	0.40	0.40	0.7	2.3	2.2	−3.1
18	8.2	6.66	−19.1	38.1	28.7	−24.8	0.41	0.52	26.1	2.7	2.8	5.8
19	44.6	54.5	22.1	482.4	452.1	−6.3	0.16	0.21	33.9	3.5	3.5	2.3
20	32.0	24.1	−24.5	215.4	167.5	−22.2	0.32	0.40	24.8	2.9	3.6	25.8
Mean			0.1			−7.7			18.7			5.9
St.dev			16.1			10.4			17.2			12.3
p-value		0.94 (>0.05)			0.005			0.002			0.10	
z-score		0.3 (<2.0)			2.8			3.2			2.2	

for tumor diagnosis, staging, and response assessment. In addition the lesion's total metabolic volume and shape are more definable and quantifiable for purposes of therapy. To the best of our knowledge, this is the first time that a patient motion-free "ground truth" is established for assessing motion effect on the activity as well as volume of liver tumor in PET/CT images.

4.B. Limitation of a fixed threshold segmentation method

All methods to segment lesions in PET/CT images suffer from the inability to determine their accuracy against a gold standard "true" tumor volume. In this study, the true lesion volumes are also unknown. However, the study of lesions within the liver free from motion smearing brings us one-step closer to the true lesion appearance. Concerning FDG-defined tumor volumes, all methods should be considered as relative determinations. The first methods used were based on defining an image threshold, such methods were initially based on empirical phantom measurements of ^{18}F laden spheres within an ^{18}F background.[30] The principal shortcoming of threshold methods is their inability to account for nonuniform distributions of activity within the lesion and the variability of the background. Most threshold methods are based on SUV_{max}, which is a single point in a PET image that is sensitive to PET image noise, motion blurring, voxel size, and image filtering. The SUV_{max} in FB is smaller than that in RS by $-15\% \pm 11\%$, because the SUV_{max} value in FB is likely smeared with surrounding lower SUV values. Once the SUV_{max} is changed, the threshold at a percentage of the SUV_{max} will be shifted accordingly, resulting in a different lesion volume. Because of the cubic relationship between the diameter and volume, a slight increase or decrease in the delineated edge with different thresholds will result in a much more pronounced difference in volume. It is important to recognize that the magnitude of the error in volume estimation as a consequence of respiratory motion is likely to be valid regardless of the accuracy with the applied threshold. This quantified motion effects on activity assessment and volume estimation of PET lesions using TBS are consistent in scale with those found in phantom simulation studies.[31]

4.C. Conservation of lesion uptake activities in BSL method

The BSL method is a recently-reported reproducible technique to determine the tumor activity of FDG above the background tissue.[25,26] In this histogram-based method, the intensity of local background is assumed to be normally distribution. So when the local background is subtracted the total lesion uptake will be identified. Partial volume effect on

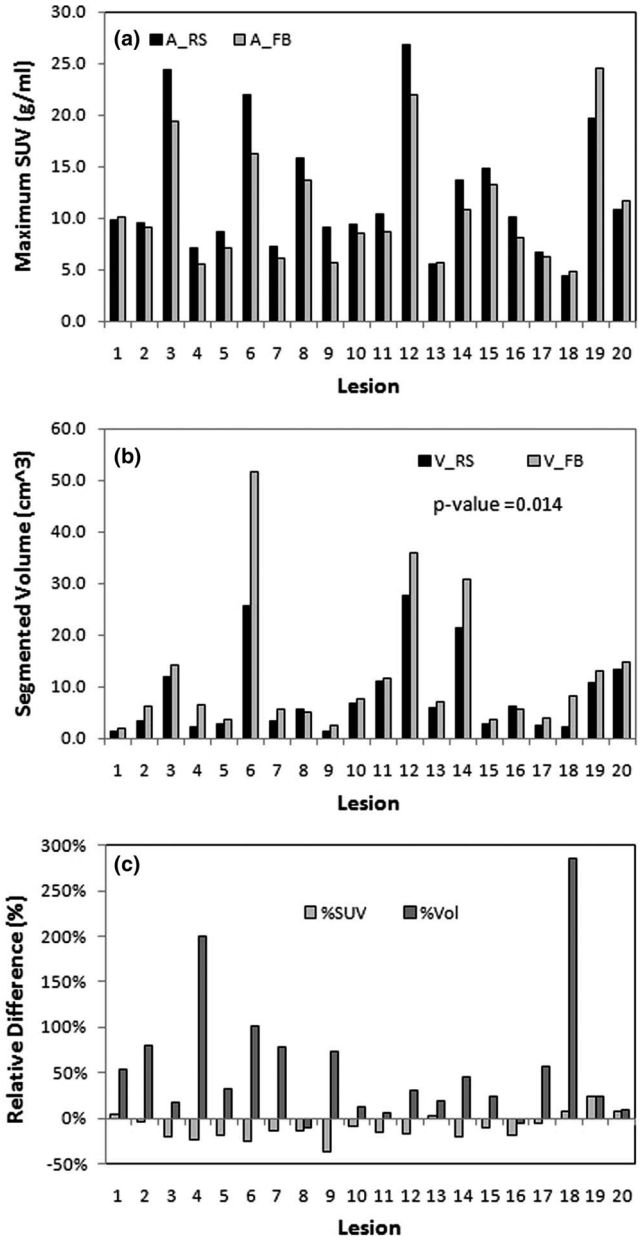

FIG. 3. Tumor activity (a), segmented volume (b), and percentage difference (c) between FB and RS using TBS. Tumor volumes in FB and RS are significantly different (47% ± 52%, $p = 0.01$). In FB, the activity (max or mean SUV) is smaller and the volume is greater, under estimating the activity and significantly overestimating the lesion size.

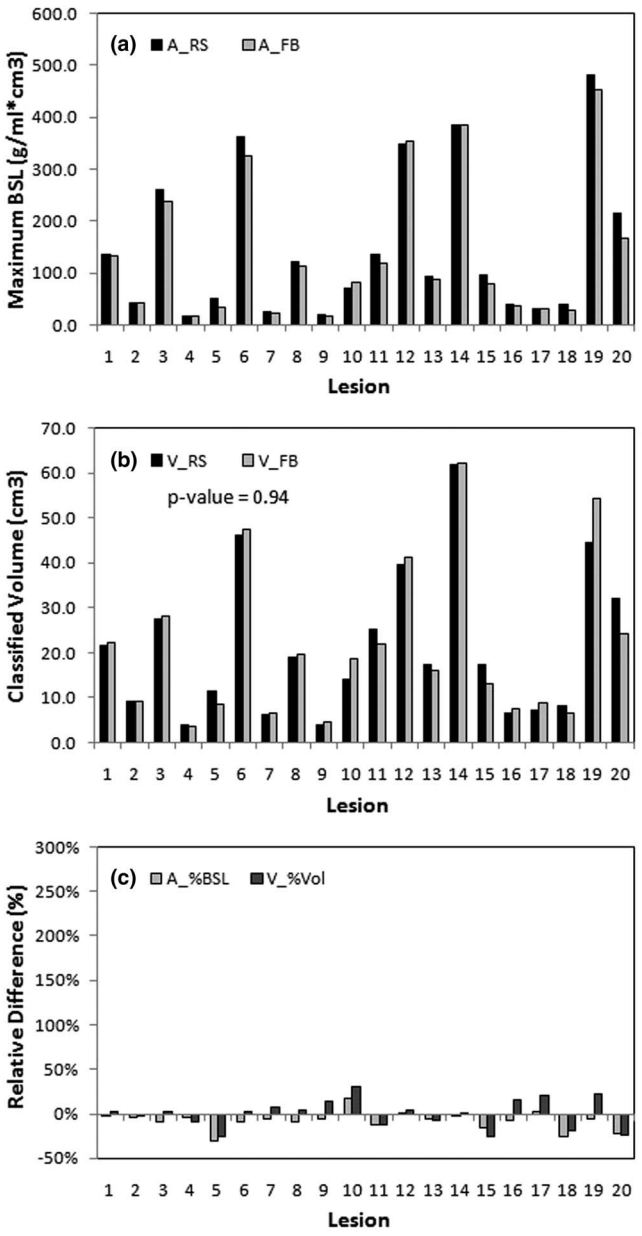

FIG. 4. Tumor activity (a), estimated volume (b), and percentage difference (c) between FB and RS using HBE. The difference in tumor volumes between FB and RS is not significant (0% ± 16%, $p = 0.94$). Both the activity and the volume are similar in FB and RS, %Diff in volume is reduced by one order of magnitude with little systematic bias in contrast to TBS result [Fig. 3(c)].

the clear or blurred edge of the lesion can alter the background distribution to a certain degree, contaminating the high-activity wing of the Gaussian background and therefore affecting the Gaussian fit and cutoff estimate (mean + 2σ). The adoption of the watershed segmentation algorithm in the BSL method provides a means to reduce the dependence of the fit on the low-activity wing of the Gaussian background. Furthermore, the mode of the distribution provides a lesion-specific background level that is determined automatically. After the Gaussian fit to the background and its subtraction from the uptake in the histogram of the VOI, the sum of the remaining signals is representative of the lesion's total uptake. In a VOI, the total uptake of a lesion should be conserved with or without motion and the apparent difference is mostly resulted from redistribution of the activity in space between FB and RS. The increase in average activity by 8% from RS to FB [Fig. 4(a) and Table III] may result from the physiological increase in activity in the liver tumors over time,[32] since RS scans were done 17 ± 17 min after the FB scans (Table I). The system-dependent partial volume effect occurs in both FB and RS PET images, while the attenuation correction bias affects FB PET, not RS.

As stated above, the uptake activities of the lesion with and without motion should be conserved, except for small

activity blurred into the local background in FB and small tumor uptake gain in RS during the time interval after FB. This conservation could be used as a simple constraint for a segmentation algorithm. While this can be applied to any segmentation method, when applied to threshold segmentation an adaptive threshold for both FB and RS PET can be established via an uptake equivalent threshold, i.e., the threshold that produces the same total uptake. Therefore, the higher threshold for lesion segmentation in FB tends to cancel out the volume increase effect in FB. In fact, Table III and Fig. 4 demonstrate a close match of the lesion activities and volumes between RS and FB using BSL to determine the adaptive threshold, supporting the uptake conservation hypothesis. The difference in lesion volume is not of significance between FB and RS PET images (Table III). Due to partial volume effect, absolute lesion volume is expected to be PET-resolution dependent. One would reasonably assume that additional partial volume effects due to motion blurring would cause a mismatch between the volumes estimated for RS and FB, however similar volume estimates between the two remain. Our results show that the volume difference has $\sigma = 16\%$, which is relatively high due to the partial volume effect and the cubic error-propagation from threshold to volume determination. BSL difference has $\sigma = 10\%$, which is not far from $\pm 10\%$ from a Monte Carlo simulation study.[24] The lesion volume conservation effect can be interpreted as the adaptive capability of the BSL to the background levels. The BSL cutoff thresholds in Table III are lesion-specific and have a bias to compensate for the motion smearing effect–the difference between FB and RS is $19\% \pm 17\%$ on average. The implication of this result is that the viable lesion volume can be assessed using FB PET image with reasonably accurate and consistent estimation. While this method does not purport to be a "gold standard" in viable lesion volume assessment, it does provide a consistent volumetric estimation based on a statistical model of the local background activity distribution within the VOI. Therefore, this is potentially useful in evaluating viable tumor volume change using PET imaging over the treatment course.

4.D. Limitations and future investigations

The percentage activity and volume variation between RS and FB from BSL method is still considered substantial ($0\% \pm 16\%$ and $-8\% \pm 10\%$) and have room for further improvement. Part of the reason for this is the variation of the manual definition of initial VOI for BSL calculation. Once the VOI definition process is completely automated and shared by both RS and FB, some uncertainty will be further removed and more-consistent results are expected. For target volume determination, the histogram-based method needs further investigation to establish a physical contour around the high FDG uptake region, for any clinical application that requires target volume delineation for treatment planning, such as radiotherapy. Using this histogram approach, estimated volume is determined and some high uptake voxels may not necessarily be located within one topologically confined region. The authors wish to underscore that the use of the term "tumor volume" in the context of this study refers to a volume for which the activity exceeds that of the background normal tissue (in this case liver). Because of this, any classification/segmentation beyond these estimates, including the spatial location of tumor boundaries, requires the use of an additional model to perform that task.

Although, the difference in BSL activities ($-8\% \pm 10\%$, in Table III) becomes smaller between FB and RS, comparing with that of TLG ($18\% \pm 36\%$, in Table II) from the TBS method, it is still statistically significant. Two effects would contribute the -8% higher BSL activity in RS than FB PET: a smaller activity decrease in FB due to motion blurring (becomes local background) and a small activity increase in RS due to further physiologically accumulation of tracer into the tumor during the time gap after FB scan. The histogram-based BSL method determines the appropriate background around a lesion to calculate the lesion activity and volume. However, it is beyond the capability of the BSL method to estimate the SUV_{max} difference between FB and RS due to motion blurring.

5. CONCLUSION

A clinical near-motion-free PET "ground truth" of patient liver lesions has been established using general anesthesia to temporarily suspend patient respiration on a ventilator and achieve a reproducible assessment of a target volume from PET images. Using PET/CT images obtained in this way allows the effects of motion-free breathing upon tumor volume estimation to be assessed. In this study free-breathing PET images show a delineated volume increase by $47\% \pm 52\%$ ($p = 0.01$) relative to RS when a 50% threshold was used. Similar motion differences would be anticipated by any threshold-based method. In contrast, the BSL method provides a conserved target volume with insignificant differences in target volumes between RS and FB PET images ($0\% \pm 16\%$, $p = 0.94$). This result suggests that viable lesion volume could be estimated using free-breathing PET images with reasonable accuracy and reliability. In addition, these results suggest that the BSL method can act as a reasonable constraint or penalty for segmentation tasks even in the presence of motion.

ACKNOWLEDGMENTS

The authors are grateful to participating physicians, residents, and technologists for assisting in image acquisition, data collection, and use of commercial software.

[a]Author to whom correspondence should be addressed. Electronic mail: lig2@mskcc.org
[b]Present address: Department of Radiology Mater Misericordiae, University Hospital, Eccles Street, Dublin 7, Ireland.

[1]S. A. Nehmeh and Y. E. Erdi, "Respiratory motion in positron emission tomography/computed tomography: A review," Semi. Nucl. Med. **38**, 167–176 (2008).
[2]G. Li, D. Citrin, K. Camphausen, B. Mueller, C. Burman, B. Mychalczak, R. W. Miller, and Y. Song, "Advances in 4D medical imaging and 4D radiation therapy," Technol. Cancer Res. Treat **7**, 67–81 (2008).
[3]S. A. Nehmeh, Y. E. Erdi, T. Pan, A. Pevsner, K. E. Rosenzweig, E. Yorke, G. S. Mageras, H. Schoder, P. Vernon, O. Squire, H. Mostafavi, S. M.

Larson, and J. L. Humm, "Four-dimensional (4D) PET/CT imaging of the thorax," Med. Phys. **31**, 3179–3186 (2004).
[4]S. A. Nehmeh, Y. E. Erdi, T. Pan, E. Yorke, G. S. Mageras, K. E. Rosenzweig, H. Schoder, H. Mostafavi, O. Squire, A. Pevsner, S. M. Larson, and J. L. Humm, "Quantitation of respiratory motion during 4D-PET/CT acquisition," Med. Phys. **31**, 1333–1338 (2004).
[5]S. J. Park, D. Ionascu, J. Killoran, M. Mamede, V. H. Gerbaudo, L. Chin, and R. Berbeco, "Evaluation of the combined effects of target size, respiratory motion and background activity on 3D and 4D PET/CT images," Phys. Med. Biol. **53**, 3661–3679 (2008).
[6]T. Yamamoto, U. Langner, B. W. Loo, Jr., J. Shen, and P. J. Keall, "Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients," Int. J. Radiat. Oncol. Biol. Phys. **72**, 1250–1258 (2008).
[7]F. Qiao, T. Pan, J. W. Clark, and O. R. Mawlawi, "Region of interest motion compensation for PET image reconstruction," Phys. Med. Biol. **52**, 2675–2689 (2007).
[8]W. T. Watkins, R. Li, J. Lewis, J. C. Park, A. Sandhu, S. B. Jiang, and W. Y. Song, "Patient-specific motion artifacts in 4DCT," Med. Phys. **37**, 2855–2861 (2010).
[9]J. H. Killoran, V. H. Gerbaudo, M. Mamede, D. Ionascu, S. J. Park, and R. Berbeco, "Motion artifacts occurring at the lung/diaphragm interface using 4D CT attenuation correction of 4D PET scans," J. Appl. Clin. Med. Phys. **12**, 3502 (2011).
[10]H. Li, C. Noel, J. Garcia-Ramirez, D. Low, J. Bradley, C. Robinson, S. Mutic, and P. Parikh, "Clinical evaluations of an amplitude-based binning algorithm for 4DCT reconstruction in radiation therapy," Med. Phys. **39**, 922–932 (2012).
[11]S. A. Nehmeh, A. A. Haj-Ali, C. Qing, C. Stearns, H. Kalaigian, S. Kohlmyer, H. Schoder, A. Y. Ho, S. M. Larson, and J. L. Humm, "A novel respiratory tracking system for smart-gated PET acquisition," Med. Phys. **38**, 531–538 (2011).
[12]F. Buther, I. Ernst, J. Hamill, H. T. Eich, O. Schober, M. Schafers, and K. P. Schafers, "External radioactive markers for PET data-driven respiratory gating in positron emission tomography," Europ. J. Nucl. Med. Mol. Imaging **40**, 602–614 (2013).
[13]L. Fin, J. Daouk, P. Bailly, J. Slama, J. Morvan, I. El Esper, J. M. Regimbeau, D. Chatelain, M. Diouf, and M. E. Meyer, "Improved imaging of intrahepatic colorectal metastases with 18F-fluorodeoxyglucose respiratory-gated positron emission tomography," Nucl. Med. Commun. **33**, 656–662 (2012).
[14]T. Li, B. Thorndyke, E. Schreibmann, Y. Yang, and L. Xing, "Model-based image reconstruction for four-dimensional PET," Med. Phys. **33**, 1288–1298 (2006).
[15]F. Qiao, T. Pan, J. W. Clark, Jr., and O. R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies," Phys. Med. Biol. **51**, 3769–3783 (2006).
[16]M. F. Kruis, J. B. van de Kamer, A. C. Houweling, J. J. Sonke, J. S. Belderbos, and M. van Herk, "PET motion compensation for radiation therapy using a CT-based mid-position motion model: Methodology and clinical evaluation," Int. J. Radiat. Oncol., Biol., Phys. **87**, 394–400 (2013).
[17]K. K. Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," Int. J. Radiat. Oncol., Biol., Phys. **76**, 583–596 (2010).
[18]S. A. Nehmeh, Y. E. Erdi, G. S. Meirelles, O. Squire, S. M. Larson, J. L. Humm, and H. Schoder, "Deep-inspiration breath-hold PET/CT of the thorax," J. Nucl. Med. **48**, 22–26 (2007).
[19]S. Nagamachi, H. Wakamatsu, S. Kiyohara, S. Fujita, S. Futami, H. Arita, S. Tamura, and K. Kawai, "The reproducibility of deep-inspiration breath-hold (18)F-FDG PET/CT technique in diagnosing various cancers affected by respiratory motion," Annls. Nucl. Med. **24**, 171–178 (2010).
[20]T. Kawano, E. Ohtake, and T. Inoue, "Deep-inspiration breath-hold PET/CT versus free breathing PET/CT and respiratory gating PET for reference: evaluation in 95 patients with lung cancer," Annls. Nucl. Med. **25**, 109–116 (2011).
[21]D. M. Lovelock, P. Wang, W. Dalecki, J. Zatcky, and J. Yamada, "Spatial accuracy of a novel technique to position a liver tumor for high-dose single-fraction radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **78**, S139 (2010).
[22]E. R. Ryan, C. T. Sofocleous, H. Schoder, J. A. Carrasquillo, S. Nehmeh, S. M. Larson, R. Thornton, R. H. Siegelbaum, J. P. Erinjeri, and S. B. Solomon, "Split-dose technique for FDG PET/CT-guided percutaneous ablation: A method to facilitate lesion targeting and to provide immediate assessment of treatment effectiveness," Radiology **268**, 288–295 (2013).
[23]H. Zaidi and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques," Europ. J. Nucl. Med. Mol. Imaging **37**, 2165–2187 (2010).
[24]S. A. Nehmeh, H. El-Zeftawy, C. Greco, J. Schwartz, Y. E. Erdi, A. Kirov, C. R. Schmidtlein, A. B. Gyau, S. M. Larson, and J. L. Humm, "An iterative technique to segment PET lesions using a Monte Carlo based mathematical model," Med. Phys. **36**, 4803–4809 (2009).
[25]I. A. Burger, H. A. Vargas, A. Apte, J. L. Humm, M. Gonen, S. M. Larson, and C. R. Schmidtlein, "A novel PET quantification method determining the background subtracted lesion activity with simple histogram analysis is superior to total lesion glycolysis," J. Nucl. Med. **54**, 316P (2013).
[26]I. A. Burger, H. A. Vargas, A. Apte, B. J. Beattie, J. L. Humm, M. Gonen, S. M. Larson, and C. R. Schmidtlein, "PET quantification with a histogram derived total activity metric: Superior quantitative consistency compared to total lesion glycolysis with absolute or relative SUV thresholds in phantoms and lung cancer patients," Nucl. Med. Biol. **41**, 410–418 (2014).
[27]D. Freedman and P. Diaconis, "On the histogram as a density estimator – L2 theory," Z. Wahrsch. **57**, 453–476 (1981).
[28]W. Sureshbabu and O. Mawlawi, "PET/CT imaging artifacts," J. Nucl. Med. Technol. **33**, 156–161 (2005).
[29]V. Bettinardi, M. Picchio, N. Di Muzio, and M. C. Gilardi, "Motion management in positron emission tomography/computed tomography for radiation treatment planning," Sem. Nucl. Med. **42**, 289–307 (2012).
[30]Y. E. Erdi, O. Mawlawi, S. M. Larson, M. Imbriaco, H. Yeung, R. Finn, and J. L. Humm, "Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding," Cancer **80**, 2505–2509 (1997).
[31]C. Liu, L. A. Pierce, 2nd, A. M. Alessio, and P. E. Kinahan, "The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging," Phys. Med. Biol. **54**, 7345–7362 (2009).
[32]I. A. Burger, C. Burger, T. Berthold, and A. Buck, "Simplified quantification of FDG metabolism in tumors using the autoradiographic method is less dependent on the acquisition time than SUV," Nucl. Med. Biol. **38**, 835–841 (2011).