Validating PET segmentation of thoracic lesions-is 4D PET necessary?

Nielsen, M. S.; Carl, J.

Published in:
Biomedical Physics & Engineering Express

DOI (link to publication from Publisher):
10.1088/2057-1976/aa5ba9

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Biomedical Physics & Engineering Express

ENDORSED BY IPEM Institute of Physics and Engineering in Medicine

PAPER • OPEN ACCESS

Validating PET segmentation of thoracic lesions—is 4D PET necessary?

To cite this article: Martin S Nielsen and Jesper Carl 2017 *Biomed. Phys. Eng. Express* **3** 015028

Biomedical Physics & Engineering Express

PAPER

Validating PET segmentation of thoracic lesions—is 4D PET necessary?

Martin S Nielsen[1] **and Jesper Carl**[2]

[1] Department of Medical Physics, Aalborg University Hospital, Denmark
[2] Department of Clinical Medicine, Aalborg University, Denmark

E-mail: martin.skovmos.nielsen@rn.dk

Keywords: FDG-PET, segmentation, lung cancer

OPEN ACCESS

RECEIVED
18 October 2016

REVISED
9 January 2017

ACCEPTED FOR PUBLICATION
24 January 2017

PUBLISHED
16 February 2017

Abstract

Respiratory-induced motions are prone to degrade the positron emission tomography (PET) signal with the consequent loss of image information and unreliable segmentations. This phantom study aims to assess the discrepancies relative to stationary PET segmentations, of widely used semi-automatic PET segmentation methods on heterogeneous target lesions influenced by motion during image acquisition. Three target lesions included dual F-18 Fluoro-deoxy-glucose (FDG) tracer concentrations as high- and low tracer activities relative to the background. Four different tracer concentration arrangements were segmented using three SUV threshold methods (Max40%, SUV40% and 2.5SUV) and a gradient based method (GradientSeg). Segmentations in static 3D-PET scans (PET_{sta}) specified the reference conditions for the individual segmentation methods, target lesions and tracer concentrations. The motion included PET images followed a 4D-PET (PET_{4D}) and a 3D-PET (PET_{mot}) scan protocol. Moreover, motion-corrected PET images (PET_{deb}) were derived from the PET_{mot} images. Segmentations in PET_{4D}, PET_{mot} and PET_{deb} were compared to the PET_{sta} segmentations according to volume changes (ΔVol) and an error estimate ($lowUptake_{error}$) for the lesion part covering the low tracer concentration. In PET_{4D} images, all segmentation methods provided $lowUptake_{error}$ estimates equivalent to PET_{sta} segmentations and, except for the Max40% segmentations, a slight volume expansion. In the PET_{mot} images, the GradientSeg method results in an average 0.43 increased volume and an overestimation of 0.33 for the $lowUptake_{error}$. The most accurate segmentations in PET_{mot}, relative to PET_{sta}, were accomplished by the 2.5SUV and SUV40% methods. In the PET_{deb} images, the GradientSeg method solitary provided segmentations equivalent to segmentation in PET_{sta} images. The use of FDG with various tracer concentrations revealed, according to PET_{sta} images, that the most constant segmentations for motion-corrected PET images (PET_{4D} or PET_{deb}) were achieved using the GradientSeg method. In the absence of PET_{4D} or PET_{deb} images, the 2.5SUV and SUV40% methods are most consistent to PET_{sta} segmentations.

1. Introduction

In general, lung cancer is classified into two main categories, defined as small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). The majority and minority of all lung cancers are classified as NSCLC (85%) and SCLC (15%) respectively [1]. For both SCLC and the NSCLC, radiotherapy is one treatment strategy depending of the extent of the cancer as indicated by lymph node involvement and potential metastases according to the TNM definition [2, 3]. The hybrid positron emission tomography (PET) and computed tomography (CT) and recently introduced PET and magnetic resonance imaging scanners benefits a minimal misregistration between image acquisitions [4, 5]. The combined use of F-18 Fluoro-deoxy-glucose (FDG) PET and CT provide high accuracy for the clinical staging of lung cancers [6–8] and lower inter-observer variation for target definition [9]. A high FDG uptake was associated with the preferential site of local relapse of NSCLC in a recent study by Calais *et al* [10]. Respiratory motion may result in motion artefacts, thus requiring motion encompassing images in PET/CT scans (4D-PET/

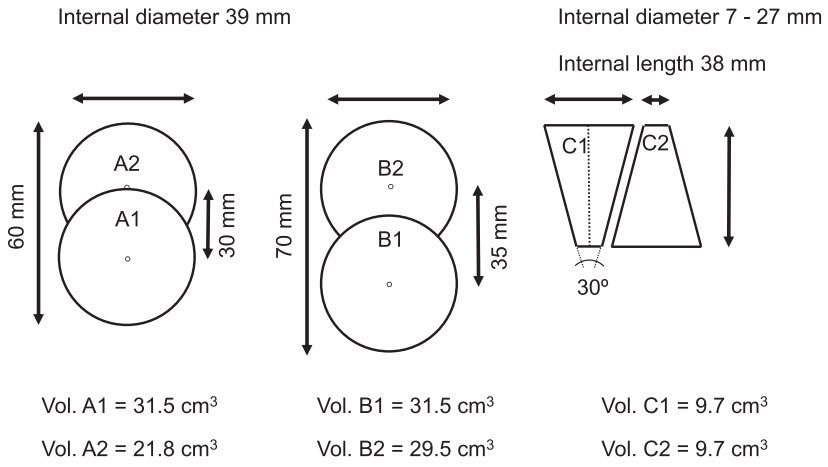

Figure 1. Target volumes (A, B and C) with dimensions and the actual internal volume size (cm^3). The target high sub-volumes are denoted by one (A1, B1 and C1) and the target low sub-volume as two (A2, B2 and C2). Targets A and B consisted of interlaced spheres, whereas the target C involved two cones aligned side-by-side in opposing direction.

CT). The 4D-PET/CT scan benefits by nearly motion free images by selecting an appropriate number of respiratory phase bins and thus, it may further improve cancer staging and target volume definition in radiotherapy [11]. The use of semi-automatic delineation (segmentation) methods may reduce the inter-observer variation in the clinical target volume (CTV) outline for radiotherapy [12, 13]. Some PET segmentation methods for outlining CTV are intensity-based, such as the standard uptake value (SUV) or the fixed/adaptive threshold activity value. Other more sophisticated methods may be established using activity statistics, resulting in segmented clusters or the use of spatial gradients in activity for segmenting boundaries [14, 15]. Currently, no consensus exists on methods for FDG-PET segmentation in lung cancer, but a threshold of 2.5 SUV or a percentage of the maximum SUV (SUV_{max}) such as 40%–50% is widely used [7, 8, 16]. Using a fixed SUV threshold for target determination still result in variations in target definition due to partial volume effect, activity recovery and FDG uptake heterogeneities [17, 18]. For PET/CT imaging of lung cancer, the effect of respiratory motion may additionally result in image intensity changes or eroded intensity gradients. Consequently, respiratory motion could result in less accuracy in established segmentation methods and therefore in uncertainty of CTV definition.

The main purpose of this work was to construct a phantom enclosing various lesions with heterogeneous PET tracer activities and to assess the effect of motion on commonly used semi-automatic PET segmentation methods according to equivalent segmentations in stationary PET images. A secondary purpose was to study the effect of commonly used segmentation methods on post-processed motion corrected 3D-PET images created by deconvolution of motion involved 3D-PET images with the target position density derived from a 4D-CT scan.

2. Materials and methods

2.1. Equipment

The target volumes consisted of three in-house manufactured interlaced double volumes, titled A, B and C submerged in a water-filled phantom of size 17 cm × 17 cm × 18 cm (figure 1). Each of the three target volumes involved high and low PET tracer uptake sub-volumes (highUptake = A1, B1 and C1 and lowUptake = A2, B2 and C2), allowing for different PET tracer uptake ratios relative to a uniform background activity. The target size of volume A and B correspond to a simulation of lung tumour up to stage T3, whereas the target C correspond to a simulation of lymph node [3]. The water-filled phantom with target volumes was connected to a QUASAR^{TM} Respiratory Motion Phantom (ModusQA Medical Devices), as illustrated in figure 2. In static mode, no motion was applied to the phantom, whereas in motion mode, a 15 mm peak-to-peak amplitude (nearly sine profile), 15 cycles per minute, orientated in the scanner longitudinal z-direction was applied to the phantom.

A total FDG PET tracer activity of 80–120 MBq ($Injected_{activity}$) was prepared for the phantom with four different FDG concentrations (ratios) relative to the background activity of SUV = 1. The SUV definition for this study was corrected for FDG activity decay at acquisition time (t) as the following:

$$SUV = Phantom_{weight}[g]\frac{Voxel_{activity}(t)[Bq\ ml^{-1}]}{Injected_{activity}(t)[Bq]} \quad (1)$$

with $Phantom_{weight}$ defined as background water and target volume materials.

The target volume uptake ratios (uptakeRatio) were organised as 4:4:1, 3:2:1, 4:2:1 and 8:3:1, where for instance, a ratio of 4:2:1 represented a highUptake SUV = 4, a lowUptake SUV = 2 and a background SUV = 1. The ratio of 4:4:1 indicated equal uptake

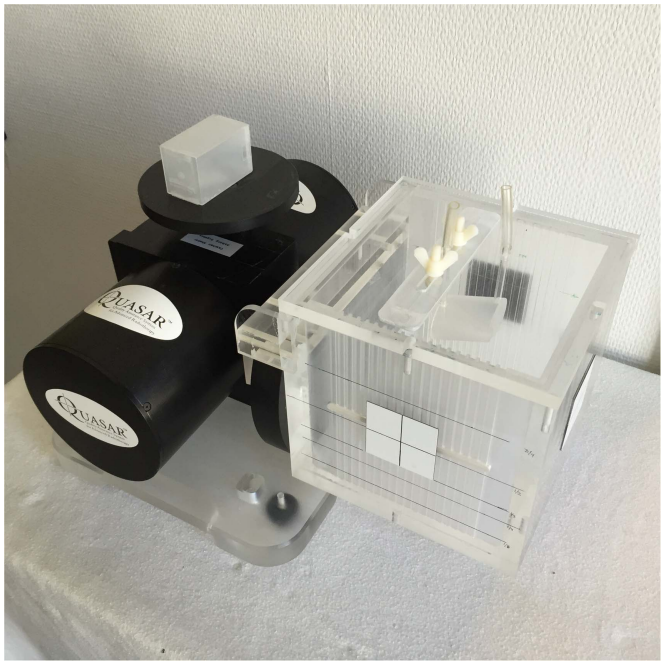

Figure 2. Measurement setup with water phantom holding target lesions, mounted on a QUASAR^{TM} Respiratory Motion Phantom, to simulate motion during image acquisitions. The target inserts were built-in to the top plate of the phantom and thereby changed between the different target setup.

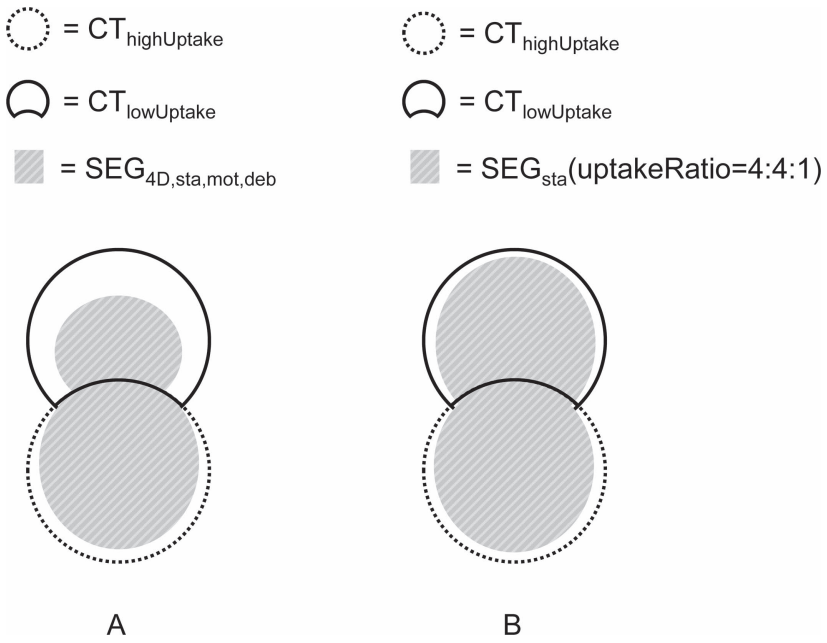

Figure 3. Relationship between the PET segmented volume $SEG_{4D,sta,mot,deb}$ and the lowUptake reference volume $CT_{lowUptake}$ used to define the $lowUptake_{error}$ quantity. Sub-figure (A) illustrates a situation where the lowUptake sub-volume was partially segmented and in sub-figure (B), a homogeneous tracer uptakeRatio = 4:4:1 illustrated an optimal lowUptake segmentation.

(SUV = 4) for both high- and lowUptake sub-volumes. The highUptake sub-volumes (A1, B1 and C1) were included in every PET segmentations, whereas the lowUptake sub-volumes (A2, B2 and C2) were estimated to be segmented in approximately half of the uptake ratios as labelled in table 1.

2.2. Imaging protocols

For each instance of FDG uptake ratio, the individual target volumes were subjected to scan protocols in both static- and motion-modes using an integrated PET/CT scanner (Discovery STE, GE Healthcare).

Table 1. SUV threshold values.

UptakeRatio	4:4:1	3:2:1	4:2:1	8:3:1
Max40%	*1.6*	*1.2*	*1.6*	**3.2**
SUV40%	*2.2*	*1.8*	**2.2**	**3.8**
GradientSeg	*NA*	*NA*	**NA**	**NA**
2.5 SUV	*2.5*	**2.5**	**2.5**	*2.5*

SUV threshold values among the tracer uptakeRatios and segmentation methods. Grey cells indicate the predicted segmentation of the highUptake volume only, whereas white cells indicate predicted segmentation of both the high- and lowUptake volumes. The non-threshold method GradientSeg was marked not available. The GradientSeg low/high-Uptake separation were predicted assuming accurate PET_{static} segmentation. For the uptakeRatio of 4:2:1 and 8:3:1, combined with GradientSeg in PET_{static} images, only the highUptake volume were segmented.

2.2.1. Static/motion 3D-PET/3D-CT

The 3D-PET scans were acquired using a 2 min acquisition time for a single bed position. In static mode the 3D-PET were labelled PET_{sta} and with phantom motion enabled PET_{mot}. The PET scans were acquired in 3D mode and reconstructed to matrix size of $256 \times 256 \times 47$ pixels (voxel size $2.73 \times 2.73 \times 3.27\,mm^3$) with a 3D-OSEM algorithm (2 iterations, 26 subsets) and a post-Gaussian filter of 3 mm full-width-half-maximum. The corresponding CT scans were acquired in helical mode at 120 kVp, 80 mAs. The CT scan length was marginally beyond the phantom volume of 17 cm plus the respiratory motion extent. The 3D-CT were labelled CT_{sta} in static mode and with phantom motion enabled CT_{mot}. The CT scans were reconstructed with a CT slice thickness of 1.25 mm and a matrix size of 512×512 pixels (voxel size $0.977 \times 0.977 \times 1.25\,mm^3$). The CT_{sta}/CT_{mot} were used for attenuation correction of the corresponding PET_{sta}/PET_{mot} images.

2.2.2. 4D-PET/4D-CT

The 4D-PET (PET_{4D}) were acquired using a total PET acquisition time of 12 min, for 6 PET bins (2 min per phase bin). The number of phase bin was carefully chosen to obtain nearly motion free PET images. Given the motion amplitude and target size, Bettinardi *et al* proposed the number of bins to obtain motion-free images in 4D-PET according to motion-amplitude and target sizes [11]. The 4D-CT (CT_{4D}) were acquired in cine axial mode at 120 kV, 80 mAs with a CT gantry rotation time of 0.5 s, cine time between images of 0.5 s and a total cine duration of 5 s. The 4D-CT was retrospectively re-sorted into 6 phase bins (maximum phase error 5%), using respiratory motion data obtained by the real-time position management system (RPM from Varian Medical System). Equivalent to 3D-CT, the CT images were reconstructed in 1.25 mm slice thickness (512×512 pixels). The matching CT phase bin from CT_{4D} were used for attenuation correction of the individual PET bins in PET_{4D}.

2.2.3. Motion-correction

In addition, a motion corrected post-processing procedure (motion deblurring) applied to the PET_{mot} scans resulted in PET images labelled PET_{deb}. The following section provides a more detailed description of the motion-correction procedure.

The motion-corrected PET_{deb} were iteratively reconstructed from the PET_{mot}, comparable to the method described by [19]. The PET_{deb} volumes were derived by deconvolving PET_{mot} with a target position probability kernel (K_{motion}), estimated from the 4D-CT. The K_{motion} was derived using a normalised position probability kernel estimated from the CT target centre-of-mass position within the 4D-CT. Using the phantom trajectory of motion aligned in the superior-inferior (z-axis), the main contribution to the position probability density was coupled to this direction. The 4D-CT acquired with the same FDG uptake ratio was used for the K_{motion} determination using a rigid target propagation from the CT phase bin 0% throughout the remaining five phases. A forward Van-Cittert iterative deconvolution method was performed as following:

$$PET_{deb}^{k+1} = PET_{deb}^{k} + [PET_{mot} - PET_{deb}^{k} \otimes K_{motion}] \quad (2)$$

with the initial condition $PET_{deb}^{k=0} = PET_{mot}$. For simplification, the deconvolution was applied in 2D space slice by slice, thus the PET_{mot} was resliced within the primary 2D motion plane (x, z). Succeeding each iteration a post-processing image correction was applied, limit the PET_{deb}^{k+1} within the original PET_{mot} minimum and maximum values within a 4 mm radius, thereby reducing Gibb's phenomenon. A visualisation of the PET_{sta}, PET_{mot} and PET_{deb} is provided for the uptakeRatio 8:3:1 in figure 4.

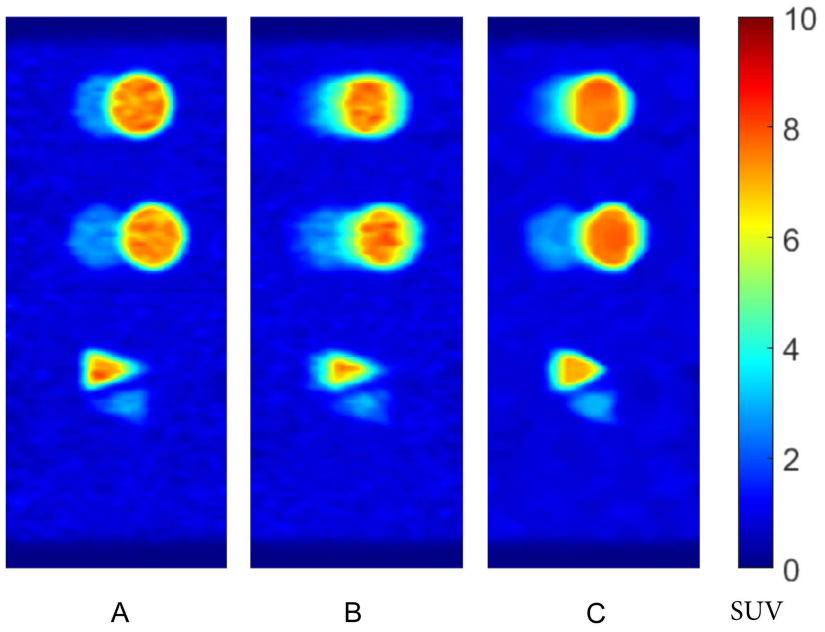

Figure 4. The three target lesions, visualised through the central image plane, for the uptakeRatio of 8:3:1 for (A) stationary PET_{sta} images, (B) motion included PET_{mot} images and (C) motion corrected PET_{deb} images.

2.3. Image registration

The PET and corresponding CT scans were co-registered using the common frame of reference (DICOM origin) from the PET/CT scan. For the individual uptakeRatio a sequence of PET registrations were organised between PET_{sta}, PET_{4D} and PET_{mot}/PET_{deb} as following: a registration from PET_{sta} to the PET_{4D} mid-ventilation (bin3) and the PET_{4D} (bin3) to PET_{mot}/PET_{deb} using the complete PET intensity range for registration with a roughly 2 cm isotropic margin covering the lesions as a volume of interest.

2.4. Segmentation

Four different PET segmentation methods were applied to every PET images, PET_{4D}(6 bins), PET_{sta}, PET_{mot} and PET_{deb} for each of the three experimental phantom volumes A–C.

I. Max40%: a fixed threshold at 40% of the maximum SUV value defined as the mean SUV over 1 cm^2 around the peak intensity pixel. Voxels above the threshold value were segmented as the target volume.

II. SUV40%: threshold of 40% maximum SUV as described above and corrected for the background level (SUV_{bg}). The segmentation included voxels beyond the threshold value defined as

$$SUV40\% = 0.40(SUV_{max} - SUV_{bg}) + SUV_{bg}. \quad (3)$$

III. GradientSeg: gradient-based watershed segmentation. A segmentation process developed by Geets *et al* [15], named IMREviewer. Segmentation with IMREviewer presented robust accuracy for lung cancer, in comparison with surgical specimens [7, 20]. The IMREviewer prepared the PET images by initially denoising and later deblurring with the PET scanner FWHM (5.4 mm) kernel, defined by a Gaussian approximation to a point spread function measured for the scanner central axis.

IV. 2.5 SUV: threshold of 2.5 times the SUV. The segmentation included voxels with an uptake value above 2.5 SUV.

In the static CT (CT_{sta}) images, the high- ($CT_{highUptake}$) and lowUptake ($CT_{lowUptake}$) sub-volumes were segmented individually using geometric 3D-structures aligned in the image according to the lesions walls. For target volumes A and B, a 3D-sphere of 39 mm (internal sphere diameter) was used for segmentations, followed by a Boolean operation to distinct the $CT_{highUptake}$ and $CT_{lowUptake}$ sub volumes. For the cone shaped target C, the first and last CT slice containing the target was segmented as 2D planar circles (diameter of 7 and 27 mm) connected by a linear interpolation. Through the image registrations, the $CT_{lowUptake}$ segmentations were transferred to PET_{4D} (bin3), PET_{mot} and PET_{deb}. A rigid alignment of $CT_{lowUptake}$ throughout the PET_{4D} were completed manually for the remaining bins.

2.5. Data analysis

The PET segmented volumes in the different scan protocols ($SEG_{4D,mot,deb}$) were compared to the static segmentations (SEG_{sta}) according to volume change (ΔVol) and an error estimate of the lowUptake sub-volume ($lowUptake_{error}$) segmentation.

Table 2. ΔVol of the PET segmentations.

Method	PET_{4D}	PET_{mot}	PET_{deb}
Max40%	0.01 [−0.01; 0.03]	0.15 [0.11; 0.19]	0.10 [0.06; 0.14]
SUV40%	0.02 [0.01; 0.03]	0.06 [0.00; 0.12]	0.06 [0.01; 0.12]
GradientSeg	0.04 [0.00; 0.07]	0.43 [0.12; 0.74]	0.04 [−0.11; 0.20]
2.5 SUV	0.04 [0.02; 0.05]	−0.04 [−0.11; 0.02]	−0.04 [−0.08; 0.00]

Mean with 95% CI segmented volume differences of the PET_{4D} (all 6 bins), PET_{mot} and PET_{deb} protocols separated into segmentation methods, including all tracer uptakeRatios and target volumes.

Table 3. The $lowUptake_{error}$ metric of the PET segmentations.

Method	PET_{4D}	PET_{mot}	PET_{deb}
Max40%	0.00 [−0.05; 0.04]	0.09 [−0.02; 0.20]	0.10 [0.00; 0.20]
SUV40%	0.04 [−0.01; 0.08]	0.12 [0.00; 0.24]	0.13 [−0.01; 0.28]
GradientSeg	0.01 [−0.03; 0.06]	0.33 [0.08; 0.58]	0.08 [−0.04; 0.21]
2.5 SUV	0.00 [−0.03; 0.04]	0.03 [−0.06; 0.13]	0.07 [−0.03; 0.17]

Mean with 95% CI of the $lowUptake_{error}$ of the PET_{4D} (all 6 bins), PET_{mot} and PET_{deb} protocols separated between segmentation methods, including all tracer uptakeRatios and target volumes (lowUptake sub volume only).

The segmented volume change was defined as the difference between PET segmentation relative to static segmentation for an equivalent segmentation method, target and tracer uptakeRatio.

$$\Delta\text{Vol} = \frac{\text{SEG}_{4D,mot,deb}[\text{cm}^3] - \text{SEG}_{sta}[\text{cm}^3]}{\text{SEG}_{sta}[\text{cm}^3]}. \quad (4)$$

The $lowUptake_{error}$ was defined by the volume similarity in terms of dice similarity coefficient (DSC) between PET segmentations and the $CT_{lowUptake}$ volumes labelled A2, B2 and C2 in figure 1.

I. $DSC_{lowUptake}$ as spatial volume similarity between the PET segmented volume and the $CT_{lowUptake}$ for the specific segmentation methods, target and uptakeRatio.

$$DSC^{sta,\,4D,mot,deb}_{lowUptake} = \frac{2\,|\,\text{SEG}_{sta,4D,mot,deb} \cap CT_{lowUptake}\,|}{|\,\text{SEG}_{sta,4D,mot,deb}\,| + |\,CT_{lowUptake}\,|}. \quad (5)$$

An illustration of the volume similarity between the PET segmentation and the lowUptake sub-volume is provided in figure 3. As the PET segmented volumes extent outside the lowUptake sub-volume, the $DSC_{lowUptake}$ never range to the maximum value of one, as $0 \leqslant DSC_{lowUptake} < 1$.

II. $lowUptake_{error}$, a quantity metric quantifying the difference between the $DSC_{lowUptake}$ in equation (5) and the $DSC_{lowUptake}$ in the static PET_{sta} image, relative to the $DSC_{lowUptake}$ measured with equal tracer uptake in both high- and lowUptake sub-volumes, for the matching segmentation method and target.

$$\text{lowUptake}_{error} = \frac{DSC^{4D,mot,deb}_{lowUptake} - DSC^{sta}_{lowUptake}}{DSC^{sta}_{lowUptake}(\text{uptakeRatio} = 4{:}4{:}1)}. \quad (6)$$

The lower and upper limits were ±1. A lower limit of −1 for the instance of $DSC^{sta}_{lowUptake} = DSC^{sta}_{lowUptake}$ (uptakeRatio = 4:4:1) - and an absence of PET segmentation within the $CT_{lowUptake}$ sub-volume i.e., $DSC^{4D,mot,deb}_{lowUptake} = 0$. The upper limit of +1 for the instance of absence lowUptake segmentation in PET_{sta} images, i.e. $DSC^{sta}_{lowUptake} = 0$, and a PET segmentation matching a highest lowUptake segmentation i.e. $DSC^{4D,mot,deb}_{lowUptake} = DSC^{sta}_{lowUptake}$ (uptakeRatio = 4:4:1).

By definition, ΔVol = 0 and the $lowUptake_{error} = 0$ for every instance of segmentation in the PET_{sta} scan protocol. The ΔVol and $lowUptake_{error}$ were compared to corresponding segmentations in the static PET_{sta} images for identically target volumes and uptakeRatios with a 5% significance level for a one-sample t-test.

3. Results

The $CT_{highUptake}$ and $CT_{lowUptake}$ volumes (mean ± 1 sd) deviations from the actual cavity volumes were 2% ± 4% for targets A and B, while the target C deviation was −5% ± 5%. The background measured SUV PET tracer activity (mean ± 1 sd) was 1.00 ± 0.06, including all FDG-PET tracer uptakeRatios. Summary tables of the ΔVol and $lowUptake_{error}$ metric are listed in tables 2 and 3, respectively. Apart from the Max40% segmentation method the PET_{4D} scans produced enlarged segmented volumes compared to the PET_{sta} scans with an increased ΔVol for the SUV40% of 2% ($p = 0.001$), gradientSeg of 4% ($p = 0.03$) and the 2.5SUV of 4% ($p < 0.001$), including all six PET bins. The $lowUptake_{error}$ of the PET_{4D} was not significantly different from $lowUptake_{error}$ of the PET_{sta} for any segmentation method. For the PET_{mot} scan protocol, the ΔVol significantly increased for the Max40% of 15% ($p < 0.001$) and GradientSeg of 43% ($p < 0.001$) methods compared

to the PET_{sta} images. No significant changes in the ΔVol were observed for the SUV40% and 2.5SUV methods in the PET_{mot} scan protocol. A significant increase in the $lowUptake_{error}$ quantity of 0.33 was detected for the PET_{mot} scan protocol using the GradientSeg method ($p < 0.001$). None of the three methods Max40%, SUV40% or 2.5SUV provided $lowUptake_{error}$ in the PET_{mot} protocol significantly different from the $lowUptake_{error}$ in the PET_{sta} protocol. In the motion corrected PET_{deb} scan protocol the Max40% ($p < 0.001$) and SUV40% ($p = 0.03$) methods both presented increased segmented volumes of 10% and 6% respectively, compared to the PET_{sta} scan protocol. The 2.5SUV method presented a lower segmented volume of -4% in PET_{deb}, in which marginally significant difference was found compared to the PET_{sta} scan protocol ($p = 0.048$). In the PET_{deb} scan the GradientSeg segmentations was not significantly different from the PET_{sta} scan protocol according to both the ΔVol and $lowUptake_{error}$ quantity. As the only segmentation method the Max40% presented a $lowUptake_{error}$ metric that was significantly different in the PET_{deb} scan protocol compared to the PET_{sta} ($p = 0.04$).

4. Discussions

This study aimed for a method to evaluate PET segmentation methods accuracies relative to static PET segmentations, for target lesions combined with motion and tracer uptake heterogeneity. The study design allocated the highUptake volume within every instance of segmentation. When motion was applied, the activity in the highUptake volume (as well as the lowUptake sub-volume) was partially displaced in the direction of the lowUptake location, simulating an overflow effect into the lowUptake location. For the instance of the cone shaped volume (target C), the overflow effect would be from the cone wide-end highUptake sub-volume into the neighbouring lowUptake cone. This study design using a phantom with plastic inserts with zero tracer uptake and motion restricted to a regular one dimension (superior-inferior) are limitations that may not be directly equivalent to the clinical conditions. Additionally, using the static segmentation as a reference condition for each segmentation method and uptakeRatio, revealed inconsistency using the GradientSeg method for target containing SUV of 3. For the instance with uptakeRatio of 3:2:1 the highUptake volume were segmented using GradientSeg method, whereas with uptakeRatio of 8:3:1 the lowUptake sub-volume was absent in the segmentation. The non-appearance of lowUptake SUV = 3 for this instance could be explained by the highUptake SUV of 8, which add high intensity gradients and thus underestimate the nearby lowUptake sub-volume. The acquisition time was retained equal between 3D and the individual 4D PET bins. This would lead to a scan time, in 4D-PET, not suitable for all patient intended for lung cancer radiotherapy. In contrast to clinical conditions, the injected activity for this study with a background of roughly 10 kBq ml^{-1}, lead to improved count statistics and hence reduced noise. The influence of a reduction of injected activity and hence reduced count statistics, in combination with motion and inhomogeneous target activity, have not been addressed in this study. The applied tracer concentrations and the uniform background activity, are likely to improve segmentation on images effected by motion. Thus further investigation of both reduced activity and increased target activity variations are needed to associate the phantom study to clinically condition. The effect of cold wall in combination with threshold PET segmentation has previously been addressed by [21] as potentially leading to overestimation of wall-less targets. The effect of irregular respiration patterns on different segmentation methods have been addressed by [22]. Despite Carles *et al* reported no substantial segmenting differences; irregular respiration during PET imaging might eventually lead to degraded PET images between the individually 4D-PET bins due to reduced count statistics.

Excluding the Max40% method, the segmented volumes in PET_{4D} were slightly increased compared to the PET_{sta} protocol. The residual motion in the PET bins connecting the inhale and exhale respiratory phases could explain the increasing volumes in PET_{4D}. However, the narrow confidence interval of the ΔVol in PET_{4D} points to a scan protocol that was consistent with segmentations in PET_{sta}. The $lowUptake_{error}$ quantities in PET_{4D} supported segmentations consistent with the PET_{sta} scan protocol as no significant deviation was detected. This finding was consistent to the result obtained by Bettinardi *et al* [11]; thus for these investigated segmentation methods, a six bin 4D-PET is recommended to be sufficient for motion-free image acquisition with an amplitude of 15 mm.

For the motion involved PET_{mot} scan protocol the segmented volumes were enlarged, except for the 2.5SUV segmentation method. In both PET_{mot} and PET_{deb} the 2.5SUV method provided segmentations that were equivalent to the PET_{sta} acquisition. In contrast to the other two threshold based segmentation methods (Max40% and SUV40%), the 2.5SUV method did not involve the maximum intensity, which degrades in motion-disturbed image acquisitions [18]. Eroded SUV values may not necessarily correspond to different volume of segmentations, as long as the SUV remains above the threshold value. Changed constellations of the FDG activities in the high- and lowUptake volumes are likely to affect the results in PET_{mot} images, particularly for threshold values close to the target SUV. The Max40% method exclusively relies on the maximum SUV, whereas the SUV40% method also facilitated background SUV correction. Due to the degraded SUV of PET_{mot}

images, a lower absolute threshold SUV value was used for Max40% segmentation compared to the SUV40%, thus segmenting slightly enlarged volumes relative to the SUV40%.

In specific the PET_{mot} protocol, the GradientSeg method presented an enlarged segmented volume and an overestimation of the lowUptake sub-volume. Since the GradientSeg method makes use of intensity gradients as segmentation boundaries the enlarged volume, including the lowUptake sub-volume, was expected as a consequence of motion blur. This finding was supported by the PET_{deb} scan protocol, which revealed that the segmentations were equivalent to PET_{sta} scan protocol due to the motion correction with a re-establishment of the image intensity gradients.

Motion corrections through PET image deblurring with a position density estimate rely on accurate 4D-CT target detection. Due to the centre-off-mass uncertainties within 4D-CT phases, the position density estimate may potentially result in an inaccurate processing of the deblurred PET image. An increased number of bins for 4D-CT may improve the temporal centre-of-mass detection and high contrast fiducials could clarify detection within each phases and limit uncertainties within K_{motion} determination. Detecting fiducial centre-of-masses within a 4D-CT scan is still limited by other factors such as the velocity during CT phase bin acquisition, orientations and fiducial style [23]. Other methods have been suggested to identify and reduce motion blur in PET images without 4D-PET, using either deformation vector field for correction or deblurring methods [24, 25]. However, segmentation in these motion corrected images were performed using a fixed threshold of the maximum intensity, close related to the Max40% method. Furthermore, local deblurring must be performed near distinct lesions due to the various respiratory motion trajectories inside the lungs [26].

In radiotherapy, including tissue with low PET activity to the planning target volume (PTV) due to non-corrected images (PET_{mot}) and a high accurate segmentation method (GradientSeg) in stationary images (PET_{sta}), could potentially lead to enlarged PTV and consequently increased normal tissue complication [27]. On the other hand, PET images which encompass the all possible target positions throughout a full respiration cycle for internal target volume (ITV) definition could be used for segmentation. In contrast to the motion blur of PET_{mot} in the current study, which encompassed all target locations, motion incorporation using super resolution-corrected PET images showed improved accuracy according to the 'true' ITV [28].

The main restrictions of the current study are the limited number of different tracer heterogeneity values and target geometries. In particular, heterogeneous PET lesions have been reported to be challenging to segment for threshold based methods and thereby suggesting sophisticated methods instead [29]. This current study exposed the similarity of the segmentation methods relative to a motionless PET acquisition. With segmentation evaluation relative to static PET segmentations, reliable ground-truth tumour volume description in static PET images is fundamental. A study by Wanet *et al* revealed high accuracy for the GradientSeg method in 4D-PET compared to the surgical specimens of lung cancers [20]. A closely related study by Yu *et al* published an optimal SUV threshold close to 3.0 with a threshold of 2.5 SUV that was not significantly different from the surgical specimens [8]. These related studies assessing the accuracies of PET segmentations according to pathological findings indicate a general recommendation for the use of GradientSeg or 2.5 SUV segmentation methods. The findings in the present study identifying the discrepancies in PET segmentation, with and without motion-correction, combined with heterogeneous tracer uptake indicates that caution should be used with maximum intensities for threshold segmentation. In addition, severe overestimation potentially exists for the GradientSeg method in non-corrected motion influenced PET images. For the instance of motion affected target lesions and the absence of 4D-PET or 3D-PET motion correction, the 2.5SUV segmentation method is recommended.

5. Conclusions

This experimental study of target lesions with heterogeneous tracer (FDG) uptake and respiratory motion demonstrated that motion-corrected PET imaging (PET_{4D} or PET_{deb}) in combination with GradientSeg method provided the most consistent and accurate PET segmentation amongst the tested segmentation methods. The GradientSeg method should be excluded in the absence of motion correction.

Acknowledgments

The work is supported by DAKFO—Danish Graduate School in Clinical Oncology and CIRRO—The Lundbeck Foundation Center for Interventional Research in Radiation Oncology and The Danish Council for Strategic Research.

Conflicts of interest

None to declare. The authors alone are responsible for the content and writing of the paper.

References

[1] Engholm G *et al* 2015 NORDCAN: Cancer Incidence, Martality, Prevalence and Survival in the Nordic Contries, Version 7.1. Association of the Nordic Cancer Registries. Danish Cancer Society (http://ancr.nu)

[2] Goldstraw P 2009 The 7th edition of TNM in lung cancer: what now? *J. Thorac. Oncol.* **4** 671–3
[3] Goldstraw P (ed) 2009 *Staging Manual in Thoracic Oncology* (Orange Park, FL: Editorial Rx Press)
[4] Heusch P *et al* 2014 Thoracic staging in lung cancer: prospective comparison of 18F-FDG PET/MR imaging and 18F-FDG PET/CT *J. Nucl. Med.* **55** 373–8
[5] Konert T *et al* 2015 PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014 *Radiother. Oncol.* **116** 27–34
[6] Edet-Sanson A, Dubray B, Doyeux K, Back A, Hapdey S, Modzelewski R, Bohn P, Gardin I and Vera P 2012 Serial assessment of FDG-PET FDG uptake and functional volume during radiotherapy (RT) in patients with non-small cell lung cancer (NSCLC) *Radiother. Oncol.* **102** 251–7
[7] Werner-Wasik M *et al* 2012 What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom *Int. J. Radiat. Oncol. Biol. Phys.* **82** 1164–71
[8] Yu J *et al* 2009 Comparison of tumor volumes as determined by pathologic examination and FDG-PET/CT images of non-small-cell lung cancer: a pilot study *Int. J. Radiat. Oncol. Biol. Phys.* **75** 1468–74
[9] Steenbakkers R J *et al* 2006 Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis *Int. J. Radiat. Oncol. Biol. Phys.* **64** 435–48
[10] Calais J, Thureau S, Dubray B, Modzelewski R, Thiberville L, Gardin I and Vera P 2015 Areas of high 18F-FDG uptake on preradiotherapy PET/CT identify preferential sites of local relapse after chemoradiotherapy for non-small cell lung cancer *J. Nucl. Med.* **56** 196–203
[11] Bettinardi V, Rapisarda E and Gilardi M C 2009 Number of partitions (gates) needed to obtain motion-free images in a respiratory gated 4D-PET/CT study as a function of the lesion size and motion displacement *Med. Phys.* **36** 5547–58
[12] Foster B, Bagci U, Mansoor A, Xu Z and Mollura D J 2014 A review on segmentation of positron emission tomography images *Comput. Biol. Med.* **50** 76–96
[13] Greco C, Rosenzweig K, Cascini G L and Tamburrini O 2007 Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC) *Lung Cancer* **57** 125–34
[14] Aristophanous M, Penney B C, Martel M K and Pelizzari C A 2007 A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography *Med. Phys.* **34** 4223–35
[15] Geets X, Lee J A, Bol A, Lonneux M and Gregoire V 2007 A gradient-based method for segmenting FDG-PET images: methodology and validation *Eur. J. Nucl. Med. Mol. Imaging* **34** 1427–38
[16] Biehl K J, Kong F M, Dehdashti F, Jin J Y, Mutic S, El N I, Siegel B A and Bradley J D 2006 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J. Nucl. Med.* **47** 1808–12
[17] Le M A, Hatt M, Pradier O, Cheze-Le R C and Visvikis D 2012 Impact of the accuracy of automatic tumour functional volume delineation on radiotherapy treatment planning *Phys. Med. Biol.* **57** 5381–97
[18] Park S J, Ionascu D, Killoran J, Mamede M, Gerbaudo V H, Chin L and Berbeco R 2008 Evaluation of the combined effects of target size, respiratory motion and background activity on 3D and 4D PET/CT images *Phys. Med. Biol.* **53** 3661–79
[19] El N I, Low D A, Bradley J D, Vicic M and Deasy J O 2006 Deblurring of breathing motion artifacts in thoracic PET images by deconvolution methods *Med. Phys.* **33** 3587–600
[20] Wanet M, Lee J A, Weynand B, De B M, Poncelet A, Lacroix V, Coche E, Gregoire V and Geets X 2011 Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens *Radiother. Oncol.* **98** 117–25
[21] Hofheinz F, Dittrich S, Potzsch C and Hoff J 2010 Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold *Phys. Med. Biol.* **55** 1099–113
[22] Carles M, Fechter T, Nemer U, Nanko N, Mix M, Nestle U and Schaefer A 2015 Feasibility of a semi-automated contrast-oriented algorithm for tumor segmentation in retrospectively gated PET images: phantom and clinical validation *Phys. Med. Biol.* **60** 9227–51
[23] Nielsen M S, Nystrom M W and Carl J 2013 Potential position errors using fiducial markers for gated image guided radiotherapy *Acta Oncol.* **52** 1472–6
[24] Reyes M, Malandain G, Koulibaly P M, Gonzalez-Ballester M A and Darcourt J 2007 Model-based respiratory motion compensation for emission tomography image reconstruction *Phys. Med. Biol.* **52** 3579–600
[25] Xu Q, Yuan K and Ye D 2011 Respiratory motion blur identification and reduction in ungated thoracic PET imaging *Phys. Med. Biol.* **56** 4481–98
[26] Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van H M, Lebesque J V and Miyasaka K 2002 Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 822–34
[27] Marks L B *et al* 2010 Radiation dose-volume effects in the lung *Int. J. Radiat. Oncol. Biol. Phys.* **76** (Suppl. 3) S70–6
[28] Hatt M, Maitre A L, Wallach D, Fayad H and Visvikis D 2012 Comparison of different methods of incorporating respiratory motion for lung cancer tumor volume delineation on PET images: a simulation study *Phys. Med. Biol.* **57** 7409–30
[29] Hatt M, Cheze-Le R C, van B A, Lambin P, Pradier O and Visvikis D 2011 Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation *J. Nucl. Med.* **52** 1690–7