Am J Nucl Med Mol Imaging 2018;8(2):127-136
www.ajnmmi.us /ISSN:2160-8407/ajnmmi0072143

Original Article

Repeatability of FDG PET/CT metrics assessed in free breathing and deep inspiration breath hold in lung cancer patients

Lotte Nygård[1], Marianne C Aznar[1], Barbara M Fischer[2], Gitte F Persson[1], Charlotte B Christensen[2], Flemming L Andersen[2], Mirjana Josipovic[1,5], Seppo W Langer[1], Andreas Kjær[2], Ivan R Vogelius[1], Søren M Bentzen[3,4]

[1]Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, 2100, Copenhagen, Denmark; [2]Department of Clinical Physiology, Nuclear Medicine & PET and Cluster for Molecular Imaging, Rigshospitalet and University of Copenhagen, Blegdamsvej 9, 2100, Copenhagen, Denmark; [3]Division of Biostatistics and Bioinformatics, University of Maryland Greenebaum Comprehensive Cancer Center, MD 21201, Baltimore, USA; [4]Department of Epidemiology and Public Health, School of Medicine, University of Maryland, 655 W Baltimore S, MD 21201, Baltimore, USA; [5]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100, Copenhagen, Denmark

Received January 5, 2018; Accepted February 13, 2018; Epub April 25, 2018; Published April 30, 2018

Abstract: We measured the repeatability of FDG PET/CT uptake metrics when acquiring scans in free breathing (FB) conditions compared with deep inspiration breath hold (DIBH) for locally advanced lung cancer. Twenty patients were enrolled in this prospective study. Two FDG PET/CT scans per patient were conducted few days apart and in two breathing conditions (FB and DIBH). This resulted in four scans per patient. Up to four FDG PET avid lesions per patient were contoured. The following FDG metrics were measured in all lesions and in all four scans: Standardized uptake value $(SUV)_{peak}$, SUV_{max}, SUV_{mean}, metabolic tumor volume (MTV) and total lesion glycolysis (TLG), based on an isocontur of 50% of SUV_{max}. FDG PET avid volumes were delineated by a nuclear medicine physician. The gross tumor volumes (GTV) were contoured on the corresponding CT scans. Nineteen patients were available for analysis. Test-retest standard deviations of FDG uptake metrics in FB and DIBH were: SUV_{peak} FB/DIBH: 16.2%/16.5%; SUV_{max}: 18.2%/22.1%; SUV_{mean}: 18.3%/22.1%; TLG: 32.4%/40.5%. DIBH compared to FB resulted in higher values with mean differences in SUV_{max} of 12.6%, SUV_{peak} 4.4% and SUV_{mean} 11.9%. MTV, TLG and GTV were all significantly smaller on day 1 in DIBH compared to FB. However, the differences between metrics under FB and DIBH were in all cases smaller than 1 SD of the day to day repeatability. FDG acquisition in DIBH does not have a clinically relevant impact on the uptake metrics and does not improve the test-retest repeatability of FDG uptake metrics in lung cancer patients.

Keywords: FDG PET, repeatability, FB versus DIBH, lung cancer

Introduction

^{18}F-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET)/computed tomography (CT) is recommended for staging and radiotherapy planning in The National Comprehensive Cancer Network (NCCN) guidelines for inoperable locally advanced non-small cell lung cancer (NSCLC). Few studies have evaluated the repeatability of PET assessed FDG uptake in lung cancer patients but the literature suggest that up to 20% variation can be expected. However, none of these studies accounted for the potential effect of breathing [1-4].

We hypothesized that the motion of the lung during the relatively prolonged PET image acquisition process could affect the repeatability and the estimate of FDG uptake metrics in the tumor lesions, possibly to a level where it could impact treatment decisions or response assessment [5, 6]. Also, clinical trials of radiation therapy with dose-escalation to the PET positive areas might be affected by breathing [7]. For this reason, we conducted a prospective test-retest study of FDG uptake in locally advanced non-small cell lung cancer patients scanned in free breathing (FB) and deep inspiration breath hold (DIBH) with FDG PET/CT for radiotherapy planning.

Specifically, the study tested the hypothesis that FDG uptake values, SUV_{max}, SUV_{mean}, SUV_{peak}, metabolic tumor volume (MTV), total lesion glycolysis (TLG) and clinical volumes were more repeatable when the FDG PET/CT scan was performed in DIBH than when it was performed without respiratory management. It was expected that SUV_{max} was higher in DIBH compared to FB and that the FDG avid tumor volume was smaller in DIBH than in FB. We also expected the FDG PET metrics to change most from FB to DIBH for smaller tumors with larger respiratory motion.

Methods and materials

Twenty patients diagnosed with inoperable non-small cell lung cancer and eligible for curative intent radiotherapy were included. Inclusion criteria were histology proven non-small cell lung cancer, age above 18 and BMI≤30. Patients with diabetes type I or patients not able to comply with 20 sec DIBH prior to the first FDG PET/CT scan, were excluded from the study.

Planned treatment consisted of radiotherapy in 2 Gray (Gy) fractions given 33 times (5 fractions per week) to a total of 66 Gy. If eligible, patients received chemotherapy in a concurrent schedule, consisting of 3 series of either cisplatin or carboplatin in combination with vinorelbine.

FDG PET/CT scans were acquired in conventional FB conditions and DIBH. Both types of scans were repeated a few days apart without any active therapy given in the interval with the purpose of assessing test-retest variability.

All patients signed an informed consent before inclusion. The study was approved by the Danish Ethics Committee, protocol number H-1-2014-011 and the Danish Data Protection Agency, 02986/30-1271.

Deep inspiration breath hold and image acquisition

The participant's inspiration level was monitored using the Real-time Position Management system (RPM) (Varian Medical Systems, Palo Alto, California, USA) on the PET/CT scanner used in the study. RPM uses an infra-red camera to track the patients' respiratory signal from an optical marker, placed on the patient's thorax surface.

To ensure a repeatable inspiration level during the image acquisition, participants received visual feedback of the RPM system's optical marker position through video goggles worn during the scanning. The patients were required to follow simple instructions and were given time to practice DIBH levels to ensure a comfortable and repeatable level of breath hold prior to the day 1 scan [8-10].

On both scanning days, patients were fasting for at least 4 hours prior to the FDG injection which was given as 4 MBq/kg. They rested reclined for 45 min ± 5 min before they were asked to empty their bladder and were subsequently positioned in the scanner. The patients were immobilised on a radiotherapy chest board (ConChest, Candor Aps, Gislev, Denmark), with arms above the head. Immobilisation and marking for radiotherapy was executed according to departmental guidelines. We aimed at performing the first DIBH FDG PET 60 min post FDG injection on both scanning days.

For the DIBH FDG PET scan, all image acquisitions were acquired over one field of view (FOV) encompassing 21.5 cm in the cranio-caudal direction while the participant was holding her/his breath at a predefined level. The location of the tumor and thus the position of the PET FOV for DIBH scanning was defined prior to the first scan of each patient from a diagnostic quality PET/CT scan obtained from the external diagnostic unit. Patients were asked to take a DIBH of 20 seconds for PET acquisition and this procedure was repeated 3 times (eight patients scanned before October 2015) and 6 times for twelve patients scanned from October 2016 and onwards in order to reduce the signal-to-noise ratio. Subsequently, these 3- or 6 PET acquisitions were combined by averaging raw counts using in-house software.

Participants were weighed before each injection of FDG on both scanning days. For patients with BMI≤25, FB scan time per bed position was 2 minutes. For patients with 25<BMI≤30, FB scan time per bed position was 2.5 minutes. If the patient's arms were down the side (due to physical constraints), the FB scan time was 3 min/PET bed.

Table 1. Patient characteristics. TNM (Tumor, Nodal, Metastases) and Clinical Stage according to The Union for International Cancer Control (UICC), version 7. Lesion characteristics (N=38) in mean and range

Patient characteristics N=19		
Gender	Female: 6; Male: 13	
Age	Median 70 years, mean 68.5	
	Range [54-85]	
T-stage	TX: 1; T1: 4; T2: 3; T3: 5; T4: 6	
N-stage	N0: 1; N1: 3; N2: 7; N3: 8	
M-stage	M0: 17; M1b: 2	
Clinical Stage	II: 1; IIIA: 8; IIIB: 8; IV: 2	
Histology	Adenocarcinoma: 9	
	Squamous cell carcinoma: 9	
	Adenosquamous carcinoma: 1	
Lesion characteristics T-sites N=22 N-sites N=16	FB	DIBH
SUV_{peak}	11.0 [1.7-29.6]	11.4 [2.0-29.4]
SUV_{max}	16.9 [3.9-44.8]	18.3 [4.4-41.7]
SUV_{mean}	10.8 [2.7-25.7]	11.8 [2.8-24.8]
MTV	6.0 [0.1-36.8]	5.2 [0.1-48.1]
TLG	108.8 [0.9-803.2]	101.2 [0.8-1192.9]
GTV	96.5 [3.6-380.4]	95.1 [3.4-376.4]

SUV: Standardized uptake value. MTV: Metabolic tumor volume. TLG: Total lesion glycolysis. GTV: Gross tumor volume. FB: Free breathing. DIBH: Deep inspiration breath hold.

The FDG PET/CT scans on days 1 and 2 were performed on the same PET/CT scanner (Siemens Biograph mCT, Siemens Healthineers, Erlangen) and according to European Association of Nuclear Medicine (EANM) guidelines [11].

The scanner settings have been described previously [12]: Attenuation and scatter corrected PET data were reconstructed iteratively using a 3D ordered-subset expectation-maximization algorithm including point spread function and time of flight information (2 iterations, 21 subsets, 2 mm Gaussian filter, 400 × 400 matrix). The pixel size/slice thickness in the reconstructed PET image was approximately 2 × 2 mm/2 mm.

The CT scans were acquired without intra venous contrast and over the same field of view (FOV) as the PET scan, using 120 kV and either 40 mAs without CareDose setting (for low dose FB CT) or CareDose settings with 225 quality reference mAs (for diagnostic quality DIBH CT). The DIBH CT scan was acquired in a DIBH of approximately 10-15 seconds using the same visual feedback as for the PET part. The FB CT scan was acquired in approximately 3-4 seconds. All PET images were attenuation corrected using the CT acquired on the same day without intravenous contrast and in the same breathing condition (DIBH-PET with DIBH-CT). This resulted in four sets of images per patient: DIBH day 1, FB day 1, DIBH day 2 and FB day 2.

On day 2, all patients received CT scanning for radiotherapy planning, consisting of a respiratory correlated 4DCT and one DIBH CT (patients 1 and 2) or three consecutive CTs in three consecutive DIBHs (patients 3-20). For the radiotherapy planning process, the gross tumor volumes (GTVs) were delineated on the midventilation [13] phase of the 4DCT.

Lesion contouring

The robustness of the following FDG PET metrics was investigated: SUV_{max}, SUV_{peak}, SUV_{mean}, TLG and MTV. Analysis took place on a Mirada XD® workstation, version 1.1.0.31 (Mirada Medical, Oxford, UK). A maximum of two tumors (primary and e.g. satellite tumor) and a maximum of two lymph nodes were defined from the first FDG PET scan in the study. If the patient had several FDG avid lymph nodes, the two nodes were selected according to size, high FDG uptake and most accessible for discriminating between FDG avid- and non FDG avid tissue. An isocontur of 50% of SUV_{max} was applied to the region of interest (ROI).

Further, a nuclear medicine physician (CBC) delineated all PET positive lesions with a visually adapted and nudged isocontour. No fixed threshold was applied in the manual delineation process. The contouring was performed in a random order of days and breathing condition and blinded to previous delineations. MTV, SUV_{mean} and TLG were investigated on the nuclear medicine delineation as well as on the 50% isocontour. SUV_{peak} and SUV_{max} were identical in both delineation methods and when analysed from the 50% of SUV_{max} isocontour. All four FDG PET/CT scans were exported from Mirada into Eclipse (version 13.6, Varian

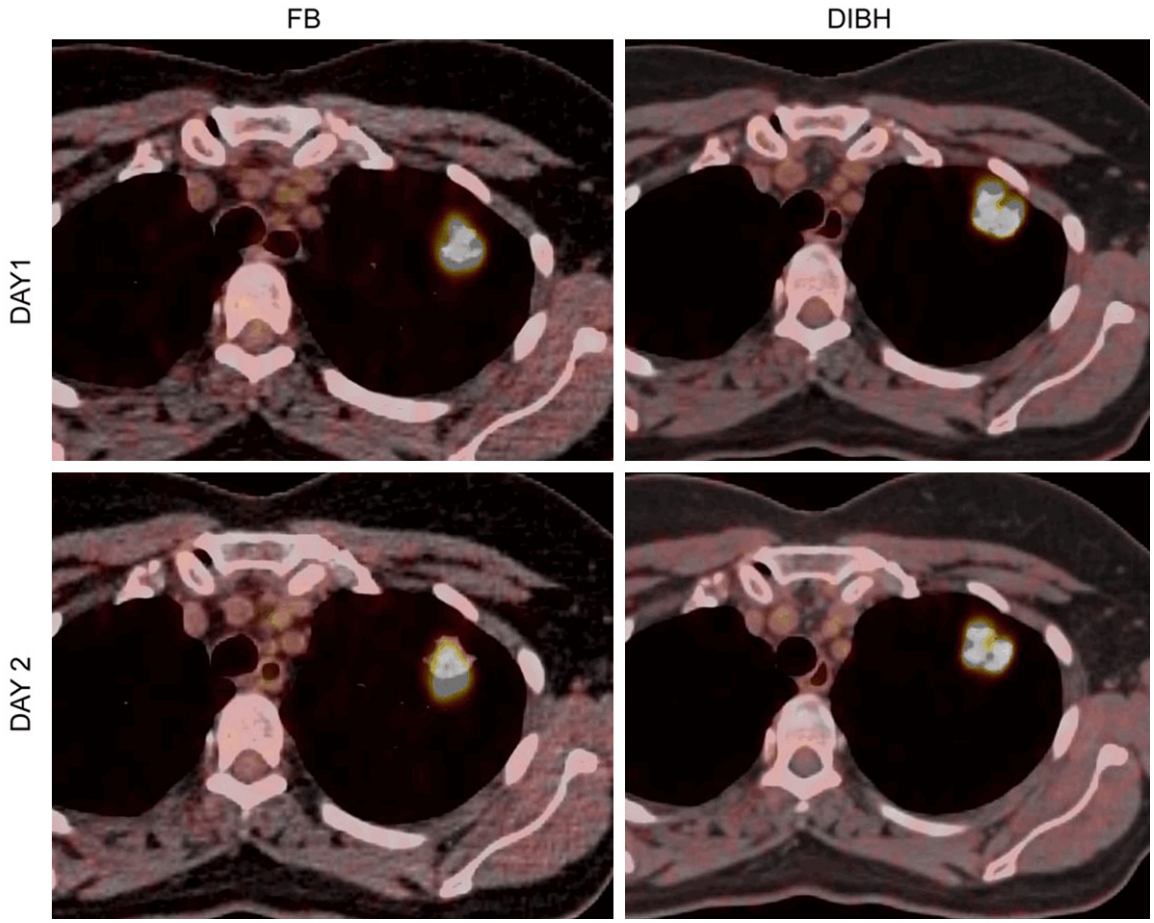


Figure 1. Illustration of a patient case. FDG PET/CT day 1 and 2 taken in both breathing conditions. FB: free breathing DIBH: deep inspiration breath hold. Notice "breathing artefact" with misalignment between CT and FDG uptake in FB on day 2 (lower left corner) compared to FB on day 1.

Medical Systems, USA). All CT scans from the PET/CT sets (i.e. both FB and DIBH from both scan days) were rigidly registered to the mid-ventilation phase of the 4DCT; in order to help with delineation on all the scans, automatic registration was applied, with focus on the volume around the delineated GTVs. After the automatic registration, a deformable registration was applied (Smart Adapt option in Eclipse) on the GTVs. The gross tumor volumes (GTVs) were propagated from the mid-ventilation phase to all four scans. Finally the GTV volumes on the PET/CTs were manually adjusted by LN and GP (senior radiation oncologist, expert in lung cancer). Tumors in the lung were contoured using "lung window" (Hounsfield units (HU) range -1000 HU to 0 HU) and lymph nodes/mediastinal tumors were contoured using a "soft tissue window" (HU range -125 HU to 225 HU).

Statistics

The study was a double-paired design with two repeated measurements of FDG uptake in DIBH, two repeated measurements of FDG uptake in FB, and two consecutive comparisons of FB versus DIBH FDG uptake.

Scatterplots and Bland-Altman plots were used to assess the error structure in the data. Logistic regression of day 1 versus day 2 in both breathing conditions was calculated and R^2 extracted as a measure of explained variance. The mean differences - measured in percent difference after visual assessment of data distributions in Bland-Altman plots-between the two days were extracted and the standard deviation (SD) calculated from day to day in the two breathing conditions:

$$\text{Difference}_{\text{FBDay1+2}} = \left(\frac{\text{FB1} - \text{FB2}}{0.5 * (\text{FB1} + \text{FB2})}\right) * 100\ [\%],$$

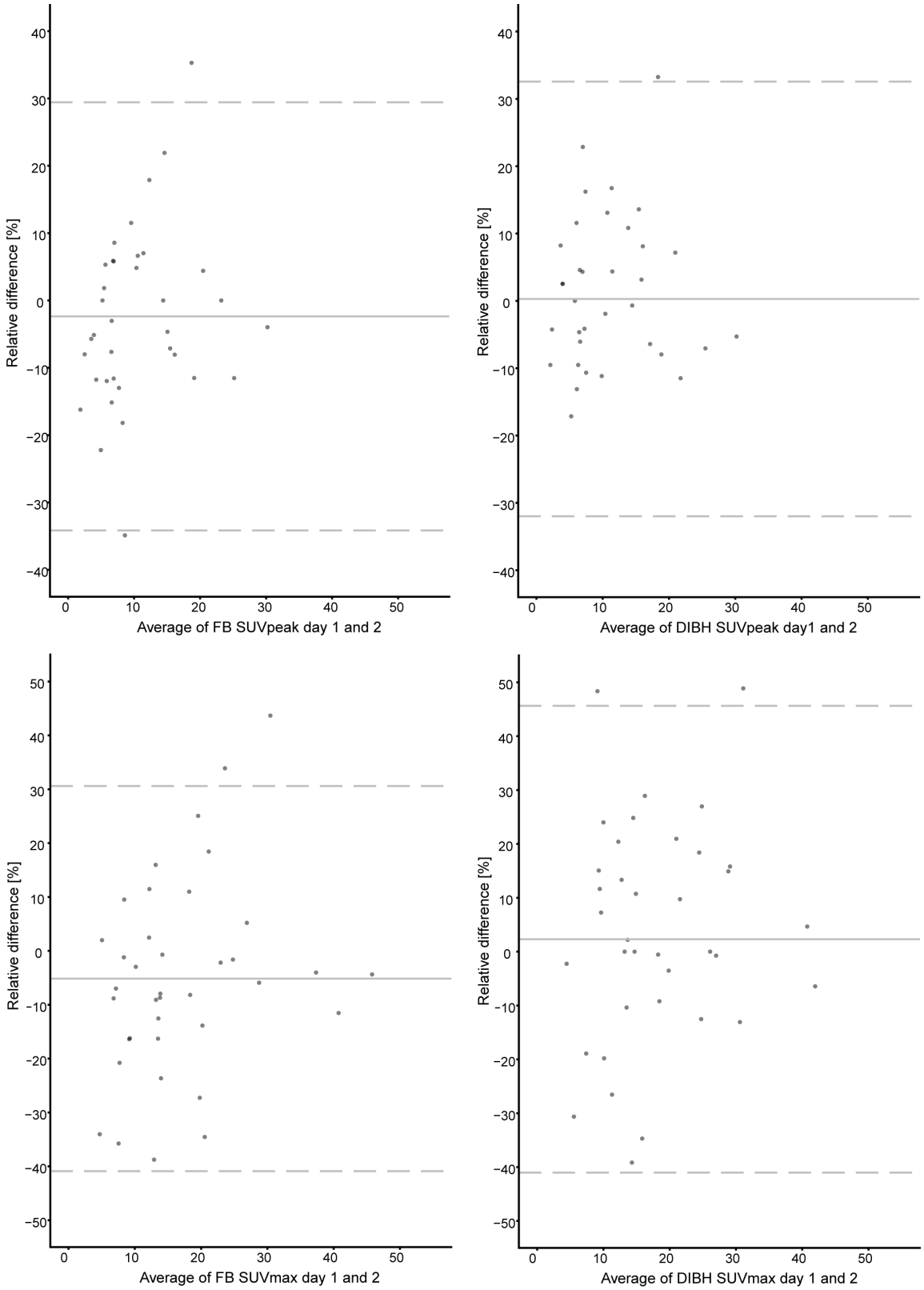


Figure 2. Bland Altman plots of the day to day repeatability of SUV_{max} and SUV_{peak} in the two breathing conditions. SUV: Standardized uptake value. FB: Free breathing. DIBH: Deep inspiration breath hold. Full line: Mean difference [%] from day to day. Dashed lines: Limits of agreement (± 1.96 × SD). SD: Standard deviation from **Table 2**.

Table 2. Logistic regression analysis from day to day (R^2) and Relative difference from day to day (SD-descriptives) in the two breathing conditions. SD in percent (%) if not otherwise specified

	50% isocontur of SUV_{max}						Expert delineation				
							Nuclear medicine				Radiation oncologist
	SUV_{peak}	SUV_{max}	SUV_{mean}	MTV	MTV	TLG	SUV_{mean}	MTV	MTV	TLG	GTV
FB_{Day1+2}											
R^2	0.88	0.88	0.85	0.95	0.95	0.96	0.63	0.99	0.99	0.98	0.99
SD [%]	16.2	18.2	18.3	35.5	2.5 [cm^3]	32.4	21.2	43.8	11.0 [cm^3]	35.6	17.6
$DIBH_{Day1+2}$											
R^2	0.88	0.80	0.77	0.94	0.94	0.92	0.66	0.96	0.96	0.96	0.99
SD [%]	16.5	22.1	22.1	49.5	3.2	40.5	16.3	27.0	14.1	26.5	18.8

FB: Free breathing. DIBH: Deep inspiration breath hold. SD: Standard deviation. SUV: Standardized uptake value. MTV: Metabolic tumor volume. TLG: Total lesion glycolysis. CT: Computer tomography. GTV: Gross tumor volume. 38 lesions delineated from FDG PET.

Table 3. Paired comparison of FB versus DIBH. Relative difference between the two breathing conditions [%]

	Day 1 (FB versus DIBH)			Day 2 (FB versus DIBH)		
50% isocontur of SUV_{max}	Mean difference [%]	SEM [%]	*p*-value	Mean difference [%]	SEM [%]	*p*-value
SUV_{peak}	-4.4	1.2	0.001	-1.8	1.4	0.2
SUV_{max}	-12.6	2.7	<0.0001	-5.2	2.5	0.04
SUV_{mean}	-11.9	2.5	<0.0001	-4.5	2.4	0.06
MTV	27.6	5.7	<0.0001	9.5	6.4	0.1
TLG	17.6	4.3	0.0002	-0.6	4.8	0.9
Nuclear medicine delineation						
SUV_{mean}	-4.1	2.6	0.11	-9.0	2.2	0.0001
MTV	-0.2	6.2	0.97	19.2	6.5	0.004
TLG	-4.4	4.5	0.32	10.6	5.2	0.041
Radiation oncologist CT delineation						
GTV	10.7	3.3	0.002	1.8	2.6	0.5

FB: Free breathing. DIBH: Deep inspiration breath hold. SEM: Standard error of the mean (=95% CI from the T-test divided by (1.96 × 2=3.92)). SUV: Standardized uptake value. MTV: Metabolic tumor volume. TLG: Total lesion glycolysis. CT: Computer tomography. GTV: Gross tumor volume.

and the equivalent equation for DIBH. The relative differences between uptake metrics between the two breathing modalities on the same scan were calculated using T-test statistics and the average difference in percent was extracted as measure of effect size.

SPSS Statistics for Windows version 22.0 (Armonk, NY: IBM Corp) and R [14] were used for all analyses and a two-sided *p*-value below 0.05 was considered statistically significant.

Results

Between December 2014 and March 2016, twenty patients were included in the study and all completed all scans on both scanning days and in both breathing conditions. In one patient the PET structures could not be divided meaningfully into separate lesions due to disease extent. This patient was excluded from the analysis and was referred for palliative chemotherapy instead of the originally planned treatment. Patient and lesion characteristics are shown in **Table 1**.

One patient was classified as having TX [15] disease due to a former lung cancer in 2013 which was surgically removed. The patient presented in the current study (2015) with relapse of squamous cell carcinoma located adjacent to the trachea, a tumor mass in the tissue surrounding esophagus and a FDG avid lymph node in station 1R.

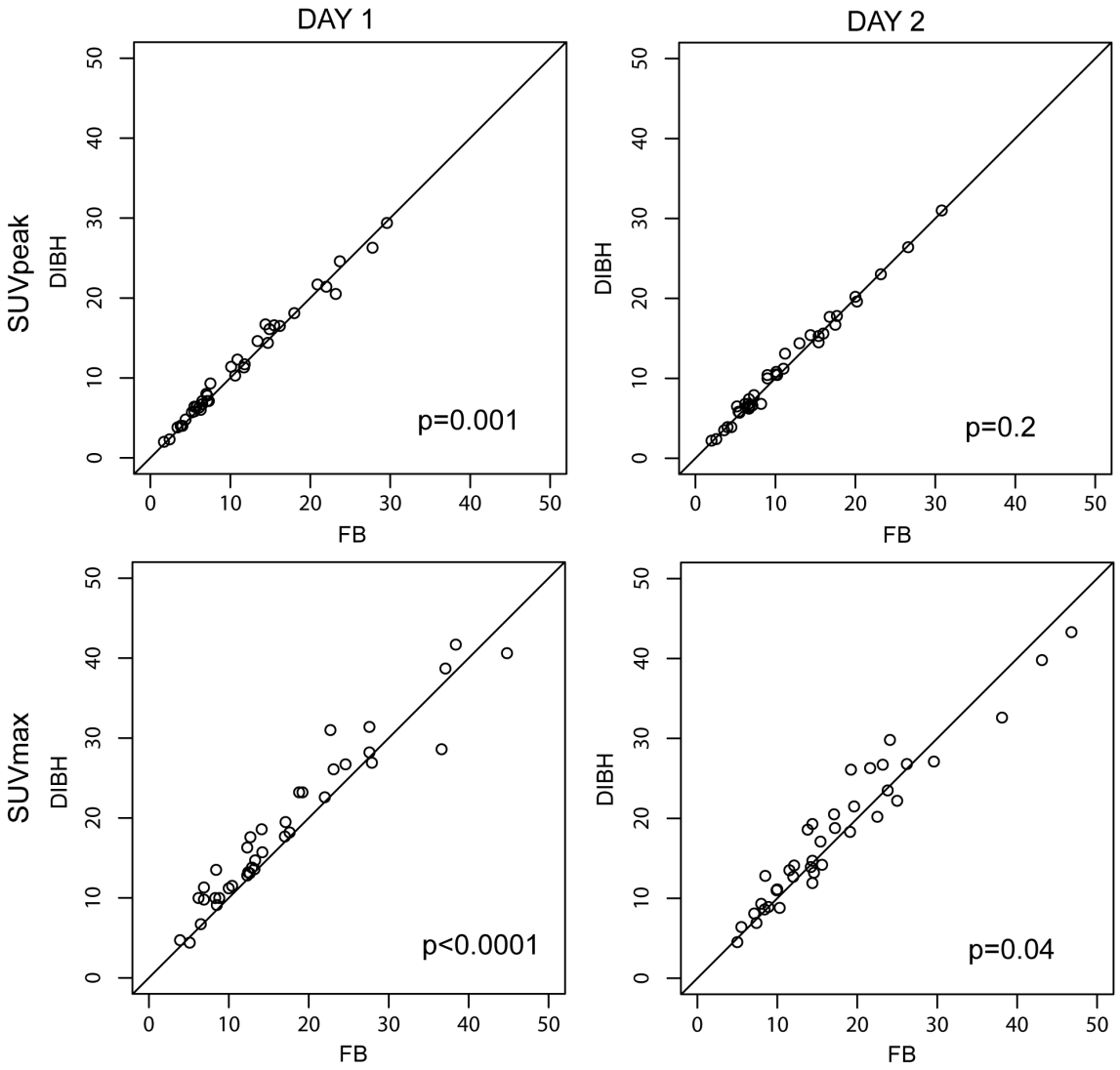


Figure 3. Scatter plots of DIBH versus FB in SUV_{max} and SUV_{peak} on both scanning days (Day 1 and Day 2). SUV: Standardized uptake value. FB: Free breathing. DIBH: Deep inspiration breath hold. P: *P*-value.

Figure 1 shows a patient's primary tumor scanned in FB and in DIBH. The signal-to-noise ratio seems better in the DIBH scans compared to FB and the geographical location is arguably better.

Average time between FDG injection and first DIBH on day 1 was 1 h 01 m, range [1 h 0 m; 1 h: 05 m]. On day 2, average time from injection to first DIBH was 1 h 2 m, range [1 h 0 m; 1 h 13 m]. In FB, average time on day 1 from injection to FB scan was 1 h 7 m, range [1 h 3 m; 1 h: 15 m]. On day 2, average time from injection to FB scan was 1 h 8 m, range [1 h 4 m; 1 h 18 m].

Figure 2 and **Table 2** show that the repeatability (standard deviation (SD)) was slightly larger in DIBH than in FB for the 50% of SUV_{max} isocontur and GTV delineations, whereas SD was lower in DIBH for the nuclear medicine delineations. R^2 from the logistic regression was larger in FB across delineation methods except for SUV_{mean} in the nuclear medicine delineation.

The two breathing conditions were compared on day 1 and day 2 to test the hypothesis of higher SUV values and lower volumes in DIBH compared to FB. **Table 3** shows that the hypothesis is supported on day 1 for the delineations with a threshold of 50% of SUV_{max} and in the GTV contours but not for the nuclear medicine delineation. The two breathing conditions are significantly different on the day 2

image sets but not on day 1 for the nuclear medicine delineation. Only SUV_{max} remained significantly different between the two breathing conditions. **Figure 3** illustrates the differences in SUV_{peak} and SUV_{max} on day 1 and day 2 with SUV_{peak} being less sensitive to respiration than SUV_{max}.

The majority of patients included in the study had primary tumors located in the upper lobe. The relative difference in SUV values between FB and DIBH as a function of amplitude of tumor motion in the cranio caudal (CC) direction (available from the radiotherapy planning 4D CT scan) is shown in Supplementary Figure 1. The difference of SUV values in FB and DIBH did not depend on the amplitude of primary tumor motion.

Discussion

FDG PET repeatability in FB and DIBH was quantified in 19 patients with inoperable NSCLC eligible for curatively intended chemoradiotherapy. Day-to-day standard deviations (SDs) were relatively large in both breathing conditions and, contrary to expectations, SD were marginally *larger* in DIBH than in FB. A published study found relatively narrower SDs in stage IV lung cancer patients with lesions within and outside the thorax [2] but breathing was not accounted for. One could expect that in parenchymal lung cancer lesions, the SD would indeed be larger.

On day 1, SUV_{peak}, SUV_{max}, SUV_{mean}, MTV and TLG were found to be significantly different in DIBH compared to FB. Surprisingly, the second day's scans seemed to consistently demonstrate a smaller impact of breathing condition. One explanation could be that patients were more relaxed on the day 2 and did not perform as well in the DIBH scan as on the first day. This is despite that the DIBH level in the RPM system was the same. Arching of the back to obtain the same DIBH level from day to day has been reported in breast cancer patients [16] and could potentially be the reason for lower mean differences in day two compared to day 1. Only SUV_{max} remained significantly different in the two breathing conditions.

The potential benefit of conducting PET/CT scans in DIBH may depend on the amplitude of tumor motion in the individual patient. To explore such dependence, we used the clinically assessed motion amplitudes from the 4D CT scan. These amplitudes are used to calculate patient-specific margins for radiotherapy. In our cohort, 8 tumors exhibited a displacement of >0.5 cm. However, the motion amplitude does not appear to be a dominant factor in the uncertainty of FDG uptake metrics. Additionally, the lesions with large motion amplitude in FB do not appear to have a relevant gain in reproducibility with DIBH scanning beyond the average reported here, see Supplementary Table 1.

Even if the value and repeatability of the conventional FDG uptake metrics are not improved at a clinically relevant scale when scanning in DIBH, the method could be of relevance if high geometrical precision is required [17-19] as shown with the example of **Figure 1**.

Another application where improved spatial accuracy may be of importance is when performing more detailed analysis of FDG uptake patterns to elucidate tumor phenotype, such as in the emerging field of radiomics [20-23].

One patient with atypical variation in SUV values from scan to scan was observed. The most possible explanation was extravasation of FDG at the injection site in the elbow. We adjusted these outlying values by applying a cylindrical VOI with a diameter of 1 cm and a long axis of 2 cm in the centre of the descending aorta [24], see Supplementary Tables 2 and 3.

This was a post-hoc correction performed after observing the outlying measurement. The adjustment resulted in narrower SDs and higher regression coefficients in all SUV values but does not change our overall conclusions.

In conclusion, the difference between SUV uptake metrics between DIBH and FB was small compared to the day-to-day variation. The impact of breathing condition was clinically negligible compared to the other uncertainties in conventional SUV metrics. Further, scanning in DIBH did not improve the test-retest stability of FDG uptake metrics. Volume parameters in the nuclear medicine delineation and GTV definitions were equivalently repeatable from day-to-day as metrics defined by 50% of SUV_{max}.

Taken together, DIBH seems unnecessary for measuring conventional FDG uptake metrics in these lung cancer patients.

Acknowledgements

This work was supported by Danish Cancer Society grant no. R72-A4605-13-S2 and NCI grant no. P30 CA 134274-04. The authors further acknowledge financial support from the Kirsten and Freddy Johansen Foundation.

Disclosure of conflict of interest

None.

Address correspondence to: Lotte Nygård, Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, 2100, Copenhagen, Denmark. E-mail: lotte.nygaard@regionh.dk

References

[1] Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by ^{18}F-FDG PET in malignant tumors. J Nucl Med 2008; 49: 1804-8.

[2] Kramer GM, Frings V, Hoetjes N, Hoekstra OS, Smit EF, de Langen AJ, Boellaard R. Repeatability of quantitative whole-body ^{18}F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non-small cell lung cancer patients. J Nucl Med 2016; 57: 1343-1349.

[3] Weber WA, Ziegler SI, Thödtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med 1999; 40: 1771-7.

[4] Aerts HJ, Bosmans G, van Baardwijk AA, Dekker AL, Oellers MC, Lambin P, De Ruysscher D. Stability of ^{18}F-deoxyglucose uptake locations within tumor during radiotherapy for NSCLC: a prospective study. Int J Radiat Oncol Biol Phys 2008; 71: 1402-7.

[5] Pépin A, Daouk J, Bailly P, Hapdey S, Meyer ME. Management of respiratory motion in PET/computed tomography: the state of the art. Nucl Med Commun 2014; 35: 113-22.

[6] Na F, Wang J, Li C, Deng L, Xue J, Lu Y. Primary tumor standardized uptake value measured on F18-Fluorodeoxyglucose positron emission tomography is of prediction value for survival and local control in non-small-cell lung cancer receiving radiotherapy: meta-analysis. J Thorac Oncol 2014; 9: 834-42.

[7] van Elmpt W, De Ruysscher D, van der Salm A, Lakeman A, van der Stoep J, Emans D, Damen E, Öllers M, Sonke JJ, Belderbos J. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. Radiother Oncol 2012; 104: 67-71.

[8] Kini VR, Vedam SS, Keall PJ, Patil S, Chen C, Mohan R. Patient training in respiratory-gated radiotherapy. Med Dosim 2003; 28: 7-11.

[9] Damkjær SM, Aznar MC, Pedersen AN, Vogelius IR, Bangsgaard JP, Josipovic M. Reduced lung dose and improved inspiration level reproducibility in visually guided DIBH compared to audio coached EIG radiotherapy for breast cancer patients. Acta Oncol (Madr) 2013; 52: 1458-1463.

[10] Josipovic M, Persson GF, Dueck J, Bangsgaard JP, Westman G, Specht L, Aznar MC. Geometric uncertainties in voluntary deep inspiration breath hold radiotherapy for locally advanced lung cancer. Radiother Oncol 2016; 118: 510-4.

[11] Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, Verzijlbergen FJ, Barrington SF, Pike LC, Weber WA, Stroobants S, Delbeke D, Donohoe KJ, Holbrook S, Graham MM, Testanera G, Hoekstra OS, Zijlstra J, Visser E, Hoekstra CJ, Pruim J, Willemsen A, Arends B, Kotzerke J, Bockisch A, Beyer T, Chiti A, Krause BJ. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging 2015; 42: 328-354.

[12] Nygård L, Vogelius IR, Fischer BM, Klausen TL, Langer SW, Lonsdale MN, Persson GF, Bentzen SM. Early lesion-specific (18)F-FDG PET response to chemotherapy predicts time to lesion progression in locally advanced non-small cell lung cancer. Radiother Oncol 2016; 118: 460-464.

[13] Wolthaus JW, Sonke JJ, van Herk M, Belderbos JS, Rossi MM, Lebesque JV, Damen EM. Comparison of different strategies to use four-dimensional computed tomography in treatment planning for lung cancer patients. Int J Radiat Oncol Biol Phys 2008; 70: 1229-1238.

[14] Team RDC. R: a language and environment for statistical computing. http://www.r-project.org/2008; http://www.r-project.org/.

[15] Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-porta R, Postmus PE, Rusch V. The IASLC lung cancer staging project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. J Thorac Oncol 2009; 2: 706-714.

[16] Mittauer KE, Deraniyagala R, Li JG, Lu B, Liu C, Samant SS, Lightsey JL, Yan G. Monitoring ABC-assisted deep inspiration breath hold for left-sided breast radiotherapy with an optical tracking system. Med Phys 2014; 42: 134-143.

[17] Bentzen SM, Gregoire V. Molecular imaging-based dose painting: a novel paradigm for radiation therapy prescription. Semin Radiat Oncol 2011; 21: 101-10.

[18] Even AJ, van der Stoep J, Zegers CM, Reymen B, Troost EG, Lambin P, van Elmpt W. PET-based dose painting in non-small cell lung cancer: comparing uniform dose escalation with boosting hypoxic and metabolically active subvolumes. Radiother Oncol 2015; 116: 281-286.

[19] Møller DS, Nielsen TB, Brink C, Hoffmann L, Lutz CM, Drøgemüller Lund M, Hansen O, Schytte T, Khalil AA, Knap MM, Nyhus CH, Ottosson W, Sibolt P, Borissova S, Josipovic M, Persson G, Appelt AL. Heterogeneous FDG-guided dose-escalation for locally advanced NSCLC (the NARLAL2 trial): design and early dosimetric results of a randomized, multi-centre phase-III study. Radiother Oncol 2017; 124: 311-317.

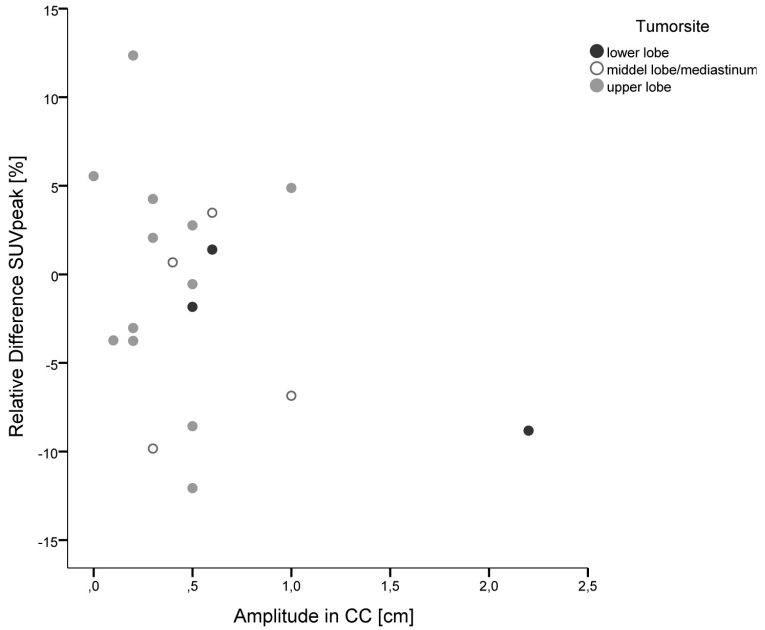
[20] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014; 5: 4006.

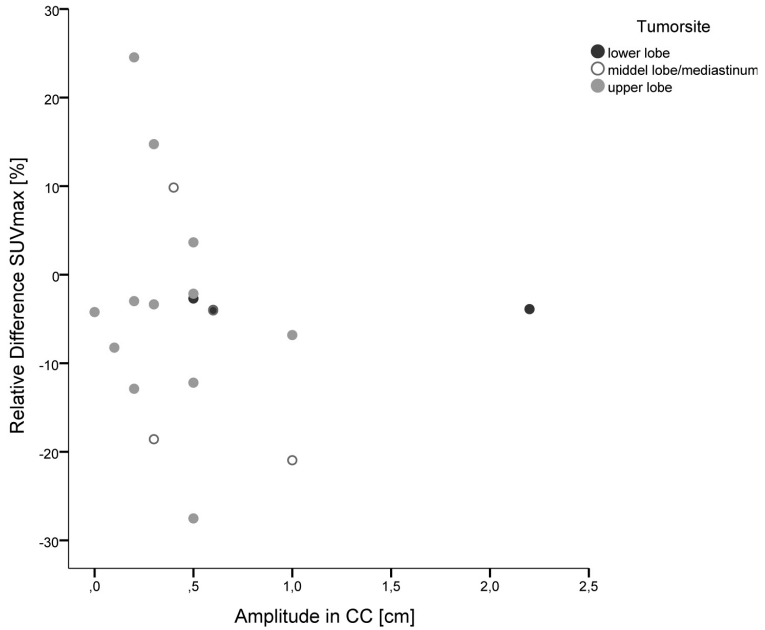
[21] Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH, Gillies R, Quackenbush J, Aerts HJWL. Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 2017; 77: 3922-3930.

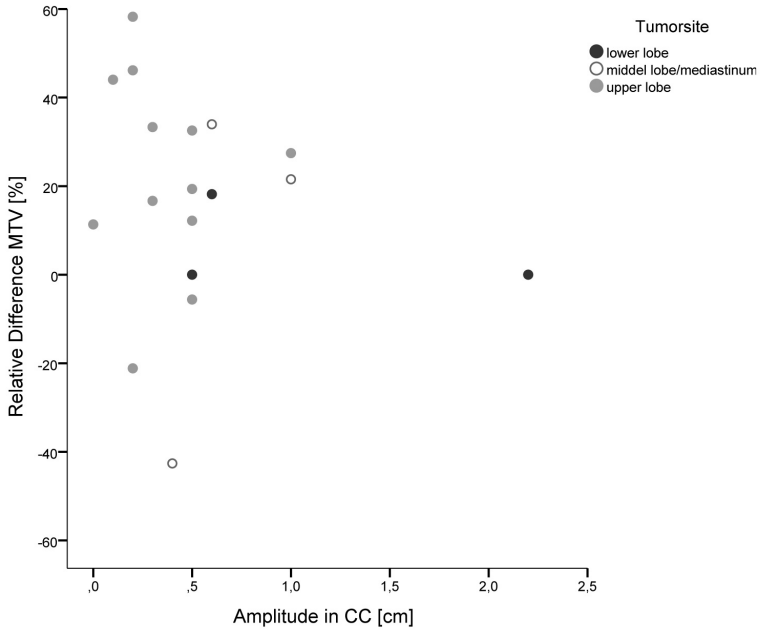
[22] Yip SS, Kim J, Coroller TP, Parmar C, Velazquez ER, Huynh E, Mak RH, Aerts HJ. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med 2017; 58: 569-576.

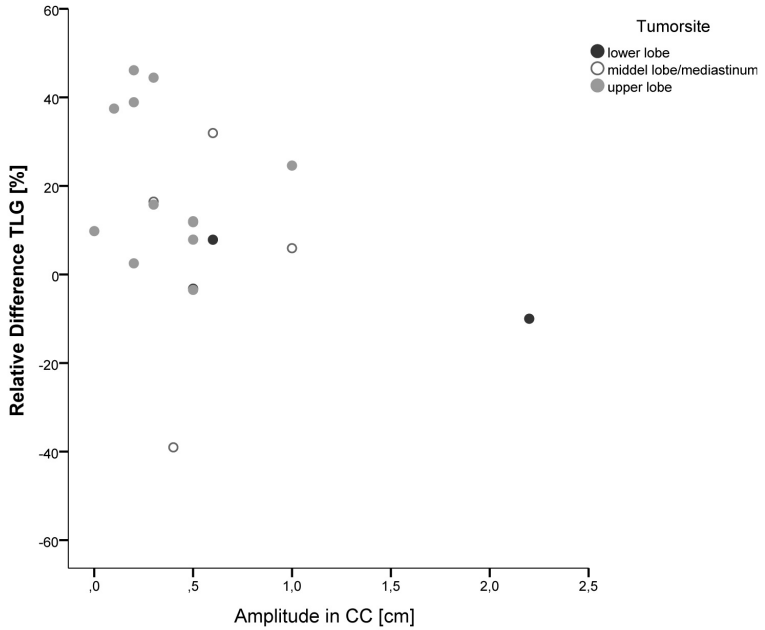
[23] Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, Hoekstra CJ, Boellaard R, Dekker AL, Gillies RJ, Aerts HJ, Lambin P. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 2013; 52: 1391-7.

[24] O JH, Lodge MA, Wahl RL. Practical PERCIST: a simplified guide to PET response criteria in solid tumors 1.0. Radiology 2016; 280: 576-84.


Tumorsite
lower lobe
middel lobe/mediastinum
upper lobe
Relative Difference SUVpeak [%]
Amplitude in CC [cm]


Tumorsite
lower lobe
middel lobe/mediastinum
upper lobe
Relative Difference SUVmax [%]
Amplitude in CC [cm]


Tumorsite
lower lobe
middel lobe/mediastinum
upper lobe
Relative Difference MTV [%]
Amplitude in CC [cm]


Tumorsite
lower lobe
middel lobe/mediastinum
upper lobe
Relative Difference TLG [%]
Amplitude in CC [cm]

Supplementary Figure 1. Amplitude. Scatter plot of relative difference [%] between FB and DIBH in scan 1 of SUV values (50% isocontour of SUV_{max}) as a function of amplitude of tumor motion in the cranio-caudal direction [cm] from the 4DCT radiotherapy planning scan. SUV: Standardized uptake value. TLG: Total lesion glycolysis. MTV: Metabolic tumor volume. FB: free breathing. DIBH: deep inspiration breath hold.

Supplementary Table 1. Day to day Difference/mean difference from day to day in FB and DIBH in eight patients with tumor motion >0.5 cm in either CC or AP direction. Tumor motion amplitude was measured clinically from 4DCT data for calculation of treatment margins

Patient number	CC Motion in FB [cm]	AP Motion in FB [cm]	Tumor Site	SUV_{peak} Day 1/Day 2 (FB/DIBH) [%]	SUV_{max} Day 1/Day 2 (FB/DIBH) [%]	SUV_{mean} Day 1/Day 2 (FB/DIBH) [%]	MTV Day 1/Day 2 (FB/DIBH) [%]	TLG Day 1/Day 2 (FB/DIBH) [%]
1	**0.6**	0.2	Lower lobe	-13/4.3	-23.7/-10.4	-22.5/-4.3	18.2/22.2	9.9/27.9
2	**1.0**	0.3	Upper lobe	-7.6/-9.5	-16.2/-20.0	-14.6/-16.5	18.8/44.4	8.6/25.6
4	0.5	**0.6**	Upper lobe	6.6/16.7	-0.7/29.0	4.2/23.3	17.4/-36.4	22.5/-10.0
5	**1.0**	**0.85**	Mediastinum	-8.1/-6.4	-13.9/-12.5	-10.8/-9.4	0/-6.7	-10.9/-17.5
6	**2.2**	**0.6**	Lower lobe	-11.6/-10.7	-9.1/-34.7	-8.1/-27.2	-13.3/80	-26/47.1
13	**0.6**	0.3	Mediastinum	-40.8*/-44.7*	-34.6/-51.0*	-37.9*/-49.5*	-15/-4.4	-52.5/-53.6
16	0.3	**0.6**	Mediastinum	5.3/11.6	-1.2/15.1	0/16.1	24/-26	21.0/-10.1
18	0.4	**0.6**	Mediastinum	-11.5/8.0	-5.9/-0.7	-6.6/-2.9	-9.2/13.6	-16.8/-16.4

CC: cranio caudal direction. AP: Anterior posterior direction. Bold: tumor amplitude >0.5 cm. SUV: Standardized uptake value. MTV: Metabolic tumor volume. TLG: Total lesion glycolysis. *: repeatability poorer than expected (± 1.96 × SD) from results in **Table 2**.

Supplementary Table 2. Logistic regression analysis from day to day (R2) and Relative difference [%] from day to day (SD in descriptives) in the two breathing conditions using adjusted data for patient number 10 by Practical PERCIST [24]

Adjusted	50% isocontur of SUV_{max}				Nuclear medicine delineation	
	SUV_{peak}	SUV_{max}	SUV_{mean}	TLG	SUV_{mean}	TLG
FB_{Day1+2}						
R^2	0.93	0.93	0.91	0.98	0.70	0.98
SD [%]	14.3	16.5	16.6	31.8	20.5	34.8
$DIBH_{Day1+2}$						
R^2	0.93	0.87	0.84	0.95	0.81	0.97
SD [%]	14.7	20.7	20.8	39.6	14.6	25.5

FB: Free breathing. DIBH: Deep inspiration breath hold. SD: Standard deviation. SUV: Standardized uptake value. MTV: Metabolic tumor volume. TLG: Total lesion glycolysis.

Supplementary Table 3. Paired comparison of FB versus DIBH. Relative differences between the two breathing modalities [%]. Adjusted data for patient number 10 by Practical PERCIST [24]

Adjusted	Day 2 (FB versus DIBH)		
50% isocontur of SUV_{max}	Mean difference [%]	SEM [%]	*p*-value
SUV_{peak}	-2.0	1.4	0.1
SUV_{max}	-5.4	2.5	0.03
SUV_{mean}	-4.7	2.4	0.05
TLG	-0.8	4.8	0.86
Nuclear medicine delineation			
SUV_{mean}	-5.3	2.1	<0.0001
TLG	10.1	5.2	0.05

FB: Free breathing. DIBH: Deep inspiration breath hold. SEM: Standard error of the mean (=95% CI from the T-test divided by (1.96 × 2=3.92)). SUV: Standardized uptake value. TLG: Total lesion glycolysis.