BJR|Open https://doi.org/10.1259/bjro.20190029

Received: 14 June 2019 | Accepted: 09 July 2019

Cite this article as:
Carlin D, Weller A, Kramer G, Liu Y, Waterton JC, Chiti A, et al. Evaluation of diffusion-weighted MRI and (18F) fluorothymidine-PET biomarkers for early response assessment in patients with operable non-small cell lung cancer treated with neoadjuvant chemotherapy. *BJR Open* 2019; **1**: 20190029.

ORIGINAL RESEARCH

Evaluation of diffusion-weighted MRI and (18F) fluorothymidine-PET biomarkers for early response assessment in patients with operable non-small cell lung cancer treated with neoadjuvant chemotherapy

[1]DOMINIC CARLIN, PhD, [1,2]ALEXANDER WELLER, MD, FRCR, [3]GEM KRAMER, PhD, [4]YAN LIU, MD, MS, PhD, [5]JOHN C WATERTON, MA PhD CSci FRSC, [6,7]ARTURO CHITI, MD, [6]MARTINA SOLLINI, PhD, [3,8]A JOOP DE LANGEN, PhD, [2]MARY E R O'BRIEN, MD, [4]MARIA URBANOWICZ, [4]BART KM JACOBS, MD, PhD and [1,2]NANDITA DESOUZA, MD, FRCR

[1]CRUK Imaging Centre, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK
[2]The Royal Marsden Hospital, Downs Road, Sutton, Surrey SM2 5PT, UK
[3]Department of Respiratory Diseases, VU University Medical Center, Amsterdam, The Netherlands,
[4]EORTC Headquarters, Brussels, Belgium
[5]Centre for Imaging Sciences, Division of Informatics Imaging & Data Sciences, School of Health Sciences, Faculty of Biology Medicine & Health, University of Manchester, Manchester Academic Health Sciences Centre, Oxford Road Manchester M13 9PL UK,
[6]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[7]Nuclear Medicine, Humanitas Clinical and Research Centre, Milan, Italy
[8]Department of Thoracic Oncology, Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands,

Address correspondence to: Dr Nandita deSouza
E-mail: *nandita.desouza@icr.ac.uk*

Objective: To correlate changes in the apparent diffusion coefficient (ADC) from diffusion-weighted (DW)-MRI and standardised uptake value (SUV) from fluorothymidine (^{18}FLT)-PET/CT with histopathological estimates of response in patients with non-small cell lung cancer (NSCLC) treated with neoadjuvant chemotherapy and track longitudinal changes in these biomarkers in a multicentre, multivendor setting.

Methods: 14 patients with operable NSCLC recruited to a prospective, multicentre imaging trial (EORTC-1217) were treated with platinum-based neoadjuvant chemotherapy. 13 patients had DW-MRI and FLT-PET/CT at baseline (10 had both), 12 were re-imaged at Day 14 (eight dual-modality) and nine after completing chemotherapy, immediately before surgery (six dual-modality). Surgical specimens (haematoxylin-eosin and Ki67 stained) estimated the percentage of residual viable tumour/necrosis and proliferation index.

Results: Despite the small numbers,significant findings were possible. ADC_{median} increased ($p < 0.001$) and SUV_{mean} decreased ($p < 0.001$) significantly between baseline and Day 14; changes between Day 14 and surgery were less marked. All responding tumours (>30% reduction in unidimensional measurement pre-surgery), showed an increase at Day 14 in ADC75th centile and reduction in total lesion proliferation (SUV_{mean} x proliferative volume) greater than established measurement variability. Change in imaging biomarkers did not correlate with histological response (residual viable tumour, necrosis).

Conclusion: Changes in ADC and FLT-SUV following neoadjuvant chemotherapy in NSCLC were measurable by Day 14 and preceded changes in unidimensional size but did not correlate with histopathological response. However, the magnitude of the changes and their utility in predicting (non-) response (tumour size/clinical outcome) remains to be established.

Advances in knowledge: During treatment, ADC increase precedes size reductions, but does not reflect histopathological necrosis.

INTRODUCTION

Non-small cell lung cancer (NSCLC) stage II and occasionally IIIA is routinely managed with surgery and adjuvant chemotherapy.[1] Persistently poor survival rates[2] have driven the use of induction (neoadjuvant) chemotherapy to reduce pre-surgical tumour burden[3–5]; results are favourable and equivalent to those where adjuvant chemotherapy is used.[6,7] During neoadjuvant chemotherapy, response assessment relies on size-based (RECIST) measurements,[8] which often only decrease 2–3 months after treatment initiation.[9] Moreover, improved

survival may be seen without a reduction in tumour size because of decreased metastatic propensity with cytostasis and post-treatment oedema.[9,10] Therefore, early indicators of treatment response sensitive to biological changes in the tumour are desirable. Imaging techniques, such as diffusion-weighted MRI (DW-MRI) and PET, probe the tumour microenvironment and provide this potential.

DW-MRI provides an apparent diffusion coefficient (ADC), a biomarker of cellularity,[11] which has been shown to increase in many cancers on treatment due to cell death.[12,13] (18F)-fluorothymidine (FLT) uptake on PET, is also a promising biomarker for treatment response evaluation since its uptake is related to tumour cell proliferation.[14,15] Some observational studies indicate that early reduction in FLT uptake correlates with later changes in tumour size[16–18] others do not.[19]

This study was part of a larger programme validating imaging biomarkers for early response assessment in drug development.[20,21] Neither ADC nor FLT uptake have been jointly evaluated with reference to histopathology in NSCLC: we hypothesised that they were related to pathological indices of response (ADC to necrosis and FLT Standardised Uptake Value (SUV) to proliferation) and correlated their changes in with histological measures of response (necrosis) and non-response (residual proliferative activity). We also studied longitudinal patterns of ADC and FLT-SUV change and linked them to RECIST-based size criteria for response and non-response. This international multicentre study was performed under quality-controlled conditions for multimodality, quantitative imaging[22,23] to ensure comparable measurements on multiple scanners at multiple sites.

METHODS

Patients

This prospective, multicentre, single-arm imaging trial (EORTC-1217) listed on clinical trials.gov (NCT02273271) recruited patients with confirmed clinical stage II-IIIA NSCLC eligible for curative intent surgery at three participating centres (Royal Marsden Hospital, UK; Humanitas, Milan; VUMC, Amsterdam, The Netherlands). 14 patients (nine male, five female) aged 53–78 years (mean 65.9 ± 7.1 years) gave written, informed consent (Table 1).

Patients received platinum-based neoadjuvant chemotherapy without pemetrexed. They had DW-MRI and FLT-PET/CT between 1.10.2015 and 06.04.2017 at three time-points: i) baseline, before chemotherapy; ii) 14 days post-treatment and iii) within 1 week of surgery, 3–5 weeks after completing chemotherapy. Figure 1 summarises the study design.

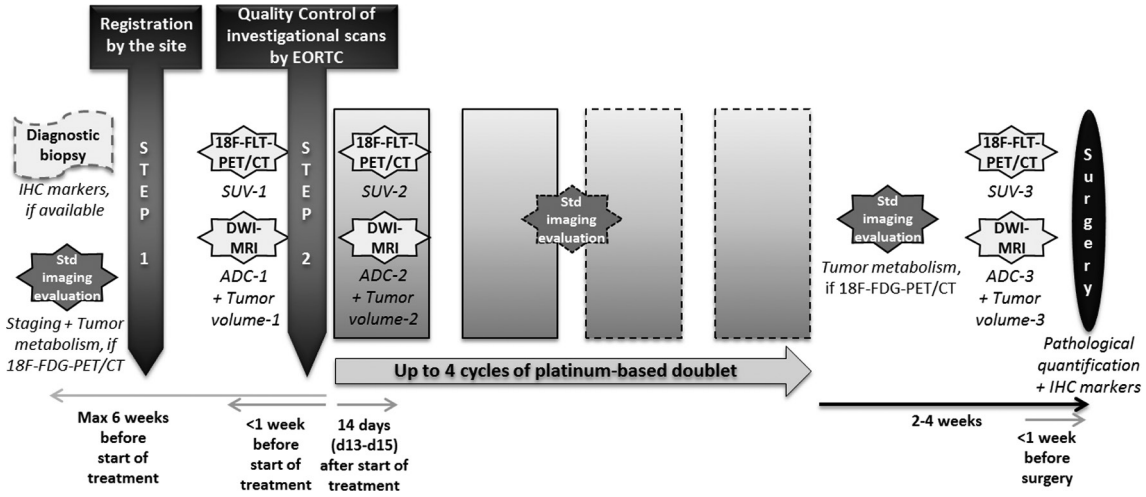
MRI

Patients were scanned on 4 clinical 1.5T MR platforms [Optima (GE Healthcare, Waukesha, WI); Avanto (Siemens AG, Erlangen, Germany); 2 Achieva (Philips Healthcare, Best, The Netherlands)] using phased-array body coils. All centres performed regular test-object scans for quality assurance (QA)/quality control (QC) and had previously participated in a technical validation phase to establish measurement reproducibility when performed in accordance with prescribed QA procedures.[24,25]

Table 1. Baseline characteristics of all non-small cell lung cancer (NSCLC) patients involved in the trial. Patient three withdrew consent prior to imaging and chemotherapy start

Patient	Gender	Age	Histopathology	Stage	Number of cycles	Platinum-based compound	Cytotoxic Compound	Reason for chemotherapy discontinuation
1	F	78	NSCLC	Stage IIIA	2	Carboplatin	Vinorelbine	Toxicity: lung infection and pulmonary embolism
2	M	71	NSCLC	Stage IIIA	3	Carboplatin	Gemcitabine	Toxicity: neutropenia
4	F	71	NSCLC	Stage IIIA	3	Carboplatin	Vinorelbine	Normal completion
5	M	57	NSCLC	Stage IIIA	3	Carboplatin	Vinorelbine	Normal completion
6	M	69	NSCLC	Stage IIA	3	Cis(1)/Carbo(2)	Vinorelbine	Normal completion
7	F	53	NSCLC	Stage IIIA	3	Cisplatin	Vinorelbine	Normal completion
8	M	68	NSCLC	Stage IIB	2	Cisplatin	Gemcitabine	Progressive disease
9	M	69	NSCLC	Stage IIIA	3	Carboplatin	Vinorelbine	Normal completion
10	M	74	NSCLC	Stage IIA	3	Carboplatin	Vinorelbine	Normal completion
11	F	60	NSCLC	Stage IIIA	3	Cis(2)/Carbo(1)	Gemcitabine	Normal completion
12	F	69	NSCLC	Stage IIIA	2	Cisplatin	Vinorelbine	Progressive disease
13	M	62	NSCLC	Stage IIIA	3	Cisplatin	Gemcitabine	Toxicity: thrombocytopenia
14	M	61	NSCLC	Stage IIIA	3	Cis(2)/Carbo(1)	Gemcitabine	Normal completion

Figure 1. Design of this prospective, multicentre, single-arm imaging trial from registration to surgery.



Axial T_1 weighted turbo spin-echo in breath-hold and three-dimensional T_2 weighted turbo spin-echo with variable flip-angle images were obtained as per local protocols.

DW-MRI was acquired during free-breathing using a single-shot echo-planar technique with short-tau inversion recovery fat suppression. Two blocks of 30 slices were acquired through the chest (thickness 5 mm, no gap), field-of-view 320 × 280 mm, four signal averages or four acquisitions, 3 b-values (100, 500, 800 s/mm^2).

FIt PET/CT

FLT was synthesised either in-house (Humanitas Clinical and Research Center, and VUMC) or by an external supplier (PETNET solutions Incorporated, Nottingham, UK). PET/CT was performed on a Discovery 690 (General Electric Healthcare, Waukesha, WI), Siemens Biograph 128m and Biograph 64 or Philips Gemini TF TOF 64 scanner in patients fasted for 4 h. Images were acquired without respiratory gating in a single bed position 60 min after injection of a 4MBq/kg FLT bolus. A non-contrast, non-breathhold CT scan was also obtained. PET/CT images were corrected for attenuation using low-dose CT scans according to EANM guidelines.[26] Scanner accreditation through the EARL initiative (EANM Research Ltd; earl.eanm.org) guaranteed measurements' reproducibility and comparability, across patient's scans and among centers.

IMAGING BIOMARKERS ANALYSIS

Each patient was imaged using the same scanner across all visits. Central review of images was performed. EARL approved reconstruction was used for PET/CT image analysis. ADC and SUV were calculated at baseline (ADC_1, SUV_1), Day 14 (ADC_2, SUV_2), and pre-surgery (ADC_3, SUV_3).

For DW-MRI, ROIs were manually segmented on computed b = 800 s/mm^2 images with anatomical reference to T1-W, T2-W, ADC and b = 100 s/mm^2 images on all slices with identifiable tumour. ROIs were copied onto corresponding ADC maps in Osirix. Whole tumour ADC histograms were generated. The median ADC (ADC_{median}, primary endpoint), the 25th and 75th centiles and interquartile range (IQR) were recorded at each time-point. DWI-derived lesion volume at each time-point was estimated by multiplying total number of voxels by voxel dimensions.

All lesions were delineated using a 50% isocontour of the SUV_{peak} corrected for local background to derive tumour volumes of interest (VOI) designated as the proliferative volume. SUV was calculated within the VOI. SUV_{mean} (primary endpoint), SUV_{max}, SUV_{peak} and total lesion proliferation (TLP, product of SUV_{mean} and proliferative volume) were recorded.

TUMOUR RESPONSE ASSESSMENT

Percentage change of imaging biomarkers at Day 14 (ΔADC_{early}, ΔFLT_{early}) and pre-surgery (ΔADC_{post}, ΔFLT_{post}) in comparison to baseline values were calculated, where ΔADC_{early} = 100%x(ADC_2–ADC_1)/ADC_1, ΔSUV_{early} = 100%x(SUV_2–SUV_1)/SUV_1, ΔADC_{post} = 100%x(ADC_3–ADC_1)/ADC_1 and ΔSUV_{post} = 100%x(SUV_3–SUV_1)/SUV_1

Response between baseline and surgery was evaluated on CT images using longest uni-dimensional axis of the primary lesion defined by RECIST1.1 criteria (reduction in target lesion diameter >30%).[8] Patients with appearance of new lesions who did not have further imaging were also classed as non-responders. The DWI and FLT images were not used to re-stage the patients, as this was outside the primary aim.

HISTOPATHOLOGY

Tumours were assessed on haematoxylin-eosin (HE) stained slides from surgical resection specimens by a central review pathologist (μ). Tumour surface area (mm^2) was calculated on a maximum of 10 representative slides per lesion, along with areas of viable tumour, necrosis and fibrosis. Total tumour surface area was calculated by summing values from all slides; total viable tumour surface area and total necrotic tissue area were obtained similarly. The percentage of viable tumour or necrosis for each lesion was estimated from 100%x (total viable tumour surface area) or (total necrotic tissue surface area) respectively/(total tumour surface area).

Table 2. Imaging and surgical status for individual participants

Patient ID	Baseline DWI	Baseline FLT	Day 14 DWI	Day 14 FLT	Pre-surgical DWI	Presurgical FLT	Surgical sample
1	√		√				
2	√	√	√		√	√	√
3	WITHDREW						
4	√	√		√		√	√
5	√		√		√		
6		√					
7	√	√	√	√	√	√	√
8	√	√	√	√			
9	√	√	√	√	√	√	√
10	√	√	√	√	√	√	√
11	√	√	√	√	√	√	√
12	√	√	√	√			
13	√	√	√	√		√	√
14	√	√	√	√	√	√	√

Table 3. Change in long axis with treatment indicating three responders by RECIST criteria at the pre-surgical time-point

Patient ID	Long axis baseline (mm)	Long axis Day 14 (mm)	Change Day 14 (%)	Long axis pre-surgery (mm)	Change pre-surgery (%)
1	59	53	−10.1		
2	60	49	−18.3	50	−16.7
3					
4	33	30	−9.1	25	−24.2
5	41	40	−2.4	44	7.3
6					
7	36	43	19.4	32	−11.1
8	68	68	0		
9	43	42	−2.3	35	−18.6
10	56	55	−1.8	52	−7.1
11	63	40	−36.5	16	−74.6
12	71	69	−2.8	70	−1.4
13	70	58	−17.1	48	−31.4
14	73	54	−26.0	48	−34.2

Table 4. Median and (range) of primary lesion values of DW-MRI and FLT-PET metrics at baseline for all patients, responders and non-responders. Note: all patients for DW-MRI and FLT-PET each include a patient who did not have response assessed

DW-MRI	ADC_{median} ($x10^{-3}$ mm^2/s)	ADC_{25th} ($x10^{-3}$ mm^2/s)	ADC_{75th} ($x10^{-3}$ mm^2/s)	ADC_{IQR} ($x10^{-3}$ mm^2/s)
Baseline (*n* = **12**)	1.14	0.91	1.33	0.42
	(0.80–1.49)	(0.56–1.26)	(0.93–1.42)	(0.20–0.56)
Responders (n = 3)	0.89	0.74	1.07	0.33
	(0.82–1.29)	(0.72–1.09)	(0.93–1.59)	(0.20–0.50)
Non-Responders (n = 8)	1.16	0.91	1.36	0.46
	(0.80–1.49)	(0.56–1.26)	(1.05–1.73)	(0.22–0.56)
FLT-PET	SUV_{max}	SUV_{peak}	SUV_{mean}	TLP
Baseline (*n* = **11**)	6.21	4.78	3.54	141
	(4.9–14.0)	(3.1–12.0)	(2.4–8.2)	(38–2870)
Responders (n = 3)	11.2	9.23	6.74	379
	(9.9–14.0)	(8.4–12.0)	(6.7–8.2)	(141 – 522)
Non-Responders (n = 7)	5.62	4.72	3.11	95.3
	(4.9–11.7)	(3.1–9.9)	(2.4–7.6)	(38–2870)

Tumour cell proliferation was assessed by immunohistochemistry (Ki67 staining).[27] Positive nuclei were calculated on each slide in at least 100 cells in the representative areas of the tumour and expressed as a percentage. The mean Ki67 index from all reviewed slides was the final sample score.

STATISTICAL CONSIDERATIONS

Co-primary imaging endpoints were ΔADC_{early} defined on ADC_{median} and ΔSUV_{early} defined on SUV_{mean}; alternative ADC and SUV measures were secondary endpoints. The primary pathology endpoint measures were percentage of viable tumour cells; percentage of necrosis and proliferative activity (Ki67 index) were secondary endpoints. Responder status was based on unidimensional size and included patients who did not undergo surgery. 31 lesions were needed to demonstrate with 95% confidence (one-sided) with 90% power that the absolute correlation between the imaging biomarker change and the pathological response was >0.5 (H0: rho ≤0.5) if the true correlation is 0.8 (H1: rho >0.5).

Statistical analysis used SAS9.4 (STAT14.3). Graphs were generated using Excel (version14.0.7212.5000) and validated in SAS. Correlations between continuous endpoints were assessed using the Spearman rank correlation test with Fisher transformation and assumed to be positive between ΔSUV_{early} and %viable tumour ($\rho_0 = 0.5$), and negative between ΔADC_{early} and %viable tumour ($\rho_0 = -0.5$). Equivalent two-sided 90% confidence intervals are reported to check futility.

The one-sided type I error was fixed at 1% for all secondary and exploratory analyses. When relevant, equivalent two-sided 98% confidence intervals are reported.

Longitudinal comparisons of ADC, SUV and volumes were analysed with a two-sided Wilcoxon signed-rank test.

RESULTS

14 patients were successfully screened and 13 finally participated (one withdrew). 13 patients had DW-MRI and[16]FLT-PET/CT at baseline (10 had both), 12 were re-imaged at Day 14 (eight dual-modality) and nine after completing chemotherapy, immediately before surgery (six dual-modality). four patients did not have surgery, three due to progressive disease and one due to toxicity. An overview of imaging scans and surgery status is presented in Table 2.

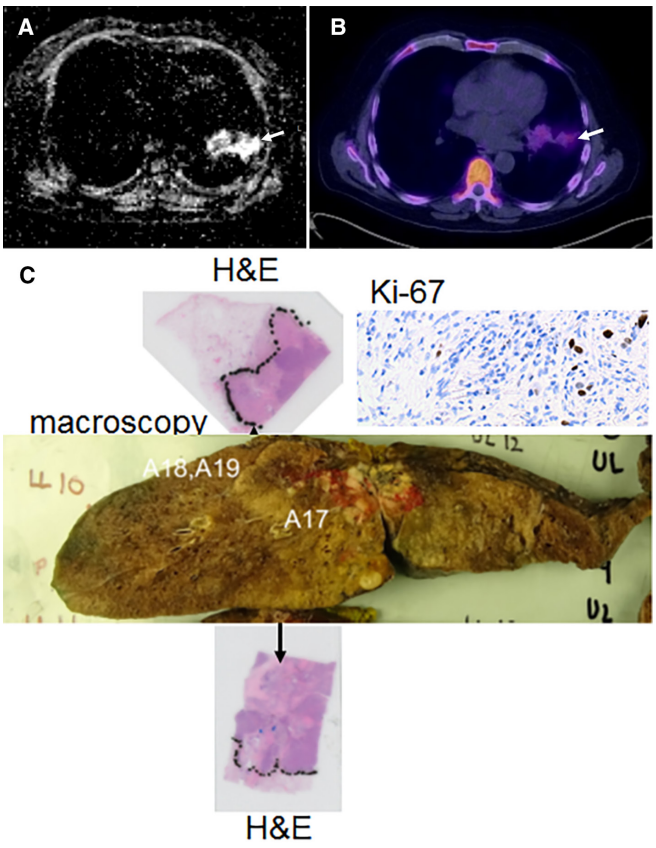
Of the 12 patients imaged at Day 14, three met the criteria for response of the primary lesion (30% reduction in unidimensional measurement) on CT performed prior to surgery (Table 3). Eight patients were classed as non-responders; the numbers of patients imaged at each time-point with each imaging modality are given in Table 4. The remaining patient was not included in the response assessment because toxicity to Carboplatin plus Vinorelbine prompted a change to the targeted agent Erlotinib.

Re-staging the patients with either DWI or FLT was outside the remit of the study. Also repeat mediastinoscopy was not routinely performed. Repeat CT scan pre-surgery detected a new lesion in one patient of our cohort, who was therefore classified as a non-responder.

CORRELATION OF IMAGING PARAMETERS AND PATHOLOGICAL MEASURES

Early changes in imaging parameters were compared with pathological measures (Figure 2). In patients who underwent surgery and had imaging at Day 14 (*n* = 8), there was no meaningful

Figure 2. A 70-year-old patient with lung cancer treated with platinum based neoadjuvant chemotherapy: ADC map derived from DW-MRI prior to surgery (a), FLT-PET at the same time-point (b), show residual tumour with lateral consolidation (solid white arrows). Following resection, this was corroborated on the corresponding slice of the macroscopic surgical specimen (c, centre). Residual tumour was confirmed on H&E stained sections. Ki-67 stain shows the low level of proliferative activity (brown stain) within the centre of the tumour.



correlation at the 1% significance level between ΔADC_{early} and the %viable cells ($r = 0.50$, $p = 0.99$, 90% CI = [−0.30; 0.87]) or ΔSUV_{early} and the %viable cells ($r = 0.07$, $p = 0.85$, 90% CI = [−0.64; 0.71]). The two-sided confidence intervals did not reach the $\rho = -0.8$ and $\rho = 0.8$ respectively correlation as originally hypothesised. Similarly, there were no meaningful correlations between presurgical ADC_{median} and %necrosis ($r = 0.26$, $p = 0.56$, 98% CI = [−0.80; 0.92]). The observed correlation between Ki-67 and presurgical SUV_{mean} was higher than $\rho = 0.8$ as originally hypothesised ($r = 0.86$, $p = 0.08$, 98% CI = [0.05; 0.98]).

LONGITUDINAL PATTERNS OF RESPONSE TO TREATMENT

Baseline ADC and SUV values from DWI and FLT imaging respectively at each time-point are presented in Table 4**.** The relative changes in imaging metrics are presented in Table 5 for DWI and Table 6 for FLT. The ADC_{median} increased ($p < 0.001$) and SUV_{mean} decreased ($p < 0.001$) significantly between baseline and Day 14, however the change between Day 14 and surgery was less marked (Table 5), indicating that this parameter changes early. The changes in ADC_{median}, ADC_{75th} and ADC_{IQR} at Day 14 relative to baseline were generally larger in responders than non-responders (Figure 3a–c) although small numbers precluded statistical evaluation. Early changes in ADC_{25th}, however, were comparable between responders and non-responders.

The decreases in SUV_{mean}, SUV_{max}, SUV_{peak} and TLP at Day 14 relative to baseline were generally larger in responders than non-responders (Figure 3d–f); again small numbers precluded statistical evaluation.

CHANGES WITH RESPECT TO MEASUREMENT REPEATABILITY

Changes in imaging biomarkers at Day 14 were also assessed with reference to previously established test-retest estimates of measurement repeatability in order to exclude changes that represent measurement variability.[24,25] (Figure 4). Increases in ADC_{median} greater than the limits of measurement repeatability were seen in four patients (2/3 responders, 2/7 non-responders). The measurable increases in ADC_{75th} were seen on all three responders but only in 1/7 non-responders, while ADC_{25th} only increased in 1 responder and two non-responders. Measurable decreases at Day 14 were observed for SUV_{mean} in seven patients (all 3 responders and 4/6 non-responders) SUV_{max} (2/3 responders, 5/6 non-responders) and in six patients for SUV_{peak} (2/3 responders, 4/6 non-responders), and TLP (all 3 responders and 3/6 non-responders).

COMPARISON OF IMAGING VOLUMES

Tumour volumes generally decreased between baseline and Day 14 for all imaging modalities, with mean ± standard deviation

Table 5. Mean ± standard deviation and range of percentage change of DWI metrics relative to baseline at Day 14 and within 1 week of surgery

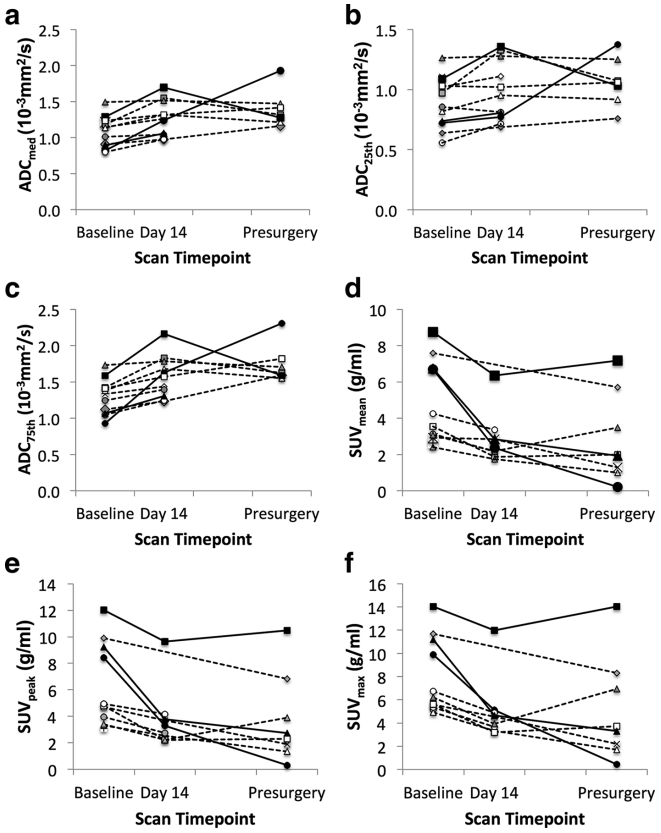
	Mean ± SD ΔDWI % (Range)				
	Volume	ADC_{median}	ADC_{25th}	ADC_{75th}	ADC_{IQR}
Day 14	−18.4 ± 30.8	18.2 ± 15.1	12.2 ± 13.1	22.6 ± 20.1	54.7 ± 90.3
All patients ($n = 11$)	(-86.6–29.2)	(1.4–51.1)	(-5.0–36.7)	(3.0–75.6)	(3.0–319.6)
Pre-surgery	−30.1 ± 41.7	27.7 ± 48.5	18.3 ± 32.7	35.0 ± 52.5	81.8 ± 126.7
All patients ($n = 7$)	(-99.8–39.6)	(-1.4–135.3)	(-5.4–90.0)	(-1.4–148.7)	(-2.4–356.9)

Table 6. Mean ± standard deviation and range of percentage change of FLT metrics relative to baseline at Day 14 and within 1 of surgery

	Mean ± SD ΔFLT (Range)				
	Volume	SUV_{max}	SUV_{peak}	SUV_{mean}	Total lesion proliferation
Day 14	−21.0 ± 38.6	−35.0 ± 13.9	−28.4 ± 33.1	−35.1 ± 18.9	−49.5 ± 23.1
All patients (n = 9)	(-68.0–55.2)	(-58.3–−14.7)	(-61.1–47.9)	(-65.7–−3.0)	(-89.7–−18.4)
Pre-surgery	−19.6 ± 41.0	−43.2 ± 37.0	−45.4 ± 34.7	−44.3 ± 34.9	−52.8 ± 32.9
All patients (n = 8)	(-91.7–69.2)	(-95.6–11.9)	(-96.3–14.4)	(-97.0–13.6)	(-98.9–−4.7)

volume reductions of 20.9±33.8% for DWI (n = 9, p = 0.03), 21.0±38.6% for FLT (n = 9, p = 0.16) and 26.4±30.5% for CT (n = 10, p = 0.03); volume reductions between techniques were not significantly different at the 1% significance level. For patients in whom the pre-surgery scan was available, DWI (n = 7) and CT-derived (n = 8) tumour volumes showed average reductions of 41.7±30.8% for DWI and 58.0±23.3% for CT from baseline. The mean pre-surgical FLT volume reduction (n = 8) was 19.6±41.0% compared to baseline (p = 0.1). Comparative volume changes are illustrated in Figure 5.

Figure 3. Absolute values of (a) ADC_{median}, (b) ADC_{25th} and c) ADC_{75th} , (d) SUV_{mean}, (e) SUV_{peak}, and (f) SUV_{max} at each time-point (solid lines =responders, dashed lines=non-responders) show the increase in ADC values and decrease in FLT-SUV parameters with time on treatment.



There was no significant correlation between the ΔADC_{median} at Day 14 and the DWI-assessed tumour volume change measured at Day 14 (r = −0.39, one-sided p = 0.68) or prior to surgery (r =

Figure 4. Comparison of change at Day 14 of treatment relative to baseline of a) DWI and b) FLT metrics for responders and non-responders with respect to measures of repeatability (dashed lines). Responders are shown with black filled markers and the y-axis for ΔADC_{early} has been broken to include a patient with high ΔADC_{IQR}. The increase in ADC_{75} and ADC_{IQR} and decrease in FLT-TLP was outside the limits of measurement variability in responders.

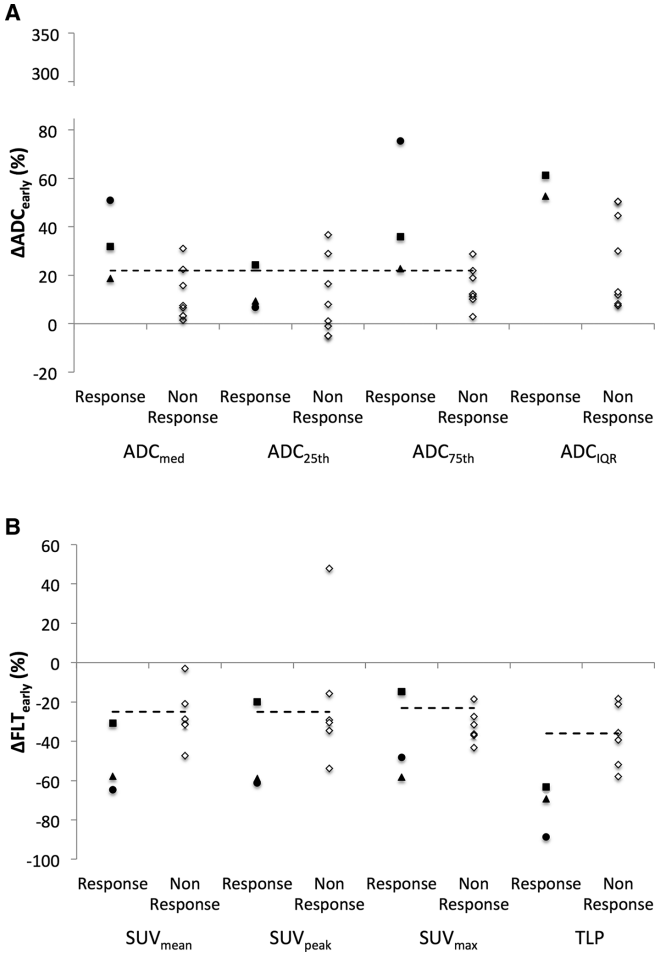
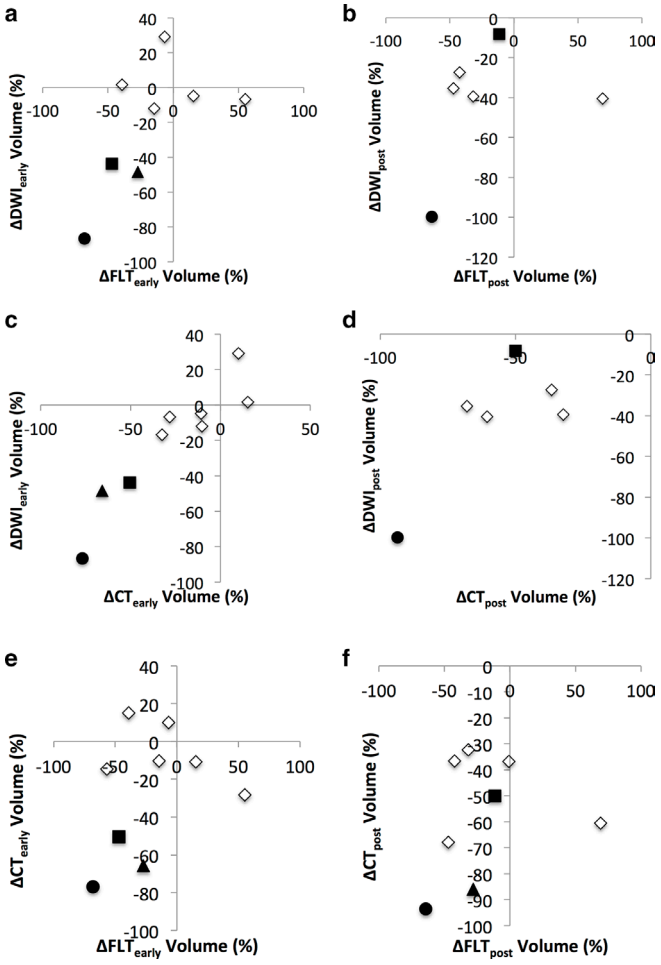


Figure 5. A comparison of changes at Day 14 (top row) and pre-surgery (bottom row) from baseline between tumour volumes derived from DWI and FLT (a, b), DWI and CT (c, d) and FLT and CT (e, f). Responders are represented by black filled markers (*symbols: square=patient 11, triangle=patient 13, circle = patient 14*).



−0.14, one-sided $p = 0.93$). There was also no significant correlation between the ΔSUV_{mean} at Day 14 and the FLT tumour volume change measured at Day 14 ($r = -0.08$, one-sided $p = 0.95$) or prior to surgery ($r = 0.25$, one-sided $p = 0.75$).

CORRELATION OF IMAGING TECHNIQUES

There was no meaningful correlation between ADC_{median} and SUV_{mean} at baseline ($r = -0.27$, one-sided $p = 0.78$), Day 14 ($r = -0.06$, one-sided $p = 0.88$) or pre-surgery ($r = -0.49$, one-sided $p = 0.55$). There was also no significant correlation between ΔADC_{median} and ΔSUV_{mean} ($r = -0.24$, one-sided $p = 0.44$).

DISCUSSION

This pilot study showed no significant association between changes in imaging biomarkers (ADC and FLT-SUV) and pathological measures of response.. Despite the small numbers, this is a robust finding because even the confidence intervals did not reach the correlations originally hypothesised. Therefore, it is unlikely that these imaging biomarkers relate to the pathological features as quantified here. It is also unlikely that we would have seen an association between the imaging parameters and histopathology had the planned sample size been achieved. The original study design did not include a futility analysis, but had a futility analysis been included, we could not have ethically justified continued recruitment in light of the poor response to neoadjuvant chemotherapy in these patients. Our primary pathological measure was the percentage of viable residual tumour, which may not have been the best histopathological endpoint, as pathological evaluation was semi-quantitative. Tumour cellularity derived from digital pathology analyses is a more robust measure that could be considered for future validation of imaging biomarkers on histopathology. Secondary analysis found a positive association between pre-surgical FLT-SUV and Ki-67 in our data, however the literature on the relationship between these two parameters is ambiguous. Positive correlations have been found between the two in gliomas,[28] breast cancer[29] and a mixture of lung nodules,[15] but a number of studies in a variety of tumour types have found no relationship.[30–32]

Our exploratory analysis suggests that changes in imaging biomarkers (measured in a multicentre, multivendor setting against established reproducibility criteria) occur before unidimensional tumour size changes of >30%. ADC generally increased over time, which is linked to an increase in necrotic and apoptotic cell death induced by treatment.[33] Consistent with previous studies, patients who responded to treatment in our sample generally had greater ADC_{median} increases at Day 14 relative to baseline, but statistical analysis was not relevant given the low sample size and small numbers of responders.[34,35] In other NSCLC data, early (after 1 cycle of chemotherapy) increases in ADC have been associated with increased progression-free and overall survival[34] and increased tumour volume reduction.[35] In our data, responders also had larger early increases in ADC_{75th} than non-responders, which could reflect an increase in necrotic domains following treatment. Although relatively unexplored in NSCLC, changes in ADC histogram parameters have been associated with improved response in other tumour types.[36–38]

FLT-SUV generally decreased by Day 14 but changed inconsistently thereafter. Responding patients had amongst the highest FLT-SUV values at baseline, which is expected as chemotherapy targets proliferative cells. They also had greater early decreases in SUV_{mean} and TLP; again the small numbers precluded statistical analysis. A previous study ($n = 9$) assessing NSCLC treatment response using FLT-PET[19] found that FLT parameters did not distinguish between responders and non-responders after 1 cycle of chemotherapy, so that the use of FLT-PET to indicate response remains debatable.

The use of test-retest metrics to establish measurement variability is critical, particularly in a multicentre setting where multiple scanner platforms add variability. Repeatability, ("closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement"), can be represented by limits-of-agreement interval.[39] Our previous test-re-test data from multiple centres including those in this study in larger series for ADC[24] and FLT,[25] gave repeatabilities of 22 and 25% respectively. Therefore, we are

confident that the changes measured here exceed measurement variability. For some outcomes (such as ADC_{75th}), changes were typically observed more often for patients classified as responders. Importantly, changes in imaging biomarkers preceded volumetric changes. If future research confirmed a strong association between (absence of) early change in the biomarker that remained within measurement variability, then these biomarkers could be used for early detection of (in)effective treatment.

NSCLC responded poorly to platinum-based neoadjuvant chemotherapy (four patients became inoperable). Unfortunately, response assessment based on size is confounded by peritumoral atelectasis and inflammation as these regions are often indistinguishable from residual tumour on CT. This is more problematic when segmenting volumes: CT volumes increased with treatment in two of the three patients classified as responders, likely caused by the inclusion of inflammatory tissue.

The main limitations of this study were firstly, the stringent inclusion criteria that required patients to be operable at the outset. The use of neoadjuvant chemotherapy in these cases represented a departure from standard-of-care and severely limited recruitment. Recruitment also was severely affected by 3 cases of disease progression on chemotherapy. Secondly, although the preclinical literature indicates that ADC is a biomarker of necrosis/apoptosis, in a clinical setting histopathological analysis is influenced by technical aspects and observer interpretation, so that the data is less reliable.[40] Correlation of imaging biomarkers with histology is difficult at a whole lesion level because in large tumour specimens as here, selected sections may not be representative of the entire lesion. Image analysis of whole digitised pathology specimens would address these issues in future. Thirdly, progressive disease due to new metastases not increase in primary tumour size, confounded response classification in two cases. Finally, two patients withdrew consent for MRI either during or following MRI examination. The tolerability of multiple imaging examinations in patient groups with poor performance status and compromised respiratory function is an important consideration when planning future studies.

In conclusion, this study adds to the body of evidence documenting longitudinal changes in imaging biomarkers and indicates their lack of correlation with traditional histological markers of response and non-response. Changes in ADC and FLT following neoadjuvant chemotherapy in NSCLC occur as early as 14 days after initiating treatment and exceed measurement variability in responders. However, the utility of early changes in imaging biomarkers as well as the baseline biomarker levels in predicting (non-) response as defined by clinical outcome or tumour size remains to be established.

ACKNOWLEDGEMENTS

Trial support was provided by a large number of personnel at each site, we are grateful to them all: **EORTC headquarters:** Axelle Nzokirantevye (data manager), Sandra Collette, Saskia Litiere (statisticians). Joana Brilhante (project coordinator). **Humanitas Research Hospital**: Dr Katia Marzo (radiopharmacist), Dr Luca Toschi (medical oncologist), Dr Edoardo Bottoni (thoracic surgeon), Dr Cristiana Bonifacio (radiologist), Dr Egesta Lopci (nuclear medicine physician), Dr Daoud Rahal (pathologist). **The Royal Marsden Hospital, London**: Dr Sanjay Popat (medical oncologist), Veronica Morgan (radiographer), Ana Ribero (nuclear medicine technician) **St Georges Hospital, London**: Dr Brendan Tinwell (pathologist), Ms Carol Tan (Thoracic surgeon) **VUMC Amsterdam**: Prof Otto Hoekstra (Nuclear Medicine physician)

CONFLICT OF INTEREST

John Waterton has received compensation from Bioxydyn Ltd, a for-profit company engaged in the development and provision of imaging biomarker services. All other authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

INFORMED CONSENT

Written informed consent was obtained from all subjects (patients) in this study.

REFERENCES

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013; **49**: 1374–403. doi: https://doi.org/10.1016/j.ejca.2012.12.027
2. Non-Small cell lung cancer survival rates. *by Stage* 2017;Available from.
3. Liu SV, Melstrom L, Yao K, Russell CA, Sener SF. Neoadjuvant therapy for breast cancer. *J Surg Oncol* 2010; **101**: 283–91. doi: https://doi.org/10.1002/jso.21446
4. Julien LA, Thorson AG. Current neoadjuvant strategies in rectal cancer. *J Surg Oncol* 2010; **101**: 321–6. doi: https://doi.org/10.1002/jso.21480
5. Balmanoukian A, Ettinger DS. Managing the patient with borderline resectable lung cancer. *Oncology* 2010; **24**: 234–41.
6. Pless M, Stupp R, Ris H-B, Stahel RA, Weder W, Thierstein S, et al. Induction chemoradiation in stage IIIA/N2 non-small-cell lung cancer: a phase 3 randomised trial. *The Lancet* 2015; **386**: 1049–56. doi: https://doi.org/10.1016/S0140-6736(15)60294-X
7. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 2017; **28**(suppl_4): iv1–21. doi: https://doi.org/10.1093/annonc/mdx222
8. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1. *Eur J Cancer* 2009; **45**: 228–47. doi: https://doi.org/10.1016/j.ejca.2008.10.026
9. Nishino M, Dahlberg SE, Cardarella S, et al. Volumetric tumor growth in advanced non-small cell lung cancer patients with EGFR mutations during EGFR-tyrosine kinase inhibitor therapy: developing criteria

to continue therapy beyond RECIST progression. *Cancer* 2013; **119**: 3761–8.
10. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clinical Cancer Research* 2009; **15**: 7412–20. doi: https://doi.org/10.1158/1078-0432.CCR-09-1624
11. Thörmer G, Otto J, Horn L-C, Garnov N, Do M, Franz T, et al. Non-invasive estimation of prostate cancer aggressiveness using diffusion-weighted MRI and 3D proton MR spectroscopy at 3.0 T. *Acta radiol* 2015; **56**: 121–8. doi: https://doi.org/10.1177/0284185113520311
12. XR L, Cheng LQ, Liu M, et al. DW-MRI ADC values can predict treatment response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *Med Oncol* 2012; **29**: 425–31.
13. Okuma T, Matsuoka T, Yamamoto A, Hamamoto S, Nakamura K, Inoue Y. Assessment of early treatment response after CT-guided radiofrequency ablation of unresectable lung tumours by diffusion-weighted MRI: a pilot study. *Br J Radiol* 2009; **82**: 989–94. doi: https://doi.org/10.1259/bjr/13217618
14. Rasey JS, Grierson JR, Wiens LW, Kolb PD, Schwartz JL. Validation of FLT uptake as a measure of thymidine kinase-1 activity in A549 carcinoma cells. *J Nucl Med* 2002; **43**: 1210–7.
15. Buck AK, Halter G, Schirrmeister H, et al. Imaging proliferation in lung tumors with PET: 18F-FLT versus 18F-FDG. *J Nucl Med* 2003; **44**: 1426–31.
16. Ullrich RT, Zander T, Neumaier B, Koker M, Shimamura T, Waerzeggers Y, et al. Early Detection of Erlotinib Treatment Response in NSCLC by 3′-Deoxy-3′-[18F]-Fluoro-L-Thymidine ([18F]FLT) Positron Emission Tomography (PET. *PLoS One* 2008; **3**: e3908. doi: https://doi.org/10.1371/journal.pone.0003908
17. Kenny L, Coombes RC, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3′-deoxy-3′-[18F]fluorothymidine positron emission tomography. *Eur J Nucl Med Mol Imaging* 2007; **34**: 1339–47. doi: https://doi.org/10.1007/s00259-007-0379-4
18. Tsuyoshi H, Morishita F, Orisaka M, Okazawa H, Yoshida Y. 18F-Fluorothymidine PET is a potential predictive imaging biomarker of the response to gemcitabine-based chemotherapeutic treatment for recurrent ovarian cancer. *Clin Nucl Med* 2013; **38**: 560–3. doi: https://doi.org/10.1097/RLU.0b013e318292ee9c
19. Crandall JP, Tahari AK, Juergens RA, Brahmer JR, Rudin CM, Esposito G, et al. A comparison of FLT to FDG PET/CT in the early assessment of chemotherapy response in stages IB–IIIA resectable NSCLC. *EJNMMI Res* 2017; **7**: 8. doi: https://doi.org/10.1186/s13550-017-0258-3
20. Soloviev D, Lewis D, Honess D, Aboagye E. 18F]FLT: An imaging biomarker of tumour proliferation for assessment of tumour response to treatment. *Eur J Cancer* 2012; **48**: 416–24. doi: https://doi.org/10.1016/j.ejca.2011.11.035
21. Sinkus R, Van Beers BE, Vilgrain V, DeSouza N, Waterton JC, Consortium Q-C. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. *Eur J Cancer* 2012; **48**: 425–31. doi: https://doi.org/10.1016/j.ejca.2011.11.034
22. deSouza NM, Winfield JM, Waterton JC, Weller A, Papoutsaki M-V, Doran SJ, et al. Implementing diffusion-weighted MRI for body imaging in prospective multicentre trials: current considerations and future perspectives. *Eur Radiol* 2018; **28**: 1118–31. doi: https://doi.org/10.1007/s00330-017-4972-z
23. Hristova I, Boellaard R, Galette P, Shankar LK, Liu Y, Stroobants S, et al. Guidelines for quality control of PET/CT scans in a multicenter clinical study. *EJNMMI Physics* 2017; **4**: 23. doi: https://doi.org/10.1186/s40658-017-0190-7
24. Weller A, Papoutsaki MV, Waterton JC, Chiti A, Stroobants S, Kuijer J, et al. Diffusion-Weighted (DW) MRI in lung cancers: ADC test-retest repeatability. *Eur Radiol* 2017; **27**: 4552–62. doi: https://doi.org/10.1007/s00330-017-4828-6
25. Kramer GM, Liu Y, de Langen AJ, Jansma EP, Trigonis I, Asselin M-C, et al. Repeatability of quantitative 18F-FLT uptake measurements in solid tumors: an individual patient data multi-center meta-analysis. *Eur J Nucl Med Mol Imaging* 2018; **45**: 951–61. doi: https://doi.org/10.1007/s00259-017-3923-x
26. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. Fdg PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015; **42**: 328–54. doi: https://doi.org/10.1007/s00259-014-2961-x
27. Tinnemans MMFJ, Lenders M-HJH, Ramaekers FCS, Schutte B, ten Velde GPM, Wagenaar SS, et al. Evaluation of proliferation parameters in in vivo bromodeoxyuridine labelled lung cancers. *Virchows Archiv* 1995; **427**: 295–301. doi: https://doi.org/10.1007/BF00203398
28. Yamamoto Y, Ono Y, Aga F, Kawai N, Kudomi N, Nishiyama Y. Correlation of 18F-FLT uptake with tumor grade and Ki-67 immunohistochemistry in patients with newly diagnosed and recurrent gliomas. *Journal of Nuclear Medicine* 2012; **53**: 1911–5. doi: https://doi.org/10.2967/jnumed.112.104729
29. Crippa F, Agresti R, Sandri M, Mariani G, Padovano B, Alessi A, et al. 18F-FLT PET/CT as an imaging tool for early prediction of pathological response in patients with locally advanced breast cancer treated with neoadjuvant chemotherapy: a pilot study. *Eur J Nucl Med Mol Imaging* 2015; **42**: 818–30. doi: https://doi.org/10.1007/s00259-015-2995-8
30. Johnbeck CB, Knigge U, Langer SW, Loft A, Berthelsen AK, Federspiel B, et al. Prognostic value of 18F-FLT PET in patients with neuroendocrine neoplasms: a prospective head-to-head comparison with 18F-FDG PET and Ki-67 in 100 patients. *Journal of Nuclear Medicine* 2016; **57**: 1851–7. doi: https://doi.org/10.2967/jnumed.116.174714
31. Honndorf VS, Schmidt H, Wehrl HF, Wiehr S, Ehrlichmann W, Quintanilla-Martinez L, et al. Quantitative Correlation at the Molecular Level of Tumor Response to Docetaxel by Multimodal Diffusion-Weighted Magnetic Resonance Imaging and [18 F]FDG/[18 F]FLT Positron Emission Tomography. *Mol Imaging* 2015; **14**: 7290.2014.00045. doi: https://doi.org/10.2310/7290.2014.00045
32. Schelhaas S, Wachsmuth L, Viel T, Honess DJ, Heinzmann K, Smith D-M, et al. Variability of proliferation and diffusion in different lung cancer models as measured by 3'-Deoxy-3'-18F-Fluorothymidine PET and diffusion-weighted MR imaging. *Journal of Nuclear Medicine* 2014; **55**: 983–8. doi: https://doi.org/10.2967/jnumed.113.133348
33. Papaevangelou E, Almeida GS, Jamin Y, Robinson SP, deSouza NM. Diffusion-Weighted MRI for imaging cell death after cytotoxic or apoptosis-inducing therapy. *Br J Cancer* 2015; **112**: 1471–9. doi: https://doi.org/10.1038/bjc.2015.134
34. Tsuchida T, Morikawa M, Demura Y, Umeda Y, Okazawa H, Kimura H. Imaging the early response to chemotherapy in advanced lung cancer with diffusion-weighted magnetic resonance imaging compared to fluorine-18 fluorodeoxyglucose positron emission tomography and computed tomography. *J. Magn. Reson. Imaging* 2013; **38**: 80–8. doi: https://doi.org/10.1002/jmri.23959

35. Yabuuchi H, Hatakenaka M, Takayama K, Matsuo Y, Sunami S, Kamitani T, et al. Non–Small cell lung cancer: detection of early response to chemotherapy by using contrast-enhanced dynamic and diffusion-weighted MR imaging. *Radiology* 2011; **261**: 598–604. doi: https://doi.org/10.1148/radiol.11101503
36. Choi MH, Oh SN, Rha SE, Choi J-I, Lee SH, Jang HS, et al. Diffusion-Weighted imaging: apparent diffusion coefficient histogram analysis for detecting pathologic complete response to chemoradiotherapy in locally advanced rectal cancer. *J. Magn. Reson. Imaging* 2016; **44**: 212–20. doi: https://doi.org/10.1002/jmri.25117
37. Kim Y, Kim SH, Lee HW, Song BJ, Kang BJ, Lee A, et al. Intravoxel incoherent motion diffusion-weighted MRI for predicting response to neoadjuvant chemotherapy in breast cancer. *Magn Reson Imaging* 2018; **48**: 27–33. doi: https://doi.org/10.1016/j.mri.2017.12.018
38. Meng J, Zhu L, Zhu L, Ge Y, He J, Zhou Z, et al. Histogram analysis of apparent diffusion coefficient for monitoring early response in patients with advanced cervical cancers undergoing concurrent chemo-radiotherapy. *Acta radiol* 2017; **58**: 1400–8. doi: https://doi.org/10.1177/0284185117694509
39. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology standards for quantitative imaging biomarkers. *Radiology* 2015; **277**: 813–25. doi: https://doi.org/10.1148/radiol.2015142202
40. Potts SJ, Young GD, Voelker FA. The role and impact of quantitative discovery pathology. *Drug Discov Today* 2010; **15**(21-22): 943–50. doi: https://doi.org/10.1016/j.drudis.2010.09.001