Diffusion-weighted (DW) MRI in Lung Cancers: ADC Test-retest Repeatability

ABSTRACT:

Purpose: To determine the test-retest repeatability of Apparent Diffusion Coefficient (ADC) measurements across institutions and MRI vendors, plus investigate the effect of post-processing methodology on measurement precision.

Methods: Thirty malignant lung lesions >2cm in size (23 patients) were scanned on two occasions, using echo-planar-Diffusion-Weighted (DW)-MRI to derive whole-tumour ADC (b=100, 500 and 800smm^{-2}). Scanning was performed at 4 institutions (3 MRI vendors). Whole-tumour volumes-of-interest were copied from first visit onto second visit images and from one post-processing platform to an open-source platform, to assess ADC repeatability and cross-platform reproducibility.

Results: Whole-tumour ADC values ranged from 0.66-1.94x10^{-3}mm^{2}s^{-1} (mean=1.14). Within-patient coefficient-of-variation (*wCV*) was 7.1% (95% CI 5.7–9.6%), limits-of-agreement (LoA) -18.0 to 21.9%. Lesions >3cm had improved repeatability: *wCV* 3.9% (95% CI 2.9–5.9%); and LoA -10.2 to 11.4%. Variability for lesions <3cm was 2.46 times higher. ADC reproducibility across different post-processing platforms was excellent: Pearson's R^2 = 0.99; CoV 2.8% (95% CI 2.3-3.4%); and LoA -7.4 to 8.0%).

Conclusion: A free-breathing DW-MRI protocol for imaging malignant lung tumours achieved satisfactory within-patient repeatability and was robust to changes in post-processing software, justifying its use in multi-centre trials. For response evaluation in individual patients, a change in ADC >21.9% will reflect treatment-related change.

Key words: Lung Cancer. Apparent Diffusion Coefficient (ADC). Repeatability. Reproducibility. Diffusion-weighted MRI

KEY POINTS:

1. In lung cancer, free-breathing DWI-MRI produces acceptable images with evaluable ADC measurement.
2. ADC repeatability coefficient-of-variation is 7.1% for lung tumours >2cm.
3. ADC repeatability coefficient-of-variation is 3.9% for lung tumours >3cm.
4. ADC measurement precision is unaffected by the post-processing software used.
5. In multicentre trials, 22% increase in ADC indicates positive treatment response.

ABBREVIATIONS:

DW-MRI = Diffusion-weighted magnetic resonance imaging

ADC = Apparent diffusion coefficient

NSCLC = Non small-cell lung cancer

SCLC = Small-cell lung cancer

EORTC = European Organization for Research and Treatment of Cancer

CRUK = Cancer Research UK

UK = United Kingdom

GE = General Electric

STIR = Short-tau inversion recovery

NSA = Number of signal averages

LoA = Limits of Agreement,

wCV = Within subject Coefficient of Variation

ICC = Intra-class correlation

CCC = Concordance correlation coefficient

IDL = Interactive digital language

DICOM = Digital Imaging and Communications in Medicine

EPSRC = *Engineering and Physical Sciences Research Council*

NIHR = National Institute for Health Research (UK)

NHS = National Health Service

ICR = Institute of Cancer Research (UK)

RMH = Royal Marsden Hospital (UK)

MRC = Medical Research Council (UK)

INTRODUCTION

Diffusion-weighted MRI derived Apparent Diffusion Coefficient (ADC) is emerging as a potentially valuable imaging biomarker for quantifying treatment response in a number of tumour types, including in lung cancers. It is being applied as an end point in an increasing number of clinical trials both outside (1-3) and within the lung (Table 1, (4, 5)). In order to utilize change in ADC as a response biomarker, uncertainty in its quantitation must be lower than the change following treatment, which in lung tumours ranges between 16-90% (Table 1). Therefore in order to detect meaningful change with treatment, it is desirable that the uncertainty of the ADC measurement is <16%. This measurement uncertainty must include calculation of marker precision and bias estimation. The latter is carried out in test objects (6-8), while the former (defined as '*the closeness of agreement between measured quantity values obtained by replicate'*) is obtained through test-retest *repeatability* measurements under specified conditions. There are no reports of ADC measurement repeatability in lung cancers in a multi-centre setting, although inter and intra-observer coefficients of variation estimated from repeated measurements on the same data sets (reproducibility) have been estimated at 3.7-11.4% (depending on lesion size and location in the chest) (9, 10).

Methods of deriving ADC suffer from inconsistent methodology across different centres, both in data acquisition and analysis; a wide variety of lesion segmentation methodologies and software packages have been presented for ADC quantitation (Table 1 (11-20)). However, even when acquisition and analysis methods are standardised, uncertainty resulting from different scanner platforms and different post-processing algorithms between institutions, which is inherent in multicentre trials, is unknown. The EORTC/CRUK imaging biomarker validation roadmap stresses the huge importance of multicentre multivendor repeatability/reproducibility studies to ensure that imaging biomarkers can translate beyond single-centre use (21). The purpose of this study therefore was to determine the *repeatability* of ADC measurements acquired on a test-re-test basis using a common and generalizable free breathing DW-MRI protocol, across four university hospitals and 3 different MRI vendor platforms. It was performed under EORTC Innovative Medicines Initiative (IMI) QuIC-ConCePT project (Quantitative Imaging in Cancer: Connecting Cellular Processes with Therapy), for which the variation in ADC measurement precision was investigated as a function of lesion size, post-processing methodology and ADC summary statistic employed. The ultimate aim was to validate the use of ADC for treatment response assessment in lung cancer in a multi-centre setting (www.imi.europa.eu).

MATERIALS AND METHODS

Patients

This prospective, multicentre study was performed following local Ethics Review Board approval in Italy, the Netherlands and the UK. Across four university hospitals, patients with at least one lung

tumour > 2 cm in size identified on CT and without contra-indication to MRI were invited to participate. Written informed consent was obtained for the 27 patients enrolled (15 men, 12 women, age range 41 – 86 years). Between May 2014 and September 2015, 25 of the 27 patients were scanned on two occasions >1 hour and <1 week apart (median interval 4.29 hours). In 2 patients, the repeatability scans were inconsistent with the imaging protocol for the study and their data was excluded from the analysis (Table 2). For these patients, the DW-imaging had been performed with insufficient signal averages (NSA < 4), providing lower signal-to-noise ratio than was obtained for the remaining patients. The remaining 23 patients underwent test-retest repeatability imaging according to the study protocol. Of these 23 patients, 15 patients had primary lung cancer (14 NSCLC and 1 small cell lung cancer (SCLC)), 8 had metastatic lung lesions (3 from colorectal carcinoma; 1 from renal cell cancer; and 1 uterine leiomyosarcoma) and 3 had an undocumented primary site. 6 of the 23 patients were treatment naïve (5 of whom had lung cancer and 1 metastatic renal cell cancer). The other 17 had either received treatment >1 week prior to enrolment (chemotherapy or radiotherapy to lungs in 11 patients), or treatment status was not documented by the scanning site (6 patients). For the 11 in whom prior treatment had been documented, 6 received chemotherapy alone, 3 received a combination of chemotherapy and radiotherapy and 2 had received radiotherapy alone. The mean interval between end of treatment and baseline scan was 63 weeks, so that no on-going post-treatment effects were present. Analysis was possible for 30 lung lesions >2cm in size.

Quality Assurance

Quality assurance was carried out prior to scanning and then every 3 months, using a previously described temperature controlled sucrose phantom, to confirm ADC stability on the 1.5 T MR scanners at the four sites (22).

Image Data Acquisition

All imaging experiments were performed during free breathing, using phased-array body coils (2 anterior and 2 posterior elements) on the following platforms: GE Optima 1.5T (site A); Philips Achieva 1.5T (sites B and G); Siemens Avanto 1.5T (site E). DW-MR imaging comprised twice refocused spin-echo sequences with single-shot echo planar readout, using a short-tau inversion recovery (STIR) fat suppression technique, over a large field of view. 5mm transverse slices with no slice gap were obtained through the whole lung, using 30-60 slice scanning volumes per series of DW images. Each series of DW images was either performed four times, or once with NSA=4 (GE, and Philips). Images were acquired at three b-values (100, 500 and 800 s/mm^2). DW sequence parameters included b values greater than 100s/mm^2, in order to reduce the effects of perfusion on the ADC estimate (Table 2).

Anatomical images of the whole chest were also obtained, using axial T1-weighted (T1-W) turbo spin-echo sequence and a three-dimensional (3D) T2-W turbo spin echo sequence with variable flip angle.

Image Data Analysis

Bright regions on the high-b value images that are iso- or hypo-intense to spinal cord on the ADC maps are features that have been shown to differentiate tumour from pleural fluid or pulmonary collapse and were used to delineate tumour (23) (Figure 1). With reference to the high b-value images and ADC maps (so as to differentiate tumour from adjacent atelectasis), tumour size (maximum lesion dimension) was evaluated on the anatomical T1W imaging using an electronic calliper (performed in OsiriX, Pixmeo, Geneva, Switzerland).

Segmentation was performed on every slice on which tumor was present, to encompass the whole tumour cross section, by an experienced radiologist (AW) using both a region growing technique (where a user defined seed is grown and checked by the operator for registration, to include all nearest neighbour pixels that lie within the mean +/- a specified number of standard deviations of the original seed mean value - ADEPT, Institute of Cancer Research, UK) and a freehand drawing technique (OsiriX). Freehand segmentation in Osirix was performed with reference to the region growing boundaries defined in ADEPT, in order to define anatomically co-registered tumour regions on the two software packages (Figure 2).

Whole-tumour segmentation was performed on the computed (or 'virtual') high b (=800 smm^{-2}) value images rather than the acquired b=800smm^{-2} images as the former provide higher SNR, in combination with good image quality and background suppression, while at the same time ensuring exact anatomical registration with the ADC images (24). Segmentation was performed for images obtained at the first patient visit (DWI-1), for all lesions that were: (a) >2cm in size and; (b) present on at least 3 consecutive slices. These regions were copied slice by slice onto anatomically co-registered tumour regions on images obtained at the second patient visit (DWI-2), so as to generate anatomically matched test-retest ADC measurements.

ADC and computed DW-MR images were generated by applying a mono-exponential decay model to signal decay with increasing b-value (Levenberg- Marquardt algorithm) in both software packages (24). Measurements generated in OsiriX were used to calculate multi-centre, cross vendor, within-patient ADC test-retest *repeatability*.

For those lesions in which image analysis was possible on the 2 different post-processing platforms (25 of 30 lesions), the reproducibility of the ADC measurement using 2 different software packages was evaluated.

Statistical Analysis

Statistical analysis was performed in Graph-pad Prism Version 6 (GraphPad Software Inc. CA, USA). Data used for comparison was tested for normality (d'Angostino Pearson) and log-transformed if non-normal. Normally distributed data were compared using a Student's t-test.

Test-retest repeatability and measurement precision for whole tumour median ADC (ADC_{med}) was assessed with Bland-Altman plots, as well as by calculating Limits of Agreement (LoA), within subject Coefficient of Variation (*wCV*), and intra-class correlation (ICC). These parameters were calculated for ADC values measured for all lesions and separately for lesions 2- 3cm and lesions > 3cm. Differences in test-retest ADC_{med} measurement variability between scanning institutions (sites A, B, E and G) were assessed using one-way ANOVA. Variability for this purpose was defined by the difference in ADC value per lesion on test-retest scanning [(ADC-1) – (ADC-2)]. Differences in ADC measurement variability based on lesion size (<3cm versus >3cm) were assessed using the F-test (25). For the ADC values generated on the two post-processing platforms, differences between the absolute ADC values and test-retest ADC value variability were assessed for significance using paired Student's t-test. For this analysis, difference in ADC variability using each post-processing platform was calculated by considering: $[(ADC-1) – (ADC-2)]_{ADEPT}$; compared with $[(ADC-1) – (ADC-2)]_{OsiriX}$. Pearson's correlation coefficient was also calculated for ADC estimates derived using the two post processing platforms and agreement of ADC values between the platforms was assessed using Bland-Altman analysis and the concordance correlation coefficient (CCC) (8).

The influence of ADC summary statistic on repeatability used was assessed by calculating wCV, LoA, ICC and measurement variability [(ADC-1) – (ADC-2)] for whole tumour mean ADC values (ADC_{mean}), as had previously been performed for ADC_{med}. ADC_{mean} test-retest variability was compared with corresponding ADC_{med} variability using the paired Student's t-test. In addition, Pearson's R^2, concordance correlation coefficient and coefficient of variation (CoV) between paired ADC_{mean} and ADC_{med} values were calculated. The absolute ADC values (ADC_{med} and ADC_{mean}) were also compared for difference using the Student's t-test with (Holm-Bonferroni corrected) level of significance set as p = 0.0125.

RESULTS

ADC measurements of the test object from all scanners at all time-points fell within the expected range, indicating that quality assurance specifications for the study were met (22).

Evaluable lesions ranged in size from 21 to 94mm (Table 3). Median ADC (ADC_{med}) values for whole tumour were in the range 0.66 to 1.94 $x10^{-3}mm^2/s$ (mean = 1.14 $x10^{-3}mm^2/s$, sd = 0.33 $x10^{-3}mm^2/s$). Equivalent mean tumour ADC (ADC_{mean}) values were in the range 0.64 to 1.97 x $10^{-3}mm^2/s$ (mean =

1.16 x10^{-3}mm^2/s, sd 0.31 x10^{-3}mm^2/s). The highest value of ADC was recorded for patient 4, in whom artefact from a subcutaneous metallic foreign body distal to but at the same level as the treatment naive right upper lobe NSCLC is likely to have influenced diffusion weighted signal decay.

Averaged ADC$_{med}$ values and the *repeatability* of whole lesion analysis from the two imaging time-points are summarized in Table 3. Within patient ADC$_{med}$ coefficient of variation (*wCV*) for all lesions was: 7.1% (95% CI 5.7 – 9.6%); limits of agreement (LoA) were -18.0 to 21.9%; and ICC was 0.94 (95% CI 0.88 to 0.97). The equivalent repeatability results using ADC$_{mean}$ were very similar: *wCV* for all lesions was 7.0% (95% CI 5.6 to 9.3%); LoA -17.5 to 21.3%; ICC 0.95 (95% CI 0.89 to 0.97). In line with this, there was no significant difference in ADC measurement variability for ADC$_{mean}$ compared with ADC$_{med}$ (p=0.41). A strong correlation was observed between ADC$_{med}$ with ADC$_{mean}$, where: Pearson's R^2 = 0.98; CoV 3.0 % (95% CI 2.6 – 3.7%); LoA -6.3 to 10.9%; and concordance correlation coefficient (CCC) 0.99 (95% CI 0.980 to 0.992). Despite this, absolute ADC$_{med}$ values were significantly different from ADC$_{mean}$ (p=0.007), although the magnitude of this difference was small (mean ADC$_{med}$ = 1.14, range = 0.66 to 1.94 x10^{-3}mm^2/s, whereas mean ADC$_{mean}$ =1.16, range = 0.64 to 1.97 x 10^{-3}mm^2/s). Nonetheless, these results reflect a systematic shift toward higher values for ADC$_{mean}$ compared with ADC$_{med}$.

Considering the effect of lesion size, ADC$_{med}$ repeatability for lesions > 3cm (n=16) is summarised by: *wCV* of 3.9% (95% CI 2.9 – 5.9%); LoA -10.2 to 11.4%; and ICC 0.98 (95% CI 0.95 to 0.99). In comparison, ADC$_{med}$ measurement variability for lesions <3 cm (n=14) was c. 2.5 times higher, with: *wCV* of 9.6% (95% CI 7.0 – 15.2%); LoA -23.3 to 30.5%; and ICC 0.92 (95% CI 0.77 to 0.97). This difference in ADC measurement variability for lesions >3cm compared with lesions < 3cm reached statistical significance [$F_{(15,13)}$ = 0.13, p = 0.0002]. Bland Altman plots in Figure 3 (a-c) summarise these data. Comparing lesions >3cm with lesions <3cm, no significant difference was observed between these groups in terms of either the interval between scans (Mann-Whitney p=0.24) or prior treatment status (Fischer exact test p = 0.17). From the one-way ANOVA, the scanning institution had no significant effect on test-retest ADC$_{med}$ measurement variability [$F_{(3,26)}$ = 0.87, p = 0.47)], a result that is confirmed by the overlapping 95% confidence intervals for *wCV* (Table 3, Figure 3 (d)). Similarly, there was no significant difference in absolute ADC$_{med}$ values between primary and metastatic lesions (p = 0.58), nor between treatment naïve and previously treated patients (p =0.74).

ADC *reproducibility* using two different post-processing software packages was possible for DW-MRI performed at sites A, B and E. For 5 lesions scanned at site G, due to storing and transfer of the image data in a 'JPEG lossless' format, in which grey-scale bit-depth of the DICOM files is compressed, quantitative analysis was not possible on ADEPT (IDL). For the remaining 25 lesions, agreement (measured on a per-lesion basis) between ADC$_{med}$ values generated on two different post-processing platforms was excellent, with: Pearson's R^2 = 0.99; CoV 2.8% (95% CI 2.3 – 3.4%); LoA

-7.4 to 8.0%; and concordance correlation coefficient (CCC) 0.99 (95% CI 0.989 to 0.996) (Table 4). This is demonstrated graphically in the correlation, Bland Altman and box-plots in Figure 4. In addition, for the two different post-processing platforms, no significant differences were seen in terms of either the absolute ADC_{med} values generated (p = 0.13) or for test-retest ADC_{med} variability (p= 0.73).

DISCUSSION

This study demonstrates a *wCV* of < 10% for ADC (both median and mean values) in malignant lung lesions across multiple institutions, using a whole lung DW-MRI protocol during free breathing and different post-processing software packages. It is the first study to confirm multi-centre within-patient test-retest *repeatability* in malignant lung lesions and indicates that within a clinical trial, a measured ADC change of >22% is an acceptable threshold for indicating response, as it would be above the 95% limits of agreement for test-retest scanning (LoA = -18.0 to 21.9%). This change is a little greater than the change recorded following treatment in some single centre reports in the literature (Table 1). Nevertheless, the similarity of the absolute ADC values between data in these reports and our cohort endorses our ADC repeatability measurement and justifies its use in future multicentre clinical trials (6). However, due to the wide range of *individual* ADC test-retest variability, generalisability of our findings to assess response of *individual* patients in the clinical setting would require justification. Our data nonetheless demonstrates acceptable cohort *wCV*, for the purpose of measuring treatment-related change in a clinical trial. Furthermore, ADC is a very promising biomarker that will allow quantitative interrogation of tumour microstructure and cell membrane integrity (http://qibawiki.rsna.org/index.php), potentially reflecting treatment-induced changes early during therapy, where size based measurements are non-informative because they do not reflect changes in tumour biology (26).

The choice of ADC summary statistic significantly altered absolute ADC values: ADC_{med} was significant different from ADC_{mean}, despite strong correlation between the two values. This is likely due to the bimodal ADC distribution within mixed necrotic/solid tumours. This difference highlights the importance of consistent methodology within and between trials before absolute ADC measurements can be compared, so as to mitigate against risks (27). Repeatability was equivalent for both metrics, indicating that choice of either metric is acceptable. The effect of lesion size on ADC repeatability for lung tumours is in line with prior reports on reproducibility (9). Significantly better repeatability was seen for lesions >3cm than smaller lesions. This reflects the greater effect of respiratory motion on smaller lesions. When respiratory excursion in the z-axis is greater than half of tumour size, volume averaging between normal lung and tumour occurs for all locations within tumour. It is interesting to note that for lesions <3cm in size, half of this dimension was similar to the mean diaphragm excursion expected during quiet respiration, reported in prior studies to lie between 1.4 to 1.7 cm (9, 28). This effect of lesion size is also likely to have accounted for differences in *wCV*

observed between institutions in our study - Site E had greater mean lesion size than site A and tendency to a lower *wCV*, although this latter difference did not reach statistical significance. Use of motion compensation protocols when assessing small lesions may well be warranted in the future in a single centre setting. Any measures employed should take into account the dependence of respiratory motion upon tumour location in the chest (29).

Perfusion related ADC bias was minimised by using $b=100s/mm^2$ as the lowest b-value (30, 31) with the upper b-values dictated by previously observed ADC values in lung tumours ((32, 33), Table 1). The $b=500s/mm^2$ acquisition ensured that ADC values from predominantly mucinous/necrotic tumours (high ADC) were accurately represented, as in these tumours signal at $b=800s/mm^2$ has a significant noise contribution. The satisfactory ADC measurement repeatability in lung tumours has enabled roll-out into a European multicentre trial assessing NSCLC treatment response to neo-adjuvant chemotherapy [EORTC 1217 https://clinicaltrials.gov/show/NCT02273271].

The free breathing protocol used in his study is easily implemented in multicentre trials and both generalizable across centres and suitable for the lung cancer patient group, in whom breath-hold capacity is limited. Ease of implementation was strong factor in devising the protocol and it could be further refined if proposed for single centre use. For example, motion compensation measures could be applied at the expense of scan duration in the single centre setting (9), where image quality may be improved by reducing the effect of respiratory motion. One limitation of our data is that lesions < 2cm were not included in the analysis. Future evaluation of smaller lesions would be best achieved after applying a successful respiratory compensation protocol.

The effect of gradient non-linearity on ADC accuracy and reproducibility has been highlighted as area of concern for clinical trials and poor inter-scanner reproducibility has been cited as reducing the diagnostic value of ADC (34). In light of this, each patient included in this study had examinations performed on the same scanner with identical acquisition parameters at each visit. However, even when using the same scanner, changing tumour position within the B0 field and relative to the DWI scanning volume can distort tumour ADC estimates, due to non-linearity of both the spatial encoding and diffusion encoding gradients (34). Inconsistent patient position is likely to have a negative impact on repeatability, a factor that was minimised in our study by consistent patient and scanning volume positioning by dedicated research technologists.

It is interesting to note that no significant difference was observed between ADC values for treatment naïve compared with patients that had received treatment. Lesion segmentation in our study included necrotic areas of tumour for some patients, potentially leading to bias in the ADC measured, while for patient 4 (treatment naïve), metallic artefact at the same slice position as tumour caused encoding and diffusion gradient distortion. However, the small difference in ADC between the two imaging time-

points for this patient (1.34%) indicates that any artefact induced alteration of ADC did not have an adverse effect on repeatability.

Our data demonstrate the robustness of mono-exponential log linear ADC fitting. Prior reports have shown that post-processing of quantitative MRI parameters can have a profound impact measurement uncertainty, especially with dynamic contrast enhancement (35-37). In our analysis of ADC, the fact that the two software packages did not utilize perfectly matched regions of interest at each location within tumour (because we were unable to export regions of interest from one package to the next), demonstrated the robustness of ADC measurement in the chest. The region growing segmentation methodology is a technique previously shown to produce acceptable intra- and inter-observer reproducibility (9) and our analysis illustrates that cognitive registration of regions matched between software packages on the b=800 smm^{-2} images suffices. To be clinically meaningful, the measurement needs to be repeatable across multiple observers, software platforms and imaging platforms (2, 38). This study adds to existing data by confirming the validity of post-processing on a widely available, open-source DICOM browser such as Osirix. The data from this study provides the first step in demonstrating the viability of ADC for the purpose of treatment response evaluation in lung cancer and justifies its application to future clinical trials.

CONCLUSION

We have demonstrated satisfactory test-retest repeatability and reproducibility of ADC measurements in lung tumours, using an easily implemented free breathing DW-MRI protocol across multiple institutions. These results justify the more widespread interrogation of ADC as a potential biomarker in phase II and III clinical trials, where its role in predicting outcomes following therapy now requires evaluation (39). If proposed for more widespread use, ADC also provides a robust measurement that is not unduly influenced by different post processing software packages, showing very close agreement and satisfactory *reproducibility* between our in house analysis software and open-source DICOM browser based Osirix. Further interrogation of the methodology, including with motion compensation and high resolution single lesion coverage would be essential before applying ADC quantitation to individual patients in the clinical setting.

Figure 1: Comparison of anatomical T2w (4 (a)) and DW-MRI (4 (b) and (c)) images for a left upper lobe tumour
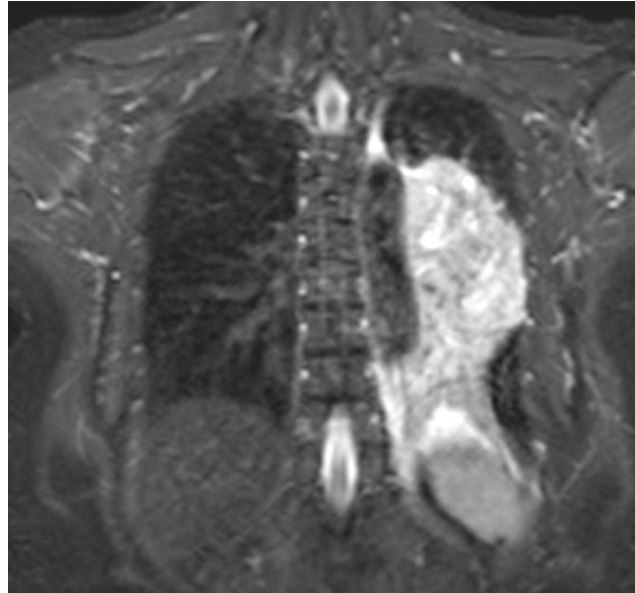

Figure 1 (a): Coronal anatomical T2-SPACE image through a left hilar mass, causing distal left lower lobe collapse. Primary tumour is not differentiated from adjacent collapse.
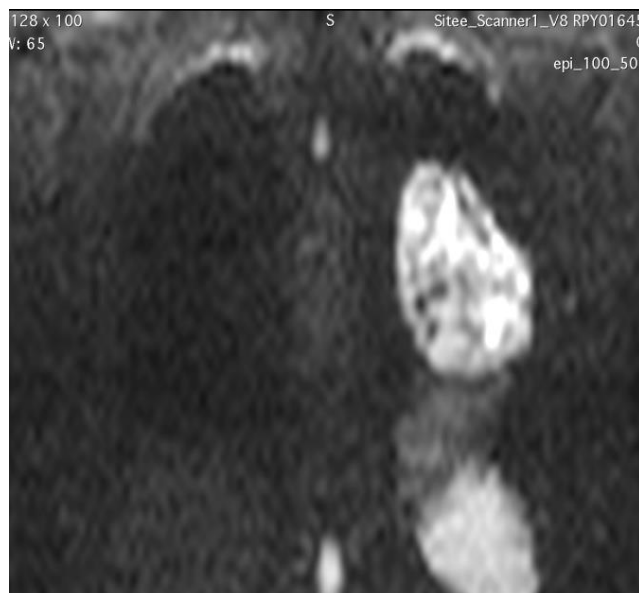

Figure 1 (b): Corresponding b=800 image which, like the ADC map delineates tumour boundaries more clearly than on the anatomic T2W images (4a)
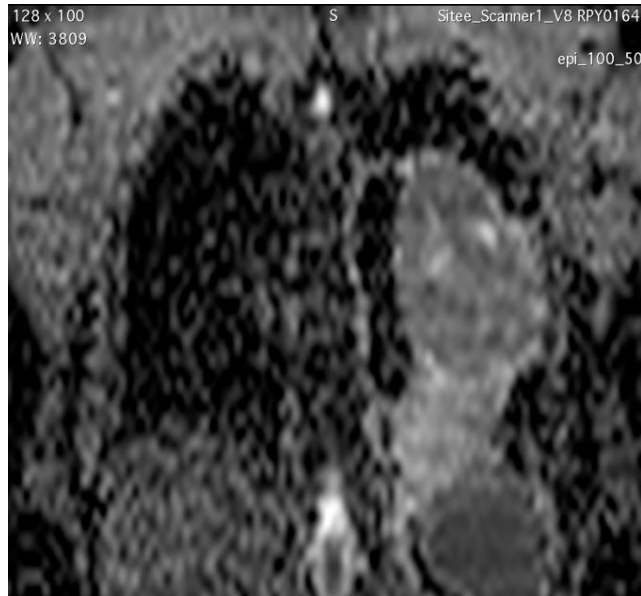
Figure 1 (c): Corresponding ADC on which tumour has higher restricted diffusion compared with distal atelectasis, demonstrating tumour boundaries clearly.
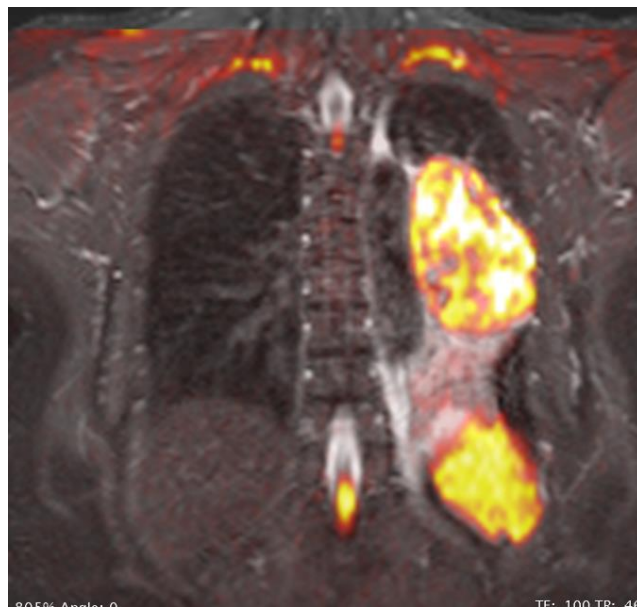

Figure 1 (d): Colour overlay of the b800 DW images onto the Coronal T2W images, acquired in expiration with a respiratory trigger from the dome of the diaphragm. Images are matched for slice location in the z-direction. Co-registration of the free breath b800 images with the end expiratory SPACE images is an interesting observation.

Figure 2: Comparative ROIs generated using axial images in two different post-processing packages:
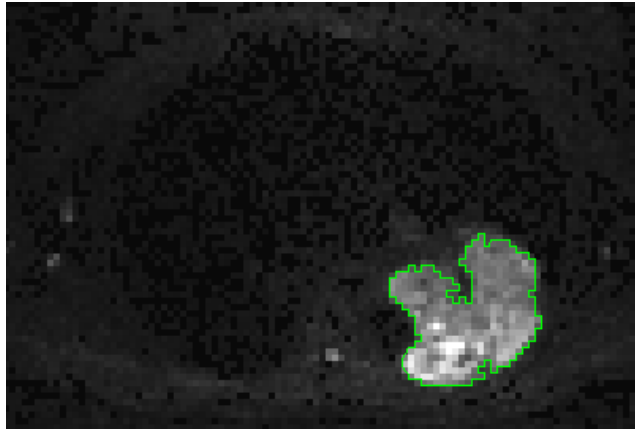

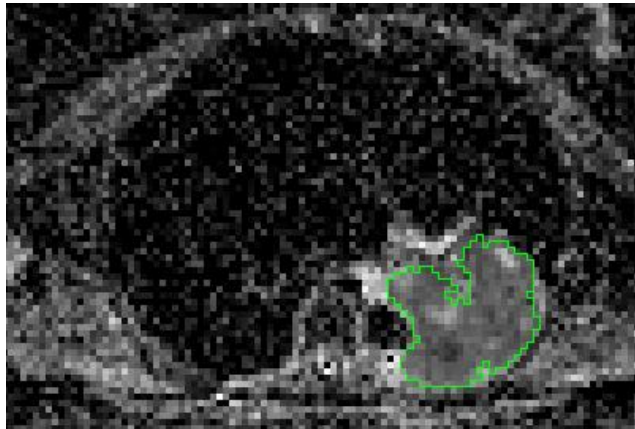**Figure 2 (a):** ROI generated using axial images in ADEPT (b800 image)

Figure 2 (b): ROI generated using axial images in ADEPT (ADC image)

Figure 2 (c): ROI generated using axial images in Osirix (b800 image)
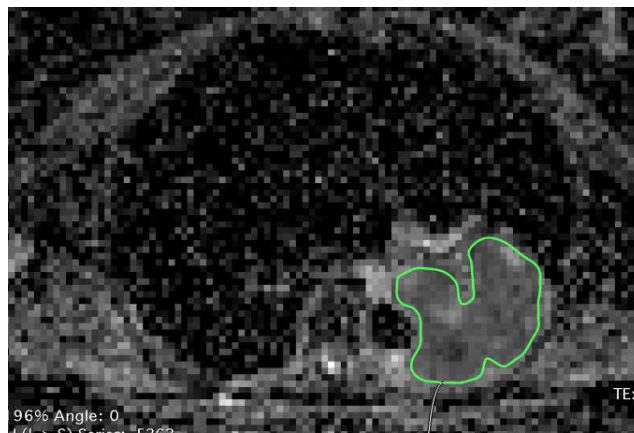

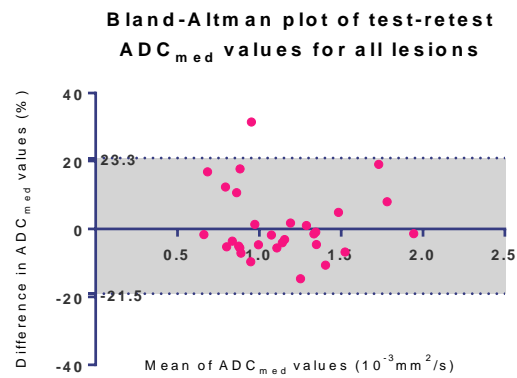**Figure 2 (d):** ROI generated using axial images in Osirix (ADC image)

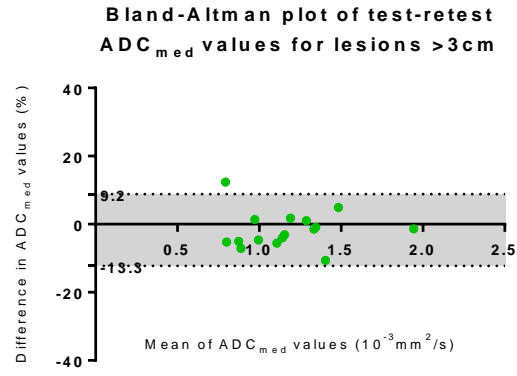**Figure 3 (a):** Bland Altman plot of test-retest repeatability for all lesions

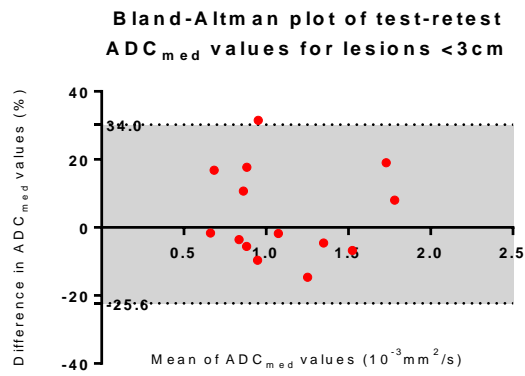**Figure 3 (b):** Bland Altman plot of test-retest repeatability for lesions >3cm

Figure 3 (c): Bland Altman plots of test-retest repeatability for lesions <3cm
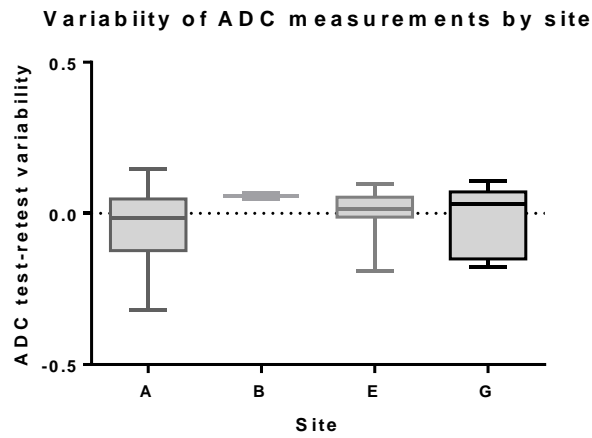
Figure 3 (d): Box plot of ADC_{med} (x10^{-3}mm^2/s) measurement variability by site (sites A, B, E and G)

Figure 4 (a): Scatterplot of ADC_{med} values ($x10^{-3}mm^2s^{-1}$) obtained for whole lesion segmentation on ADEPT (IDL based) and Osirix (c-DWI) post-processing software packages

Figure 4 (b): Bland Altman plot of ADC_{med} values obtained for whole lesion segmentation on ADEPT (IDL based) and Osirix (c-DWI) post-processing software packages

Variability of ADC measurements for different software packages

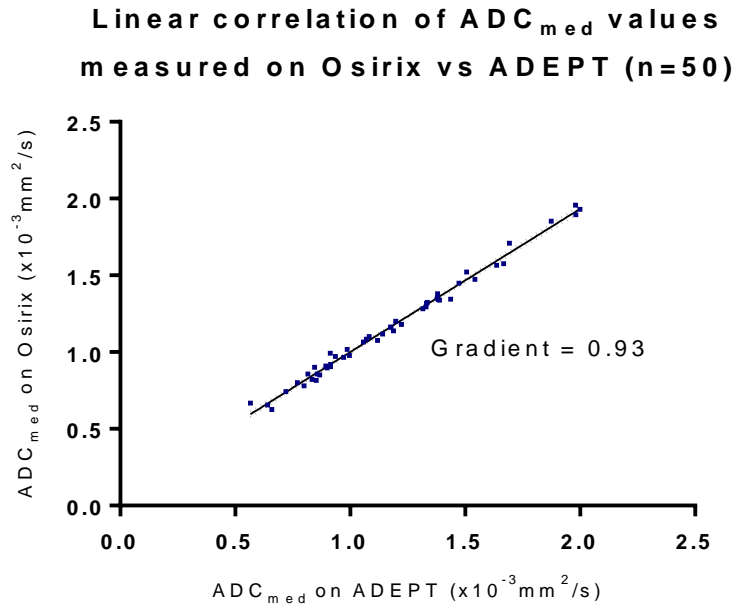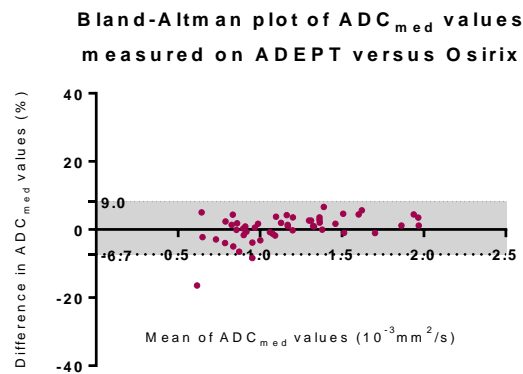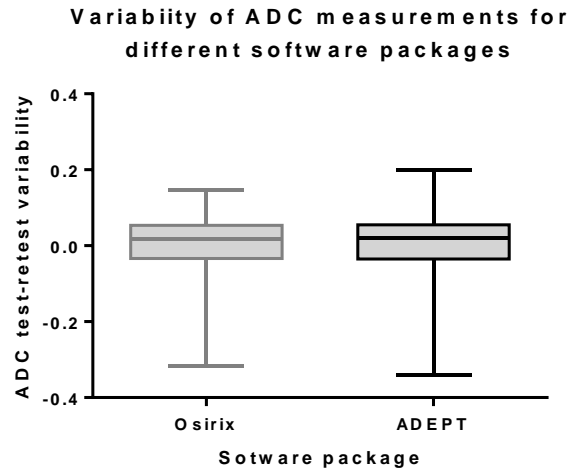**Figure 4 (c):** Boxplot confirming no significant difference in test-retest ADC$_{med}$ value variability for measurements generated by analysis on ADEPT (IDL based) and Osirix (c-DWI) software

Table 1: ADC parameters and tumour segmentation methodologies used within the literature to date.

Author, year	Patient group	ADC metric measured	% ADC increase with treatment
Reischauer 2014 (11)	9 patients, 13 lesions, NSCLC ADC 1 week after chemotherapy start	Mean ADC from freehand whole tumour segmentation performed on ADC maps	16.2% ADC increase for RECIST responders
Yu 2014 (12)	25 patients, NSCLC ADC after 1 cycle chemotherapy	Mean ADC from freehand segmentation of single central largest tumour slice performed on ADC maps	90% ADC increase for RECIST responders
Tsuchida 2013 (13)	28 patients, NSCLC ADC after 1 cycle chemotherapy	Mean ADC from freehand segmentation of single central largest tumour slice on b=800 images	21.5% cutoff for ΔADC differentiated RECIST responders from non-responders
Yabuuchi 2011 (14)	28 patients, NSCLC ADC after 1 cycle chemotherapy	Mean ADC within 3 representative regions of interest on ADC maps	'Good ADC increase' (mean ΔADC = 35.9%) had longer PFS and OS than 'poor ADC increase'
Sun 2011 (15)	21 patients, NSCLC ADC 1 week after chemotherapy	'Average' ADC from freehand segmentation of single central largest tumour slice on ADC maps	36% ADC increase for RECIST responders
Okuma 2009 (16)	17 patients. Lung tumour ADC 3 days after radiofrequency ablation (RFA)	'Average' ADC from setting an ROI in tumour on the single central largest tumour slice on ADC maps	29.6% ADC increase following RFA (higher increase for those that showed later local control)
Chang 2012 (17)	7 patients, NSCLC ADC mid- chemo-radiotherapy	'Average' ADC from 100mm^2 ROI placed central largest tumour slice on ADC maps	67.7% ADC increase for RECIST responders
Ohno 2012 (18)	64 patients, NSCLC	Mean ADC from circular	No ADC change measured

	ADC pre chemo-radiotherapy	ROIs placed on every tumour slice on the b=0 and b=1000 images	(baseline value only)
Regier 2012 (19)	41 patients, NSCLC ADC pre radiotherapy	Mean and minimum ADC values from polygonal ROIs encompassing whole tumour on ADC maps	No ADC change measured (baseline value only)
Bernardin 2014 (9)	8 patients (2 NSCLC, 2 SCLC, 4 metastatic lung lesions)	Mean and median ADC from segmentation of the central 3 slices of tumour	No ADC change measured (ADC inter-and intra-observer reproducibility)
Weiss 2016 (20)	10 patients, NSCLC ADC pre chemo-radiotherapy	Mean ADC from whole tumour and metastatic lymph node segmentation on b=1000smm^{-2} images	19-26% relative ADC increase from baseline

Table 2: DW-Imaging parameters (SS-EPI = single shot echo planar imaging)

Sequence	SS-EPI	Orientation	Axial (whole lung)
Acq. matrix	128 x 112 (87.5%)	**N^o signal averages**	1, repeated 4x
FOV read (mm)	380	**Frequency bandwidth (Hz per pixel)**	1400 – 1800
FOV phase (mm)	273mm	**PE direction**	AP
Pixel size (mm)	3 x 3	**Acceleration factor**	2
Slice gap (mm)	0	**Fat suppression**	STIR (TI : 180 ms)
Slice thickness (mm)	5	**b-values / s mm^{-2}**	100, 500, 800
TR (ms)	$\geq$ 8000	**Parallel imaging**	Yes
TE (ms)	72	**Diffusion gradient mode**	Trace (Gradient over-plus)

Table 3: Within patient repeatability of duplicate ADC_{med} ($x10^{-3}mm^2/s$) on test-retest scanning of pulmonary masses, for ADC measurement on Osirix.

Lesion	Overall mean of ADC_{med} ($x10^{-3}mm^2s^{-1}$) ($\pm$ sd)	*wCV* of ADC_{med} (%) (95% CI) [95% LoA (%)]	ADC_{med} ICC (95% CI)	Mean lesion diameter (cm) ($\pm$ sd)
All lesions (n=30), whole tumour segmentation	1.14 (0.33)	7.1 (5.7 – 9.6) [-18.0 to 21.9]	0.94 (0.88 to 0.97)	4.5 (2.4)
Site A (10 lesions)	1.08 (0.35)	9.5 (6.6 - 16.7) [-23.0 to 29.9]	0.93 (0.75 to 0.98)	3.3 (1.7)
Site B (2 lesions)	1.44 (0.11)	4.1 (2.1 -26.1) [-10.7 to 12.0]	0.97 (-0.80 to 1.00	2.6 (0.3)
Site E (13 lesions)	1.19 (0.34)	4.8 (3.5 – 7.8) [-12.5 to 14.3]	0.95 (0.86 to 0.99)	5.5 (2.5)
Site G (5 lesions)	1.01 (0.25)	7.8 (4.8 - 19.2) [-19.4 to 24]	0.90 (0.32 to 0.99)	4.9 (2.6)
Lesions > 3cm	1.17 (0.30)	3.9 (2.9 – 5.9) [-10.2 to 11.4]	0.98 (0.95 to 0.99)	6.2 (2.0)
Lesions < 3cm	1.10 (0.37)	9.6 (7.0 – 15.2) [-23.3 to 30.5]	0.92 (0.77 to 0.97)	2.5 (0.3)

Table 4: Reproducibility of duplicate pulmonary mass ADC_{med} ($x10^{-3}mm^2s^{-1}$) values generated on different post-processing platforms (IDL based ADEPT and Osirix) (CCC= concordance correlation coefficient)

Lesion	Overall mean of ADC_{med} ($x10^{-3}mm^2s^{-1}$) ($\pm$ sd)	CoV of ADC_{med} (%) (95% CI of CoV) [95% LoA]	CCC for ADC_{med} on ADEPT vs Osirix (95% CI)	Mean lesion diameter (cm) ($\pm$ sd)
n=25 lesions, segmentation performed on ADEPT	1.17 (0.36)	2.8 (2.3 – 3.4)	0.99 (0.989 to 0.996)	4.5 (2.4)
n=25 lesions, segmentation performed on Osirix	1.16 (0.34)	[-7.4 to 8.0]		

REFERENCES

1.	O'Flynn EA, DeSouza NM. Functional magnetic resonance: biomarkers of response in breast cancer. Breast Cancer Res. 2011;13(1):204.

2.	Kyriazi S, Collins DJ, Messiou C, Pennert K, Davidson RL, Giles SL, et al. Metastatic ovarian and primary peritoneal cancer: assessing chemotherapy response with diffusion-weighted MR imaging--value of histogram analysis of apparent diffusion coefficients. Radiology. 2011;261(1):182-92.

3.	Blackledge MD, Collins DJ, Tunariu N, Orton MR, Padhani AR, Leach MO, et al. Assessment of treatment response by total tumor volume and global apparent diffusion coefficient using diffusion-weighted MRI in patients with metastatic bone disease: a feasibility study. PloS one. 2014;9(4):e91779.

4.	Concatto NH, Watte G, Marchiori E, Irion K, Felicetti JC, Camargo JJ, et al. Magnetic resonance imaging of pulmonary nodules: accuracy in a granulomatous disease–endemic region. European radiology. 2016;26(9):2915-20.

5.	Shen G, Jia Z, Deng H. Apparent diffusion coefficient values of diffusion-weighted imaging for distinguishing focal pulmonary lesions and characterizing the subtype of lung cancer: a meta-analysis. European radiology. 2016;26(2):556-66.

6.	Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Statistical methods in medical research. 2015;24(1):9-26.

7.	Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277(3):813-25.

8.	Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Statistical methods in medical research. 2015;24(1):27-67.

9.	Bernardin L, Douglas NH, Collins DJ, Giles SL, O'Flynn EA, Orton M, et al. Diffusion-weighted magnetic resonance imaging for assessment of lung lesions: repeatability of the apparent diffusion coefficient measurement. European radiology. 2014;24(2):502-11.

10.	Cui L, Yin J-B, Hu C-H, Gong S-C, Xu J-F, Yang J-S. Inter-and intraobserver agreement of ADC measurements of lung cancer in free breathing, breath-hold and respiratory triggered diffusion-weighted MRI. Clinical imaging. 2016;40(5):892-6.

11.	Reischauer C, Froehlich JM, Pless M, Binkert CA, Koh DM, Gutzeit A. Early treatment response in non-small cell lung cancer patients using diffusion-weighted imaging and functional diffusion maps--a feasibility study. PloS one. 2014;9(10):e108052.

12.	Yu J, Li W, Zhang Z, Yu T, Li D. Prediction of early response to chemotherapy in lung cancer by using diffusion-weighted MR imaging. TheScientificWorldJournal. 2014;2014:135841.

13.	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. Journal of magnetic resonance imaging : JMRI. 2013;38(1):80-8.

14.	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(2):598-604.

15.	Sun YS, Cui Y, Tang L, Qi LP, Wang N, Zhang XY, et al. Early evaluation of cancer response by a new functional biomarker: apparent diffusion coefficient. AJR American journal of roentgenology. 2011;197(1):W23-9.

16.	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. The British journal of radiology. 2009;82(984):989-94.

17.	Chang Q, Wu N, Ouyang H, Huang Y. Diffusion-weighted magnetic resonance imaging of lung cancer at 3.0 T: a preliminary study on monitoring diffusion changes during chemoradiation therapy. Clinical imaging. 2012;36(2):98-103.

18.	Ohno Y, Koyama H, Yoshikawa T, Matsumoto K, Aoyama N, Onishi Y, et al. Diffusion-weighted MRI versus 18F-FDG PET/CT: performance as predictors of tumor treatment response and patient survival in patients with non-small cell lung cancer receiving chemoradiotherapy. AJR American journal of roentgenology. 2012;198(1):75-82.

19.	Regier M, Derlin T, Schwarz D, Laqmani A, Henes FO, Groth M, et al. Diffusion weighted MRI and 18F-FDG PET/CT in non-small cell lung cancer (NSCLC): does the apparent diffusion coefficient (ADC) correlate with tracer uptake (SUV)? European journal of radiology. 2012;81(10):2913-8.

20.	Weiss E, Ford JC, Olsen KM, Karki K, Saraiya S, Groves R, et al. Apparent diffusion coefficient (ADC) change on repeated diffusion-weighted magnetic resonance imaging during radiochemotherapy for non-small cell lung cancer: A pilot study. Lung cancer. 2016;96:113-9.

21.	O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nature reviews Clinical oncology. 2016.

22.	Douglas N, Winfield J, deSouza NM, Collins DJ, Orton MO Development of a phantom for quality assurance in multicentre clinical trials with diffusion-weighted MRI. Proceedings of the International Society of Magnetic Resonance in Medicine 2013:Presentation number 3114.

23.	Satoh S, Kitazume Y, Ohdama S, Kimula Y, Taura S, Endo Y. Can malignant and benign pulmonary nodules be differentiated with diffusion-weighted MRI? AJR American journal of roentgenology. 2008;191(2):464-70.

24.	Blackledge MD, Leach MO, Collins DJ, Koh DM. Computed diffusion-weighted MR imaging may improve tumor detection. Radiology. 2011;261(2):573-81.

25.	Forkman J. Estimator and tests for common coefficients of variation in normal distributions. Communications in Statistics—Theory and Methods. 2009;38(2):233-51.

26.	Weller A, O'Brien ME, Ahmed M, Popat S, Bhosle J, McDonald F, et al. Mechanism and non-mechanism based imaging biomarkers for assessing biological response to treatment in non-small cell lung cancer. European journal of cancer. 2016;59:65-78.

27.	Liu Y, deSouza NM, Shankar LK, Kauczor HU, Trattnig S, Collette S, et al. A risk management approach for imaging biomarker-driven clinical trials in oncology. Lancet Oncol. 2015;16(16):e622-8.

28.	Wade O. Movements of the thoracic cage and diaphragm in respiration. The Journal of physiology. 1954;124(2):193-212.

29.	Plathow C, Fink C, Ley S, Puderbach M, Eichinger M, Zuna I, et al. Measurement of tumor diameter-dependent mobility of lung tumors by dynamic MRI. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2004;73(3):349-54.

30.   Hogg N, Winfield J, Collins DJ, deSouza NM, Orton M. Development of a perfusion insensitivemeasurement of the apparent diffusion coefficient: a simulation. ESMRMB 2012, 29th Annual Scientific Meeting, Lisbon, 4-6 October. 2012;Book of Abstracts (57).

31.   Taouli B, Beer AJ, Chenevert T, Collins D, Lehman C, Matos C, et al. Diffusion-weighted imaging outside the brain: Consensus statement from an ISMRM-sponsored workshop. Journal of magnetic resonance imaging : JMRI. 2016;44(3):521-40.

32.   Brihuega-Moreno O, Heese FP, Hall LD. Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage. Magnetic resonance in medicine. 2003;50(5):1069-76.

33.   Saritas EU, Lee JH, Nishimura DG. SNR dependence of optimal parameters for apparent diffusion coefficient measurements. IEEE transactions on medical imaging. 2011;30(2):424-37.

34.   Tan ET, Marinelli L, Slavens ZW, King KF, Hardy CJ. Improved correction for gradient nonlinearity effects in diffusion-weighted imaging. Journal of magnetic resonance imaging : JMRI. 2013;38(2):448-53.

35.   Rata M, Collins DJ, Darcy J, Messiou C, Tunariu N, Desouza N, et al. Assessment of repeatability and treatment response in early phase clinical trials using DCE-MRI: comparison of parametric analysis using MR- and CT-derived arterial input functions. European radiology. 2015.

36.   Ashton E, Raunig D, Ng C, Kelcz F, McShane T, Evelhoch J. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. Journal of magnetic resonance imaging : JMRI. 2008;28(3):791-6.

37.   Rata M, Collins DJ, Darcy J, Messiou C, Tunariu N, Desouza N, et al. Assessment of repeatability and treatment response in early phase clinical trials using DCE-MRI: comparison of parametric analysis using MR- and CT-derived arterial input functions. European radiology. 2016;26(7):1991-8.

38.   Scaranelo AM, Eiada R, Jacks LM, Kulkarni SR, Crystal P. Accuracy of unenhanced MR imaging in the detection of axillary lymph node metastasis: study of reproducibility and reliability. Radiology. 2012;262(2):425-34.

39.   Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia (New York, NY). 2009;11(2):102-25.