



# How reliable are ADC measurements? A phantom and clinical study of cervical lymph nodes

Bastien Moreau<sup>1</sup> · Antoine Iannessi<sup>1</sup> · Christopher Hoog<sup>1</sup> · Hubert Beaumont<sup>2</sup>

Received: 20 October 2017 / Revised: 11 December 2017 / Accepted: 20 December 2017 / Published online: 23 February 2018  
© The Author(s) 2018. This article is an open access publication

## Abstract

**Objective** To assess the reliability of ADC measurements in vitro and in cervical lymph nodes of healthy volunteers.

**Methods** We used a GE 1.5 T MRI scanner and a first ice-water phantom according to recommendations released by the Quantitative Imaging Biomarker Alliance (QIBA) for assessing ADC against reference values. We analysed the target size effect by using a second phantom made of six inserted spheres with diameters ranging from 10 to 37 mm. Thirteen healthy volunteers were also scanned to assess the inter- and intra-observer reproducibility of volumetric ADC measurements of cervical lymph nodes.

**Results** On the ice-water phantom, the error in ADC measurements was less than 4.3 %. The spatial bias due to the non-linearity of gradient fields was found to be 24 % at 8 cm from the isocentre. ADC measure reliability decreased when addressing small targets due to partial volume effects (up to 12.8 %). The mean ADC value of cervical lymph nodes was  $0.87 \cdot 10^{-3} \pm 0.12 \cdot 10^{-3}$  mm<sup>2</sup>/s with a good intra-observer reliability. Inter-observer reproducibility featured a bias of -5.5 % due to segmentation issues.

**Conclusion** ADC is a potentially important imaging biomarker in oncology; however, variability issues preclude its broader adoption. Reliable use of ADC requires technical advances and systematic quality control.

## Key Points

- ADC is a promising quantitative imaging biomarker.
- ADC has a fair inter-reader variability and good intra-reader variability.
- Partial volume effect, post-processing software and non-linearity of scanners are limiting factors.
- No threshold values for detecting cervical lymph node malignancy can be drawn.

**Keywords** Magnetic resonance imaging · Diffusion · Biomarkers · Lymph · Quantitative evaluation

## Abbreviations

ADC	Apparent diffusion coefficient
CV <sub>R</sub>	Coefficient of variation
FOV	Field of view
GE	General electric
IVIM	Intra-Voxel Incoherent Motion
LoA	Limit of agreements

MRI	Magnetic resonance imaging
PET	Positron emission tomography
QC	Quality control
QIB	Quantitative Imaging Biomarker
QIBA	Quantitative Imaging Biomarker Alliance
RC <sub>R</sub>	Repeatability coefficient
ROI	Region of Interest
RSNA	Radiological Society of North America
SD	Standard deviation
SNR	Signal-to-noise ratio
T	Tesla

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00330-017-5265-2>) contains supplementary material, which is available to authorized users.

✉ Hubert Beaumont  
Hubert.beaumont@mediantechnologies.com

<sup>1</sup> Department of Radiology, Centre Antoine Lacassagne, 06100 Nice, France

<sup>2</sup> Research and Development Department, Median Technologies, Les deux arcs - 1800 route des crêtes - Bat. B, 06560 Valbonne, France

## Introduction

Recent advances in medical imaging technology and drug therapeutics have accelerated the emergence of new quantitative imaging biomarkers (QIB) [1, 2]. The multiplication of these QIBs is unfortunately not always accompanied by

stringent validations establishing that QIBs are well designed to characterize a disease and its changes with therapy. This lack of validation creates a situation where QIBs are routinely used but with limited knowledge of their performances, precluding a larger adoption in clinical trials.

Apparent diffusion coefficient (ADC) can quantify the level of free water diffusion restricted by an increase in tissue cellularity. Applications of ADC in cancer imaging has motivated intensive research and ADC is now one of the main QIBs derived from diffusion MRI.

Several studies have documented the incremental value of ADC assessment as a complement or substitute to standard sequences for the detection of malignant tumours [3], the degree of malignancy [4, 5] or to evaluate response to treatment [6–8].

Since lymph node involvement is pivotal in oncological imaging [9], ADC has been tested for its detection of malignant adenomegalies [10, 11]. Results are discordant [12, 13].

Previous literature comprises heterogeneous studies protocols and results [14]. Several sequential unitary processes are necessary to output an ADC assessment, the lack of reliability of any of these unitary processes is likely to degrade the final ADC assessment. It is therefore particularly relevant to study if ADC qualifies as a quantitative biomarker.

Over the last decade, a multidisciplinary community has organized retrospective investigations of QIBs starting by documenting methodologies [2]. In 2007, the Radiological Society of North America (RSNA) launched QIBA (Quantitative Imaging Biomarker Alliance [15]), a specialized working group aiming at improving the value and usefulness of QIBs in reducing variability across devices, patients, and practices.

One of QIBA aims consists in releasing ‘Profiles’, which are documents standardizing imaging protocols to obtain optimal, reliable and reproducible biomarker measures according to the current state of the art. The QIBA diffusion imaging profile is still a work in progress [16].

QIBA also proposes a standardized protocol for quality control in diffusion imaging, using a diffusion phantom [17, 18] consisting of a volume of 0 °C stabilized water as the reference value for ADC assessment [19, 20].

The main objective of this study was to evaluate the variability of ADC measurements *in vitro* on a phantom and *in vivo* on cervical lymph nodes. The secondary objective was to understand and quantify ADC measurement errors, in view of correcting them in future studies.

## Methods

We first tested QIBA metrics for quality control (QC) of ADC image quality, and then performed a reliability analysis of

ADC measurements. Finally we measured ADC values of cervical lymph nodes in healthy volunteers.

This prospective study was conducted at the Centre Antoine Lacassagne, cancer centre in Nice, France, between March and November 2016. We used a GE MRI scanner 1.5T MR450W and ADW Volume Share 5.4.6 software to process images (GE Healthcare).

## Quality control test

We used a DIN 6858-1 PET-CT phantom (PTW) consisting of a cylindrical Plexiglas body filled with a mixture of ice and water. Three smaller cylinders were inserted into the body, one of which was filled with water at 0 °C (Fig. 1, left side).

Homogeneity of temperature inside the cylinder was thermometer-controlled according to the process defined into the QIBA profile to achieve thermal equilibrium (>1 h) over the entire MRI exam period. For each b value, four successive acquisitions spaced in time from more than 12 min were performed, allowing retrospective checks.

The diffusion protocol was 3 directions, DW SS-EPI with  $b=0, 100, 600, 800$  s/mm<sup>2</sup>, TR=9,451 ms, TE=80 ms, Number of average = 2, FOV 320\*320 mm, contiguous slice thickness of 4 mm, encoding frequency axis R/L.

Four successive acquisitions were made for each b value, the phantom symmetry axis was laser-centred to the magnetic field positioning the 0 °C water cylinder at the center of the scanner. Acquisitions of the phantom were performed horizontally (x-axis) and vertically (y-axis). We measured circular regions of interest (ROIs) of 2.5 cm diameter and composed of 123 voxels (Fig. 2). Mean ADC and standard deviation (SD) were computed.

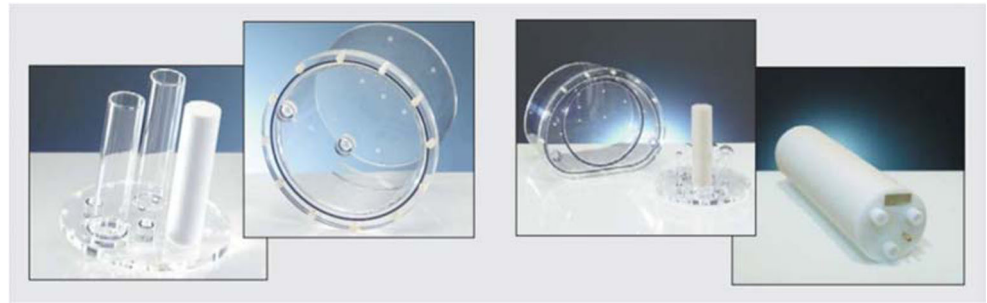
According to the equations in Table 1, we computed the measurement repeatability (R), estimated by the coefficient of variation (CV<sub>R</sub>) and the repeatability coefficient (RC<sub>R</sub>), the accuracy (ADC Bias estimate), ADC noise estimate and b-value dependency.

The signal-to-noise ratio (SNR) was computed using formula F (shown in Table 1) and involved computing the ‘Temporal Noise Image’ from the diffusion mapping at  $b = 0$ , with a 2-cm circular ROI.

Results were compared to QIBA ‘s references values [16].

In addition, we analysed the planar spatial correlation of ADC measures in shifting ROIs along the x and y axis. The ADC reference value was measured at the image center using formula C (see Table 1). We used circular ROIs of 2.2-cm diameter and 2-cm shifts from the centre either to the right (x-axis) or to the bottom (y-axis) of the image.

**Fig. 1** Phantoms used in the study. **Left** ICEWATER phantom filled with 0° C water (*DIN 6858-1, PTW, Freiburg, Germany*). **Right** SPHERE phantom at room temperature featuring spheres of various sizes between 10- and 37-mm diameters (NEMA NU2-2012, PTW, Freiburg, Germany)



**Measurement variability**

**SPHERE phantom study**

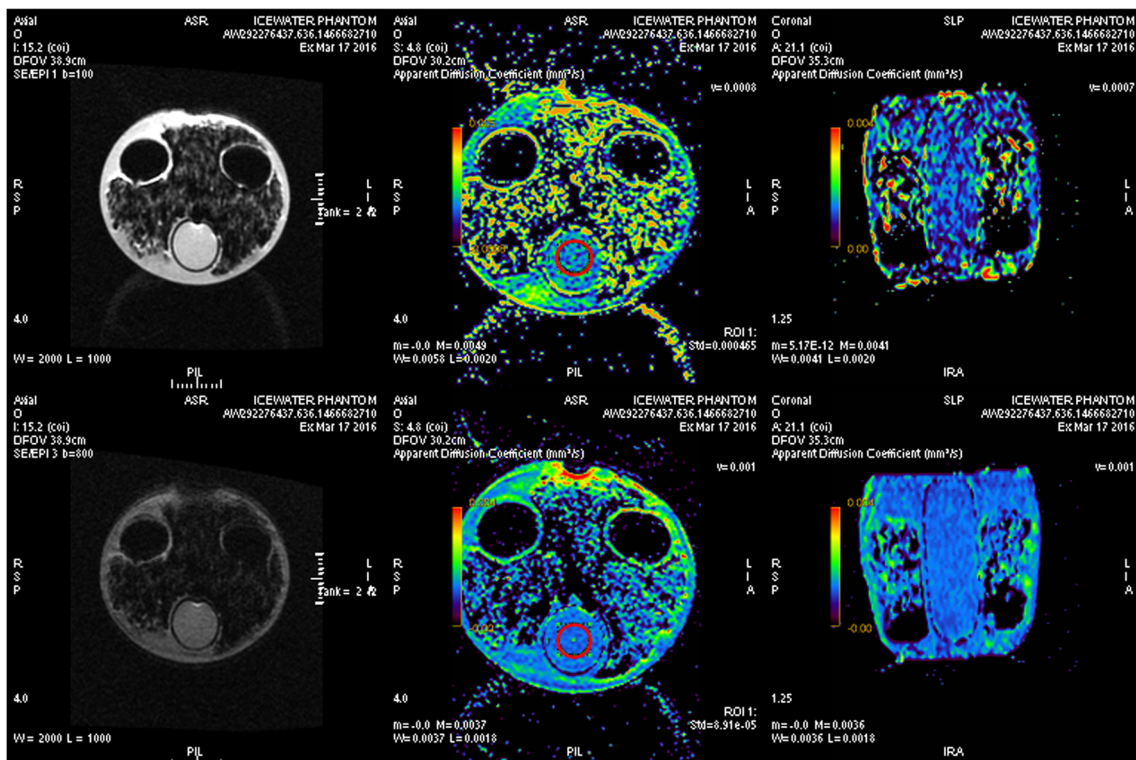
A second phantom was used (NEMA NU2-2012 (PTW)), called SPHERE Phantom (Fig. 1, right side). The SPHERE phantom embedded six different spheres (diameters 10, 13, 17, 22, 28 and 37 mm), filled with room temperature water.

We simulated clinical conditions in using the cervical level of the routine whole-body MRI, i.e. axial DW SS EPI with  $b=50$  and  $b=1,000$  s/mm<sup>2</sup>, TR=10,384 ms, TE set to minimum (around 70 ms for all scans). Number of averages=2, parallel imaging factor=2, FOV=400\*400 mm, contiguous 5-mm slice thickness, encoding frequency axis R/L. The phantom was laser centred, equidistant from all spheres. Four

acquisitions were made at 1-day intervals. All values were averaged over 4 days.

ADC measures were obtained from spherical volumes of interest (VOIs) centred on spheres (Fig. 3).

The relative ADC error was computed for each sphere size, considering that the reference ADC value was from the 37-mm sphere. We analysed the correlation between VOI size and precision of measurements in computing the CV<sub>R</sub>. Additional analysis documented the measurement error, first in measuring bias, second in computing the CV<sub>R</sub> through several concentric VOIs of decreasing size in the largest sphere, according to Table 1 (Formula A). Then partial volume effect was quantified by calculating the relative error within a VOI with a diameter equal to 80 % the diameter of a sphere compared to a VOI of identical size within the largest



**Fig. 2** Measurements on the ICEWATER phantom. Imaging of ICEWATER phantom at  $b_0$ - $b_{100}$  (top) and  $b_0$ - $b_{800}$  (bottom). From left to right: Diffusion mapping, axial view of ADC mapping and

coronal view of ADC mapping. Red circular regions of interest are set at the centre of the ice water cylinder

**Table 1** Definition of quality control metrics according to QIBA DW-MRI profile

	Metric	Definition
A	$CV_R = 100\% \cdot \frac{\sigma_R}{\mu_R}$	$CV_R$ : Coefficient of variation (%) $\sigma_R$ : standard deviation (mm <sup>2</sup> /s) of each measurements means $\mu_R$ : mean of each measurements ADC means (mm <sup>2</sup> /s)
B	$RC_R = 2.77 \cdot \sigma_R$	$RC_R$ : repeatability coefficient (mm <sup>2</sup> /s)
C	ADC bias estimate = $\mu - DC_{True}$	$DC_{True}$ : ADC=1,1.10 <sup>-3</sup> mm <sup>2</sup> /s in 0°C water
D	ADC noise Estimate = $100\% \cdot \frac{\sigma}{\mu}$ $\% \text{ bias} = 100\% \cdot \left( \frac{\mu - DC_{True}}{DC_{True}} \right)$	$\sigma$ : standard deviation of ADC values within the ROI (mm <sup>2</sup> /s) $\mu$ : mean ADC (mm <sup>2</sup> /s) within the ROI
E	ADC b value dependence = $100\% \cdot \left\  \frac{ADC_{b \text{ min}, b2} - ADC_{b \text{ min}, b1}}{ADC_{b \text{ min}, b1}} \right\ $	bmin = b0 b1 = b600 b2 = b800
F	$SNR_{nDyn} = \frac{\text{Spatial mean pixel value on Signal Image}}{\text{Spatial mean pixel value on Temporal Noise Image}}$	$SNR_{nDyn}$ : Signal to Noise Ratio

sphere. The mean and SD of ADC values were computed for all VOIs size.

**In vivo study**

Informed consent was obtained from 13 healthy volunteers. Exclusion criteria were chronic disease, history or ongoing symptoms of infection like fever, cough, rhinorrhoea, dysphagia and odynophagia, history of cervical surgery, claustrophobia and all usual contraindications for MRI. Demographic status and smoking habits were recorded for the 13 volunteers. Volunteers were scanned using the same machine as the phantom study. The acquisition was performed with a neck phased-array coil and the volunteer was instructed to breath normally.

Technical settings of diffusion sequence for volunteers were identical to those of the SPHERE phantom.

Two readers assessed ADC values of lymph nodes: a senior radiologist with more than 6 years of experience in cancer imaging and a junior radiologist.

Lymph node volumes were manually segmented on the b1000 scan, and the graphic was exported to the ADC map (Fig. 4). At least four lymph nodes were selected per volunteer, including the largest. VOIs were segmented in delineating hyper-intense diffusion areas on b1000 scans while excluding lymph nodes hilum. Each node was segmented twice by each observer using the same acquisition with an interval of 7–60 days (mean 41 days) [21]. Mean and SD ADC values were recorded.

Inter- and intra-observer agreements were calculated according to the Bland Altman method using R CRAN software. Bias and limit of agreement (LoA) were computed. Inter- and intra-observer differences in segmenting lymph node volumes and ADC values were analysed using the sum of Wilcoxon rank for paired values test. A *p*-value < 0.05 was considered significant.

**Results**

**Quality control test**

Table 2 shows QIBA’s recommended limit values for repeatability, accuracy, precision and b-value dependency.

We found that SNR computed from diffusion scans at b = 0 was 17 1, lower than the recommended limit value (50 1).

Variations in ADC measures relative to spatial positions are summarized in Fig. 5.

We found no significant correlation between ADC values and lateral shifts, but a significant correlation with vertical shifts (Pearson correlation coefficients, respectively  $\rho = 0.25$ , 95 % CI -0.45–0.76 and  $\rho = 0.95$ , 95 % CI 0.46–1).

We also found an increasing bias when shifting measurements from the scan isocenter (Table 3). The maximum 10 % error threshold recommended by QIBA was exceeded for ADC values measured at least 6 cm distant from the isocentre (Table 3).

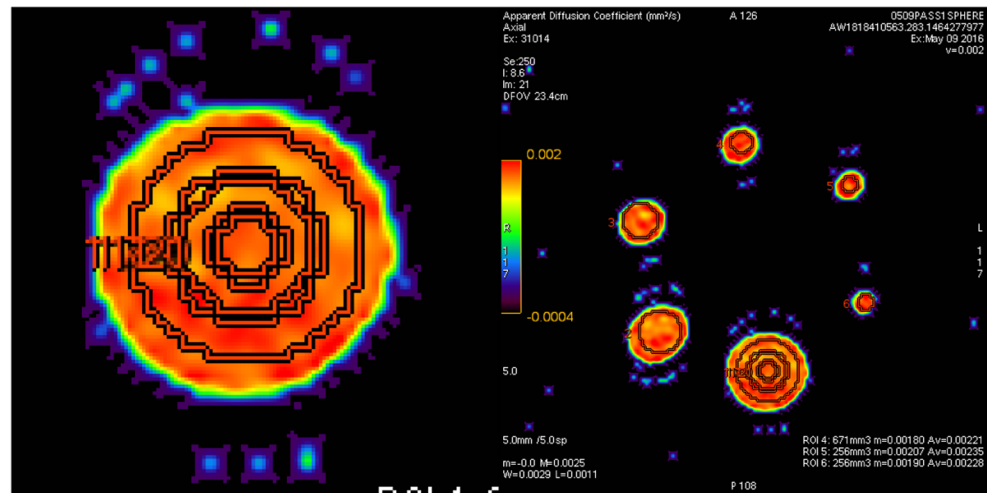
**Reliability analysis**

**SPHERE phantom study**

Firstly, our analysis showed that when VOIs are set within spheres of decreasing size, relative error and measurements variability of ADC measurements increased (Table 4). Secondly, we found no significant mean ADC difference for VOIs of decreasing sizes set within the largest sphere. In thoses case, we found less than 2 % error between the largest and smallest VOIs.

Correlation and variability analysis of ADC measurements with VOI size seemed to indicate a significant partial volume effect. Partial volume effect was visually confirmed on images.

**Fig. 3** Measurements on the SPHERE phantom. **Left** Spherical VOIs of decreasing sizes centered on the largest sphere. **Right** Spherical volumes of interest (VOIs) centered on sphere of various sizes. VOIs diameters are set to 80 % of physical sphere's diameters



### In vivo study

Thirteen volunteers were included in the in vivo study. Age ranged from 22 to 50 years (mean 32.4), and gender ratio (M/F) was 38.5 %. Two volunteers were active smokers or recent ex-smokers (15.4 %). Overall, 54 cervical lymph nodes were selected for analysis mainly on carotid-jugular sites, with a mean volume of 1 cm<sup>3</sup> (Appendix 1).

The mean value of measured ADC was  $0.87 \times 10^{-3}$  mm<sup>2</sup>/s ( $0.66\text{--}1.28 \cdot 10^{-3}$  mm<sup>2</sup>/s, SD was  $0.12 \cdot 10^{-3}$  mm<sup>2</sup>/s). We found a significant difference between the average ADC values measured by readers 1 and 2 ( $0.84 \cdot 10^{-3}$  and  $0.90 \cdot 10^{-3}$  mm<sup>2</sup>/s, respectively,  $p < 0.0001$ ).

The inter-reader analysis showed a relative bias of -5.5 %, LoA was [-18.8 %; 7.7 %]. The absolute bias was  $0.045 \cdot 10^{-3}$  mm<sup>2</sup>/s, LoA was [-0.146; 0.056].

We found a significant difference in average segmented volumes between readers 1 and 2 (respectively  $1.18 \pm 0.94$  cm<sup>3</sup> and  $1.92 \pm 1.23$  cm<sup>3</sup>,  $p < 0.0001$ ). There was a low correlation between measurement differences in terms of average ADC and volume segmentation ( $R^2 = 0.37$ ) by the two observers.

Intra-observer analysis showed, respectively, for readers 1 and 2, a relative bias = 0.6 %; LoA=[-9.2 %; 10.4 %] and relative bias = 0.5 %; LoA=[-8.8 %; 7.7 %].

Using the Beaumont et al. method [22] and based on our intra-observer reproducibility parameters, we can estimate that on longitudinal studies under strict reproducible conditions (same patient, same reader), a meaningful relative change of ADC value should be outside the range [-13 %; +15 %].

### Discussion

Our QC results showed good compliance with QIBA metrics, except for ADC bias estimate, which was slightly above the

limit, and with a variability of about 9 %. Results were independent of the value of b.

We questioned if the main part of error was due to our phantom design featuring a large off-axis volume of water and thermally suboptimal materials. Using repeated imaging of the phantom, we found, however, a good repeatability, suggesting acceptable thermal equilibrium.

SNR was also lower than QIBA's recommendation, but Malyarenko et al. [18] reported that low SNR has no impact on ADC assessments. Very low SNR without adequate post-processing would probably alter measurements as most software (including the one we used) compute ADC images in thresholding/removing low intensity voxels.

We highlighted a correlation between ADC measurement error and the distance of ROIs from the magnetic centre. The error increased with bottom-shift (up to 24 % when located 8 cm out of isocentre). Conversely, with regard to lateral-shift we found no correlation with the magnitude of errors. This result can be explained by non-uniformity of gradient-fields.

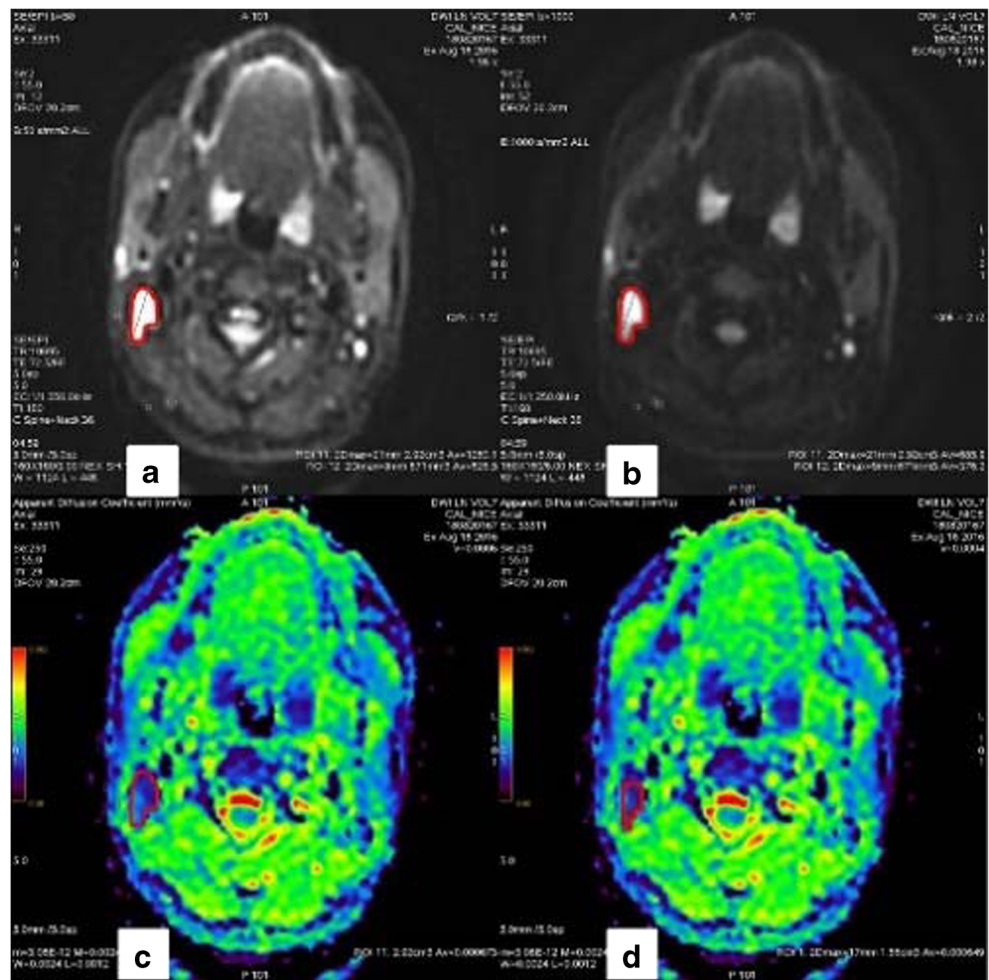
As we found a correlation between variability of ADC assessments and contour segmentations, we concluded that partial volume effect was a major contributor to the variability.

A visual review of outliers in clinical data confirmed high variation of signal intensity in the tissue surrounding these lymph nodes. Consequently, even a small variation in segmentations led to a significant modification of ADC assessments (Fig. 6)

We recommend measuring ADC by drawing ROIs smaller than the anatomical limits of the area of interest. How to optimize segmentation margins must be further investigated.

We found excellent repeatability and good reproducibility. This suggests that, if ADC intended to evaluate response to treatment, changes inferior to [-13 %; +15 %] may not be clinically relevant. However, longitudinal reproducibility would require further clinical studies to take into account all variability factors.

**Fig. 4** Measurements of cervical lymphnodes. Imaging of a healthy volunteer’s cervical lymph node. **(a)** Diffusion mapping at  $b = 50$ . **(b)** mapping at  $b = 1000$  on which the volume of interest (VOI) is contoured before being exported to other series. **(c)** and **(d)** Mapping of the apparent diffusion coefficient. In red, the VOI is determined by operator 1 **(a, b and c)** and by operator 2 **(d)**



According to our dataset, the averaged ADC value for healthy subject's cervical lymph nodes was  $0.87 \cdot 10^{-3} \pm 0.12 \cdot 10^{-3} \text{ mm}^2/\text{s}$ .

Our results are well supported by the literature.

Regarding the correlation between variability of ADC assessments and contour segmentation, heterogeneous segmentation methods are available but several studies

documented the reproducibility issues [23–25] affecting these methods. These different approaches are also reported as cumbersome and time-consuming [26].

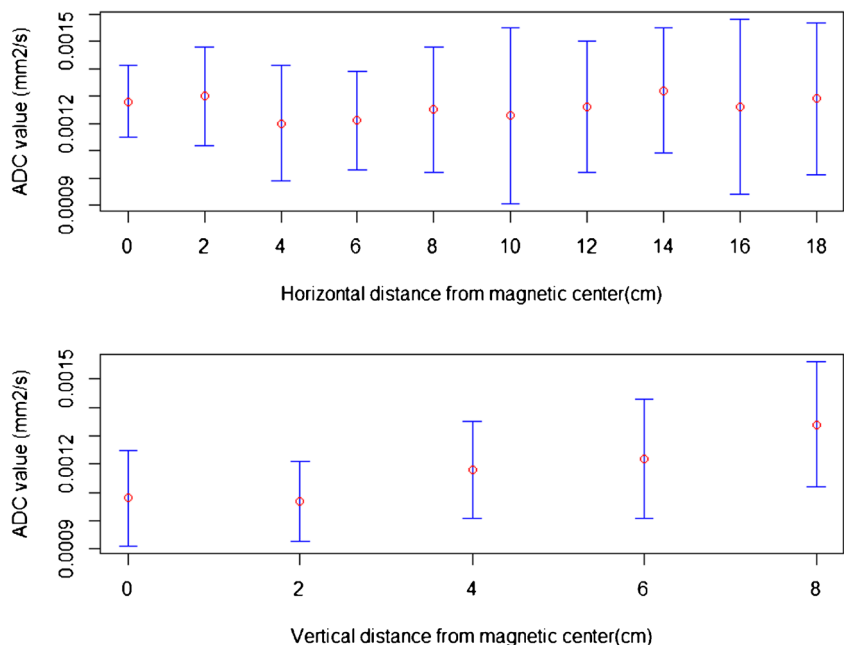
Specific phantom studies have shown that gradient-field error would be scanner-dependent [27] and not significant within 4 cm from the isocentre, explaining the good reproducibility of our ADC measurements and in other multicentric

**Table 2** Quality control after imaging the ICEWATER phantom

		QIBA claims	Results		
			b100	b600	b800
Repeatability	Coefficient of variation	< 1,5	2,5	0,5	1,3
	CV <sub>R</sub> (%)				
	Coefficient of repeatability RC <sub>R</sub> (mm <sup>2</sup> /s)	< 1,5 · 10 <sup>-5</sup>	<b>7,96<sup>-5</sup></b>	<b>1,60 · 10<sup>-5</sup></b>	<b>4,16 · 10<sup>-5</sup></b>
Accuracy	ADC Bias Estimate (%)	< 3,6	3,4	<b>4,1</b>	<b>4,3</b>
Precision	ADC Noise Estimate (%)	< 2	<b>41,1</b>	<b>9,2</b>	<b>8,8</b>
<b>b-value dependency (%) (b600-b800)</b>		< 2	0,7		

Outcome of the quality control after imaging the ICEWATER phantom. The test was done with different b values as b0-b100, b0-b600 and b0-b800. Tests not meeting QIBA quality claims are displayed in bold

**Fig. 5** Spatial correlation of ADC. Top view apparent diffusion coefficient (ADC) changes according to the horizontal distance from magnetic centre. Horizontal axis distance in cm. Vertical axis ADC value in mm<sup>2</sup>/s. **Bottom view** ADC changes according to the vertical distance from magnetic center. **Horizontal axis** Distance in cm. **Vertical axis** ADC value in mm<sup>2</sup>/s



studies. On multiple scanners, measurements at 12 cm from the isocentre showed an average error of -20 % according to vertical shift and +7 % horizontally.

Unlike our observation of an ADC value of  $0.87 \cdot 10^{-3} \pm 0.12 \cdot 10^{-3} \text{ mm}^2/\text{s}$  in healthy subject's cervical lymph nodes, Kwee et al. [12] reported a range of  $[1.15 \cdot 10^{-3} \text{ mm}^2/\text{s}; 1.18 \cdot 10^{-3} \text{ mm}^2/\text{s}]$ , with similar intra- and inter-observer variabilities. A review of 12 studies including more than 1,200 benign lymph nodes report ADC values of  $0.302 \pm 0.062 \cdot 10^{-3} \text{ mm}^2/\text{s}$  in inflammatory cervical nodes [28] and  $2.38 \pm 0.29 \cdot 10^{-3} \text{ mm}^2/\text{s}$  for abdominal nodes [29]. Kwee et al. concluded that disparity of results could be due to the various segmentation methods used.

Our results for non-diseased ADC values overlap with metastatic or lymphomatous lymph nodes ADC measures  $[0.410 \pm 0.105 \cdot 10^{-3} \text{ mm}^2/\text{s}; 1.84 \pm 0.37 \cdot 10^{-3} \text{ mm}^2/\text{s}]$  as reported by other groups [28, 29]; however, the ADC values we found match with other non-diseased ADC studies. A radiological-pathological correlation study by Vandecaveye et al. [30] on 331 cervical lymph nodes proposed an ADC threshold of  $0.94 \cdot 10^{-3} \text{ mm}^2/\text{s}$  for detecting node malignancy featuring a

specificity of 94 %. According to this threshold, 72 % of our data would have been misclassified. The use of ADC values to assess cervical lymph nodes malignancy does not reach a consensus.

Other variability factors are described in the literature.

- 1) Inter-scanner variability. Some authors [18] report less than 3 % variability while others [31] conclude that 80 % of scanners featured less than 5 % error. Another group [32] reports that the CV<sub>R</sub> was 1.5 % on phantoms, and less than 4 % for cerebral parenchyma. Impact of acquisition parameters has also been investigated [33–35].
- 2) Impacts of post-processing software were documented by Zeilinger et al. [36], reporting up to 8 % variation when ADC was processed by four different types of software. This limitation precludes longitudinal assessment of patients across different centers.

In order to address the limitations we found, variability can be minimized by standardizing ADC assessments. Efforts from scanner manufacturers are needed [8] to ease the

**Table 3** Spatial variations of apparent diffusion coefficient (ADC) measurements

Horizontal shift	<b>Distance from magnetic centre (cm)</b>	<b>0</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>8</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>16</b>	<b>18</b>	Pearson
	<i>Shifted measurement / Reference at magnetic centre (%)</i>	0.0	1.6	-6.3	-5.5	-2.3	-3.9	-1.6	3.1	-1.6	0.8	$\rho=0.25$
Vertical shift	<b>Distance from magnetic centre (cm)</b>	<b>0</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>8</b>						Pearson
	<i>Shifted measurement / Reference at magnetic centre (%)</i>	0.0	-0.9	9.3	13.0	24.1						$\rho=0.95$

Spatial variations of ADC measurements with respect to a reference VOI at center of the magnetic field. Top rows ADC measurements are shifted horizontally. Bottom rows ADC measurements are shifted to the bottom. In bold are the shift values corresponding to the distance from magnetic centre in cm. We found that ADC measurement did not change significantly when shifted right (Pearson coefficient=0.25). In opposite, ADC values increased when measurements were shifted vertically (Pearson coefficient=0.95)

**Table 4** Scaling effect of apparent diffusion coefficient (ADC) measurements

True sphere’s diameter (mm)	<b>37</b>	<b>28</b>	<b>22</b>	<b>17</b>	<b>13</b>	<b>10</b>
Diameter of measured VOIs (mm)	30	22	18	14	10	10
Mean ADC over 4 days ( $\cdot 10^{-3}$ mm <sup>2</sup> /s)	2,05	2,07	2,13	2,15	2,31	1,97
Relative error of ADC measurements (%)	0 (ref)	0,98	3,9	4,76	12,8	-3,9
Coefficient of variation (%)	0,4	0,39	1,01	2,1	2,33	2,84

Repeated measurements of SPHERE Phantom images were performed over 4 days. Spheres of different sizes were measured. Top row (in bold) True size of spheres of interest. Second row Diameter of VOIs centered on spheres of interest. Third row For each sphere, measurements have been repeated four time over four days. Mean ADC values have been computed. Fourth row Relative error (%) with respect to the VOI set into the largest sphere. Bottom row Coefficient of variation (%) of repeated measurements over four days.

To be noted Regarding right side column, unlike for other measurements, true size sphere and VOI have same size (10mm) because of the sampling limit (8 Voxel into the VOI). As a consequence, a nonlinear effect has been observed Decrease of the Relative error and the mean ADC value. This point is further developed in Table 5

calibration of diffusion sequences. Also, initiatives from the scientific community and technological improvements are expected to avoid sequences artifacts and systematic errors. Correction of non-linear gradient fields, one of the relevant issues, is the focus of ongoing research [37, 38].

Design of standardized validation methodologies and systematic quality control [39] are key to reach acceptable levels of compliance. To this end, a commercial version of diffusion phantom has been developed by QIBA [40] and an automatic quality control software is under development.

From early quantitative evaluation of diffusion, other approaches have emerged.

IVIM (Intra-Voxel Incoherent Motion) is aiming to split the different components of ADC [41]. Some authors indicate that pure molecular diffusion would be more meaningful than ADC as being independent from the perfusive component [42]. Therefore, IVIM would enable assessing tissue for which ADC is conceptually limited. Other studies develop alternative diffusion-derived QIBs in using the non-Gaussian distribution of diffusion kurtosis, Q-ball imaging and spectrum analysis in particular.

These approaches also require metrological analysis before scientific validation can be obtained, although the scientific literature shows potential added value of ADC.

The physiopathology of ADC changes still remains unknown. Some malignant tumours may feature increased ADC with respect to healthy tissue, either by spontaneous necrosis or cystic transformation, or by the destruction of parenchyma with spontaneously low ADC. Some benign tumours may lead to ADC restriction; this is the case with Warthin tumours of salivary glands due to their cellular wealth, which is superior to the normal salivary tissue [43].

Lastly, microscopic changes triggered by anti-cancer treatments may interfere with ADC assessment follow-up [44] post-chemotherapy cytotoxic oedema leading to increased restriction of ADC despite a therapeutic response, and delayed fibrosis that may lead to suspicion of recurrence by dropping the ADC. Other confounding factors such as the appearance of extracellular oedema by hyper-hydration or by regional venous obstacles may also mislead the evaluation of therapeutic responses.

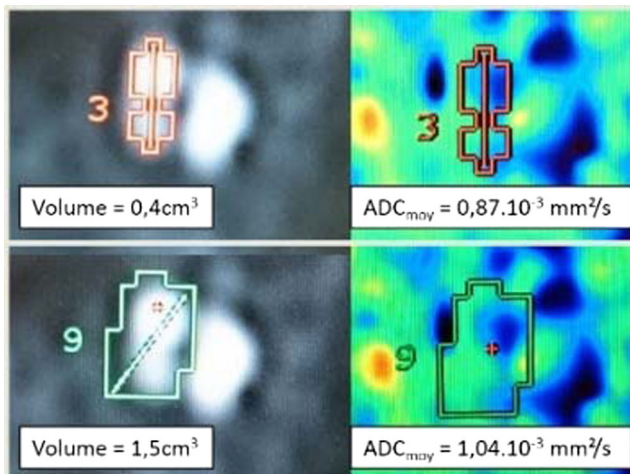
We found several limitations in our study. Phantom designs are not optimal and may have biased our results, limiting interpretations. First, regarding our ICEWATER phantom, even if the thermal equilibrium seemed satisfactory, a 1 °C error could lead to a 2.4 % ADC error, explaining the slightly higher value found compared to QIBA recommendation. Second, with regard to the SPHERE phantom: (1) without

**Table 5** Differential measurements: Largest sphere Vs spheres of smaller size

True sphere’s diameter (mm)	<b>37 vs 28</b>	<b>37 vs 22</b>	<b>37 vs 17</b>	<b>37 vs 13</b>	<b>37 vs 10</b>
Diameter of measured VOI (mm)	22 vs 22	18 vs 18	14 vs 14	10 vs 10	10 vs 10
Number of voxels sampling the measured VOI	136	92	21	8	8
Mean ADC measurements ( $\cdot 10^{-3}$ mm <sup>2</sup> /s)	2,01 2,06	2,01 2,11	2,04 2,2	2,04 2,38	2,04 1,92
ADC Relative error (%)	2,5	5,0	7,8	16,7	5,9

Repeated measurements of SPHERE Phantom images were performed over 4 days. Spheres of different sizes were measured. Top row (in bold) True size of spheres of interest (first value is the size of reference sphere, second value the size of sphere of interest). Second row Diameter of measured VOIs centered on spheres. Third row Number of voxel sampling VOIs. Fourth row For each sphere, measurements have been repeated four time over four days. Mean ADC values have been computed (first value is the size of reference sphere, second value the size of sphere of interest). Bottom row Relative error (%) with respect to the VOI set into the largest sphere as a reference.





**Fig. 6** Example of inter-observer discordance in terms of volume and apparent diffusion coefficient (ADC). Example of inter-observer discordance in terms of volume and ADC on a level II lymph node. **Top row** First reader's measurements. **Bottom row** Second reader's measurements. **Left** b1000 diffusion maps where volumes of interest (VOIs) are drawn. **Right** corresponding ADC maps. Note: The heterogeneity of the node's environment featuring areas of high ADC values (green and red in the right images) without clear correspondence on the diffusion image

thermal control, measurements were relative, limiting generalizability, (2) comparison of distant ROIs without gradient field correction was prone to bias, and (3) analysis of partial volume effect can be affected by phantom material, gas and features. Third, we did not evaluate the impact of artifacts [45]. QIBA recommends the use of corrective methods for liver and kidney analysis [8], although other groups [46, 47] report no improvement in using such methods. Fourth, the monocentric design of our study limits the generalizability of our results.

At this time, variabilities from different sources preclude a larger adoption of the ADC biomarker even though it is an important advance in cancer imaging. Generalizing quality controls and standardization of measurements is crucial to overcome these ADC variability issues.

**Acknowledgements** We acknowledge Catherine Klifa for her multiple reviews and scientific advices. All authors have equally contributed to this work.

**Funding** The authors state that this work has not received any funding.

### Compliance with ethical standards

**Guarantor** The scientific guarantor of this publication is Dr. Bastien Moreau, PhD.

**Conflict of interest** Hubert Beaumont, as co-author of this manuscript, declares relationships with the following companies: Median Technologies.

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.

**Statistics and biometry** One of the authors has significant statistical expertise.

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

**Ethical approval** Institutional Review Board approval was obtained.

### Methodology

- prospective
- experimental
- performed at one institution

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### References

1. Kessler LG, Barnhart HX, Buckler AJ et al (2015) The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res* 24:9–26
2. Sullivan DC, Obuchowski NA, Kessler LG et al (2015) Metrology standards for quantitative imaging biomarkers. *Radiology* 277: 813–825
3. Thomassin-Naggara I, Dechoux S, Bonneau C et al (2013) How to differentiate benign from malignant myometrial tumours using MR imaging. *Eur Radiol* 23:2306–2314
4. Hambrock T, Somford DM, Huisman HJ et al (2011) Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* 259: 453–461
5. Higano S, Yun X, Kumabe T et al (2006) Malignant astrocytic tumors clinical importance of apparent diffusion coefficient in prediction of grade and prognosis. *Radiology* 241:839–846
6. Taouli B, Beer AJ, Chenevert T et al (2016) Diffusion-weighted imaging outside the brain Consensus statement from an ISMRM-sponsored workshop DWI Outside the Brain. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.25196>
7. Padhani AR, Liu G, Mu-Koh D et al (2009) Diffusion-weighted magnetic resonance imaging as a cancer biomarker consensus and recommendations. *Neoplasia* 11:102–125
8. Patterson DM, Padhani AR, Collins DJ (2008) Technology Insight water diffusion MRI—a potential new biomarker of response to cancer therapy. *Nat Clin Pract Oncol* 5:220–233
9. Torabi M, Aquino SL, Harisinghani MG (2004) Current concepts in lymph node imaging. *J Nucl Med* 45:1509–1518
10. King AD, Ahuja AT, Yeung DK et al (2007) Malignant cervical lymphadenopathy diagnostic accuracy of diffusion-weighted MR imaging. *Radiology* 245:806–813
11. Choi EK, Kim JK, Choi HJ et al (2009) Node-by-node correlation between MR and PET/CT in patients with uterine cervical cancer diffusion-weighted imaging versus size-based criteria on T2WI. *Eur Radiol* 19:2024–2032
12. Kwee TC, Takahara T, Luijten PR, Nieuwelstein RAJ (2010) ADC measurements of lymph nodes Inter- and intra-observer reproducibility study and an overview of the literature. *Eur J Radiol* 75:215–220

13. Seber T, Caglar E, Uylar T et al (2015) Diagnostic value of diffusion-weighted magnetic resonance imaging differentiation of benign and malignant lymph nodes in different regions of the body. *Clin Imaging* 39:856–862
14. Deng J, Wang Y (2017) Quantitative magnetic resonance imaging biomarkers in oncological clinical trials. Current techniques and standardization challenges. *Chronic Dis Transl Med* 3:8–20
15. QIBA. In RSNA. <http://www.rsna.org/qiba/>. Accessed 1 Sept 2017
16. RSNA Perfusion, Diffusion and Flow-MRI Biomarker Ctte - QIBA Wiki. [http://qibawiki.rsna.org/index.php/Perfusion,\\_Diffusion\\_and\\_Flow-MRI\\_Biomarker\\_Ctte](http://qibawiki.rsna.org/index.php/Perfusion,_Diffusion_and_Flow-MRI_Biomarker_Ctte). Accessed 1 Sept 2017
17. Chenevert TL, Galbán CJ, Ivancevic MK et al (2011) Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. *J Magn Reson Imaging* 34:983–987
18. Malyarenko D, Galbán CJ, Londy FJ et al (2013) Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. *J Magn Reson Imaging* 37:1238–1246
19. Harris KR, Woolf LA (1980) Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water. *J Chem Soc Faraday Trans 1 Phys Chem Condens Phases* 76:377
20. Mills R (1973) Self-diffusion in normal and heavy water in the range 1–45 deg. *J Phys Chem* 77:685–688
21. Robbins KT, Clayman G, Levine PA et al (2002) Neck dissection classification update revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology-Head and Neck Surgery. *Arch Otolaryngol Head Neck Surg* 128:751–758
22. Beaumont H, Souchet S, Labatte JM et al (2015) Changes of lung tumour volume on CT - prediction of the reliability of assessments. *Cancer Imaging*. <https://doi.org/10.1186/s40644-015-0052-2>
23. Giannotti E, Waugh S, Priba L et al (2015) Assessment and quantification of sources of variability in breast apparent diffusion coefficient (ADC) measurements at diffusion weighted imaging. *Eur J Radiol* 84:1729–1736
24. Ma C, Liu L, Li J et al (2016) Apparent diffusion coefficient (ADC) measurements in pancreatic adenocarcinoma A preliminary study of the effect of region of interest on ADC values and interobserver variability Effects of ROI on ADC Measurements in PDAC. *J Magn Reson Imaging* 43:407–413
25. Blazic IM, Lilic GB, Gajic MM (2017) Quantitative assessment of rectal cancer response to neoadjuvant combined chemotherapy and radiation therapy: comparison of three methods of positioning region of interest for ADC measurements at diffusion-weighted MR imaging. *Radiology* 282(2):418–428. <https://doi.org/10.1148/radiol.2016151908>
26. Ahlawat S, Khandheria P, Del Grande F et al (2016) Interobserver variability of selective region-of-interest measurement protocols for quantitative diffusion weighted imaging in soft tissue masses Comparison with whole tumor volume measurements. *J Magn Reson Imaging* 43:446–454
27. Malyarenko DI, Newitt D, Wilmes LJ et al (2016) Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials Cross-Platform ADC Bias Characterization. *Magn Reson Med* 75:1312–1323
28. Sumi M, Sakihama N, Sumi T et al (2003) Discrimination of metastatic cervical lymph nodes with diffusion-weighted MR imaging in patients with head and neck cancer. *Am J Neuroradiol* 24:1627–1634
29. Akduman EI, Momtahan AJ, Balci NC et al (2008) Comparison between malignant and benign abdominal lymph nodes on diffusion-weighted Imaging. *Acad Radiol* 15:641–646
30. Vandecaveye V, De Keyzer F, Vander Poorten V et al (2009) Head and neck squamous cell carcinoma value of diffusion-weighted MR imaging for nodal staging 1. *Radiology* 251:134–146
31. Belli G, Busoni S, Ciccarone A et al (2016) Quality assurance multicenter comparison of different MR scanners for quantitative diffusion-weighted imaging. *J Magn Reson Imaging* 43:213–219
32. Grech-Sollars M, Hales PW, Miyazaki K et al (2015) Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain Multi-centre reproducibility of diffusion MRI using clinical sequences. *NMR Biomed* 28:468–485
33. Nogueira L, Brandão S, Nunes RG et al (2015) Breast DWI at 3 T influence of the fat-suppression technique on image quality and diagnostic performance. *Clin Radiol* 70:286–294
34. Lu Y, Hatzoglou V, Banerjee S et al (2015) Repeatability investigation of reduced field-of-view diffusion-weighted magnetic resonance imaging on thyroid glands. *J Comput Assist Tomogr* 1. <https://doi.org/10.1097/RCT.0000000000000227>
35. Wang H-Y, Wang J, Tang Y-H et al (2015) Coronal diffusion-weighted magnetic resonance imaging of the kidney agreement with axial diffusion-weighted magnetic imaging in terms of apparent diffusion coefficient values. *Chin Med J (Engl)* 128:499
36. Zeilinger MG, Lell M, Baltzer PAT et al (2016) Impact of post-processing methods on apparent diffusion coefficient values. *Eur Radiol*. <https://doi.org/10.1007/s00330-016-4403-6>
37. Malyarenko DI, Ross BD, Chenevert TL (2014) Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements ADC Spatial Bias Correction. *Magn Reson Med* 71:1312–1323
38. Malyarenko DI, Chenevert TL (2014) Practical estimate of gradient nonlinearity for implementation of ADC bias correction. *J Magn Reson Imaging* 40:1487–1495
39. Buckler AJ, Bresolin L, Dunnick NR et al (2011) Quantitative imaging test approval and biomarker qualification interrelated but distinct activities. *Radiology* 259:875–884
40. HPD, High Precision Devices I QIBA DWI Quantitative MRI (qMRI) Diffusion Phantom. <http://hpd-online.com/diffusion-phantom.php>. Accessed 1 Sept 2017
41. Le Bihan D, Breton E, Lallemand D et al (1988) Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* 168:497–505
42. Zhang Y-D, Wang Q, Wu C-J et al (2015) The histogram analysis of diffusion-weighted Intravoxel Incoherent Motion (IVIM) imaging for differentiating the Gleason grade of prostate cancer. *Eur Radiol* 25:994–1004
43. Yabuuchi H, Matsuo Y, Kamitani T et al (2008) Parotid gland tumors can addition of diffusion-weighted MR imaging to dynamic contrast-enhanced MR imaging improve diagnostic accuracy in characterization? *Radiology* 249:909–916
44. Thoeny HC, Ross BD (2010) Predicting and monitoring cancer treatment response with DW-MRI. *J Magn Reson Imaging* 322. <https://doi.org/10.1002/jmri.22167>
45. Le Bihan D, Poupon C, Amadon A, Lethimonnier F (2006) Artifacts and pitfalls in diffusion MRI. *J Magn Reson Imaging* 24:478–488
46. Kandpal H, Sharma R, Madhusudhan KS, Kapoor KS (2009) Respiratory-triggered versus breath-hold diffusion-weighted MRI of liver lesions comparison of image quality and apparent diffusion coefficient values. *Am J Roentgenol* 192:915–922
47. Choi JS, Kim M-J, Chung YE et al (2013) Comparison of breathhold, navigator-triggered, and free-breathing diffusion-weighted MRI for focal hepatic lesions diffusion-weighted MRI techniques of the Liver. *J Magn Reson Imaging* 38:109–118