European Radiology
https://doi.org/10.1007/s00330-020-07522-0

MAGNETIC RESONANCE

Repeatability and reproducibility of ADC measurements: a prospective multicenter whole-body-MRI study

Nicolas F. Michoux[1] · Jakub W. Ceranka[2] · Jef Vandemeulebroucke[2] · Frank Peeters[1] · Pierre Lu[1] · Julie Absil[3] · Perrine Triqueneaux[1] · Yan Liu[4] · Laurence Collette[4] · Inneke Willekens[5] · Carola Brussaard[5] · Olivier Debeir[6] · Stephan Hahn[6] · Hubert Raeymaekers[5] · Johan de Mey[5] · Thierry Metens[3] · Frédéric E. Lecouvet[1]

Received: 20 July 2020 / Revised: 31 August 2020 / Accepted: 13 November 2020

Abstract

Objectives Multicenter oncology trials increasingly include MRI examinations with apparent diffusion coefficient (ADC) quantification for lesion characterization and follow-up. However, the repeatability and reproducibility (R&R) limits above which a true change in ADC can be considered relevant are poorly defined. This study assessed these limits in a standardized whole-body (WB)-MRI protocol.

Methods A prospective, multicenter study was performed at three centers equipped with the same 3.0-T scanners to test a WB-MRI protocol including diffusion-weighted imaging (DWI). Eight healthy volunteers per center were enrolled to undergo test and retest examinations in the same center and a third examination in another center. ADC variability was assessed in multiple organs by two readers using two-way mixed ANOVA, Bland-Altman plots, coefficient of variation (CoV), and the upper limit of the 95% CI on repeatability (RC) and reproducibility (RDC) coefficients.

Results CoV of ADC was not influenced by other factors (center, reader) than the organ. Based on the upper limit of the 95% CI on RC and RDC (from both readers), a change in ADC in an individual patient must be superior to 12% (cerebrum white matter), 16% (paraspinal muscle), 22% (renal cortex), 26% (central and peripheral zones of the prostate), 29% (renal medulla), 35% (liver), 45% (spleen), 50% (posterior iliac crest), 66% (L5 vertebra), 68% (femur), and 94% (acetabulum) to be significant.

Conclusions This study proposes R&R limits above which ADC changes can be considered as a reliable quantitative endpoint to assess disease or treatment-related changes in the tissue microstructure in the setting of multicenter WB-MRI trials.

Key Points

- *The present study showed the range of R&R of ADC in WB-MRI that may be achieved in a multicenter framework when a standardized protocol is deployed.*
- *R&R was not influenced by the site of acquisition of DW images.*
- *Clinically significant changes in ADC measured in a multicenter WB-MRI protocol performed with the same type of MRI scanner must be superior to 12% (cerebrum white matter), 16% (paraspinal muscle), 22% (renal cortex), 26% (central zone and peripheral zone of prostate), 29% (renal medulla), 35% (liver), 45% (spleen), 50% (posterior iliac crest), 66% (L5 vertebra), 68% (femur), and 94% (acetabulum) to be detected with a 95% confidence level.*

Keywords Whole-body imaging · Diffusion magnetic resonance imaging · Reproducibility of results · Disease progression · Cancer

✉ Nicolas F. Michoux
nicolas.michoux@uclouvain.be

[1] Institut de Recherche Expérimentale & Clinique (IREC) – Radiology Department, Université Catholique de Louvain (UCLouvain) – Cliniques Universitaires Saint Luc, Avenue Hippocrate 10, B-1200 Brussels, Belgium

[2] Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[3] Radiology Department, Université libre de Bruxelles, Hôpital Erasme, Brussels, Belgium

[4] European Organisation for Research and Treatment of Cancer, Brussels, Belgium

[5] MRI Department, UZ Brussel, Jette, Belgium

[6] LISA (Laboratories of Image Synthesis and Analysis), Ecole Polytechnique de Bruxelles, Université libre de Bruxelles, Brussels, Belgium

Published online: 06 January 2021

Abbreviations

95% CI	95% confidence interval
ADC	Apparent diffusion coefficient
CoV	Coefficient of variation
DICOM	Digital Imaging and Communications in Medicine
DWI	Diffusion-weighted imaging
LoA	Limits of agreement
R&R	Repeatability and reproducibility
RC	Repeatability coefficient
RDC	Reproducibility coefficient
ROI	Region of interest
WB	Whole body

Introduction

Whole-body (WB)-MRI with diffusion-weighted imaging (DWI) is a sensitive tool for visual detection of abnormal foci within bones and soft tissue, in cancer and inflammatory diseases [1–16]. The evaluation of apparent diffusion coefficient (ADC) derived from DWI is used for lesion characterization and assessment of the response to treatment [17]. However, how ADC is affected by disease or treatment is still poorly understood. Thus, its suitability as a quantitative imaging biomarker of tissue microstructure remains to be determined, and measurement reliability remains a central issue [18–21]. Three questions must be addressed: To what extent is ADC specific for the underlying microstructure? How should it be measured? How reliable is the measurement?

Reliability refers to high repeatability and reproducibility (R&R) of diffusion measurements. The in vitro R&R of ADC has been evaluated via an ice-water phantom, and both R&R measurements have been estimated in the range of 5% [22], with differences of only a few percentage points between centers [23–25]. Although no significant effect of field strength is observed with standard *b* values [26], an increase in ADC values for increasing *b* values has been reported [22]. Additionally, ADC depends on the diffusion gradient direction and spatial positioning of the region of interest (ROI) relatively to the scanner isocenter [23, 27]. Previous studies have shown that the reproducibility decreases at 3.0 T in abdominal regions [26, 28]. In vivo, a discrepancy remains in the literature regarding R&R of ADC [29].

These results indicate that the hardware and analysis methodology in DWI affect measurements and R&R of ADC. Hardware issues are related to scanner and coil technologies specific to MRI manufacturers, which vary in gradient performances and definition of acquisition parameters [21, 27]. Issues in analysis methodology result from selection of DWI sequences, lack of standardized Digital Imaging and Communications in Medicine (DICOM) tags for storing acquisition parameters and ADC maps [30], and the processing implemented to compute the ADC map [31, 32]. An additional factor is whether spatial alignment between MRI examinations and automated segmentation of the ROI is performed [33–35].

The current study assessed the in vivo R&R of ADC in WB-MRI when hardware and analysis methodology are kept constant, canceling any bias from the contribution of these two factors to R&R of ADC, and thus improving the applicability of the present results by other centers performing trials using the same imaging method on the same MRI scanner. To achieve this objective, healthy volunteers were recruited who did not report any symptoms or previous serious medical history, and an ice-water phantom (fixed at 0 °C temperature) was used.

Materials and methods

Study design

A multicenter platform set up to conduct oncological imaging trials (PICRIB, Platform for Imaging in Clinical Research in Brussels), including 3 universities and the European Organization for Research and Treatment of Cancer (EORTC), was used for the study. Between November 2016 and July 2017, 24 healthy, asymptomatic volunteers were prospectively recruited to undergo WB-MRI examinations; these comprised 10 women (mean (range) age, 38 (25–54) years) and 14 men (mean (range) age, 37 (23–57) years) from three university hospitals (designated C1, C2, and C3). Written informed consent was obtained from all patients. Institutional review boards from each center approved the protocol and consent form. Each center recruited eight volunteers among its staff who underwent two consecutive MRI scans (tests 1 and 2) at the recruiting institution and an additional scan (test 3) at an alternative center. Test 1 and 2 examinations were performed 1 h apart, whereas test 3 was performed on either the same day (n = 16) or within 7 days (n = 8).

MRI protocol

The study followed the general framework proposed by the Quantitative Imaging Biomarker (QIB) Alliance (QIBA) to make confident claims on the magnitude of a (true) change in the ADC that may be detected [36]. Therefore, the same imaging method on the same MRI scanner and the same analysis methodology were used at each imaging session. As no model-specific parameters for acquisition of WB-DWI are provided yet by QIBA, a protocol previously optimized for quantitative study of bone and soft tissue metastases was implemented [37]. Thus, each center used the same scanner (Philips Ingenia 3.0 T, software versions 5.1.7/5.1.7.2 at C1/C3 and 5.3.0/5.3.0.3 at C2) and followed the same WB-MRI acquisition protocol: bore diameter, 70 cm; gradients, 45 mT m^{-1}, 200 T m^{-1} s^{-1}; and dual radiofrequency transmit system. The body coil was used for transmission, and

Table 1 Acquisition parameters

Characteristic	Functional sequence WB-DWI	Morphologic sequence WB 3D T1-weighted imaging
Acquisition time	15.8 min	11.2 min (3 × 3.7 min)
Sequence	Diffusion-EPI (single shot/single spin echo)	3D SPACE (turbo spin echo)
Plane	Axial	Coronal
Slices/thickness/gap	185 (35 + 3 × 50)/6 mm/0.1 mm	3 × 210/1.2 mm/—
FOV/acquisition matrix	440 mm × 352 mm/98 × 78	500 mm × 300 mm/440 × 230
Number of stations	4*	4*
Coverage in *z*-axis	1128 mm	900 mm
Phase encoding	Anterior-posterior	Feet-head
SENSE factor	3	1.7 × 2.5
Phase oversampling	0	1.3
TR/TE/NSA/TF	6048 ms/66 ms/1/—	330 ms/21 ms/2/50
Flip angle	90°	90°/variable
Bandwidth (read out)	4198 Hz/pixel	1052 Hz/pixel
Fat suppression technique	IR (TI = 250 ms)	—
Specific parameters	*b* values: 0/150/1000 s mm^{-2}	—
Gradient directions	3 orthogonal (trace)	—
Diffusion time (δ, Δ)	(21.9 ms/33.2 ms)	—

FOV, field of view; *TR*, repetition time; *TE*, echo time; *NSA*, number of signal averages; *TF*, turbo factor; *SENSE*, SENSitivity Encoding

δ, duration of diffusion gradient lobe; Δ, time between centers of the two diffusion gradient lobes

*Phantom measurements were only performed for the first station using the HeadNeck coil

multichannel coils (with a maximum of 108 channels) were used for reception, as follows: FlexCoverage Posterior coil (integrated in the tabletop); FlexCoverage Anterior coils; HeadNeck coil and base (posterior part of HeadNeck coil on tabletop). Table 1 outlines acquisition parameters for morphologic WB 3D T1-weighted and functional WB-DWI sequences [37]. Anonymized DICOM images and ADC maps were stored in the picture archiving and communication system (PACS) at one center (C1) for a standardized central analysis.

Image analysis

Two independent readers (with 19 and 3 years of experience, respectively) used ImageJ software (version 1.45 [NIH]) to delineate 2D ROIs. Mean size (all readers and readings included) of the ROIs was 165 mm^2 in the prostate central zone (CZ), 184 mm^2 in the prostate peripheral zone (PZ), 222 mm^2 in the renal medulla, 259 mm^2 in the renal cortex, 259 mm^2 in the spleen, 259 mm^2 in the posterior iliac crest, 338 mm^2 in the muscle, 338 mm^2 in the L5 vertebral body, 338 mm^2 in the femur, 338 mm^2 in the acetabulum, 410 mm^2 in the liver, 249 mm^2 in the cerebrum white matter on ADC maps using T1-weighted images as visual reference. ROIs delineated during tests 2 and 3 were positioned on the same slice (and same area of the organ) than the ROI delineated during test 1, after taking into account possible patient movement and differences in slice orientation due to slight differences in patient position in the multichannel coils. Care was taken to select a homogeneous signal region without blood vessels and recognizable artifacts. The paraspinal muscle, prostate, liver, kidney, spleen, L5 vertebra, supra-acetabular bone (acetabulum), posterior iliac crest (ilium), femur, and white matter (cerebrum) were assessed. For the prostate, ROIs were first delineated on the b0 image and then replicated on the ADC map.

MRI phantom

To assess the accuracy of tissue diffusion measurements, we obtained phantom diffusion measurements by using a bottle filled with distilled ice-water inside a tube [17]. At thermal equilibrium (0 °C), the diffusion coefficient of water is D = 1099 × 10^{-6} mm^2 s^{-1} [22]. Measurements were performed in each center, for the first station of the WB-DWI sequence using the HeadNeck coil only, because of the limited size of the phantom. R&R was estimated from four scans with a circular ROI of 80 pixels from inside the water tube on ADC maps with a fixed axial slice. ADC maps were created using software provided by the MRI manufacturer.

Statistical analysis

Analysis was performed with MedCalc software (version 19.0.3). ADC variations were assessed according to the

Table 2 Mean value of the parameter ADC reported per center and per organ of interest, for both readers: comparison with reference values from published studies

ADC							
Center	Reader	Renal cortex	Renal medulla	Liver	Spleen	Prostate PZ	Prostate CZ
C1	1	1895	1759	1211	882	1538	1298
	2	1984	1816	1144	945	1606	1418
C3	1	1925	1716	1255	929	1527	1285
	2	1953	1824	1072	906	1533	1531
C2	1	1963	1729	1261	862	1513	1267
	2	1977	1854	1185	802	1479	1405
Overall mean value [95% CI]		1950 [1927; 1973]	1772 [1742; 1802]	1198 [1162; 1234]	890 [858; 923]	1525 [1485; 1564]	1348 [1300; 1394]
Reference studies		Thoeny [56]	Thoeny [56]	Jafar [29]	Jafar [29]	Gibbs [57]	Jafar [29]
		2030	1870	1280	850	1560	1310
Center	Reader	Muscle	Femur	Acetabulum	Ilium	L5	White matter
C1	1	1503	359	232	336	284	703
	2	1517	246	235	316	285	704
C3	1	1547	343	184	291	312	688
	2	1511	259	200	300	302	687
C2	1	1563	298	177	311	287	701
	2	1508	250	210	270	265	698
Overall mean value [95% CI]		1523 [1509; 1536]	288 [275; 300]	208 [195; 221]	303 [287; 320]	290 [271; 309]	697 [691; 703]
Reference studies		Jacobs [58]	Lavdas [59]	—	Messiou [60]		Grech-Sollars [61]
		1350	496		470		735
			Jacobs [58]		Jacobs [58]	Jacobs [58]	
			210		430	330	

Values of ADC are expressed in 10^{-6} mm^2 s^{-1}. Mean size of the ROIs (all volunteers) was 311 mm^2 (muscle), 338 mm^2 (femur), 358 mm^2 (L5), 285 mm^2 (ilium), 349 mm^2 (acetabulum), 409 mm^2 (liver), 237 mm^2 (spleen), 259 mm^2 (renal cortex), 222 mm^2 (renal medulla), 184 mm^2 (prostate PZ), 165 mm^2 (prostate CZ), and 249 mm^2 (cerebrum white matter)

Bland-Altman methodology [38]. Mean bias and limits of agreement (LoA defined as mean bias ± 1.96*standard deviation (SD) on paired differences) were derived from each Bland-Altman plot. The regression line of paired differences was also computed to investigate proportional differences with the magnitude of ADC.

Coefficient of variation (CoV) was estimated as CoV = SD divided by mean (%). The repeatability coefficient (RC), measuring ADC variations when the MRI scanner is constant, and the reproducibility coefficient (RDC), measuring ADC variations when different MRI scanners are used, were derived from the LoA (Supplemental Materials Methods). Note that RC per center and overall (all centers combined) was assessed from tests 1 and 2. RDC per pair of centers and overall (all centers combined) was assessed from tests 1 and 3.

The highest RC coefficient observed from both readers (corresponding to the worst case scenario in terms of repeatability) was identified for each organ. The upper limit of the associated 95% CI was considered as defining the R&R limit above which a clinically significant change in ADC in an individual patient can be detected with confidence when the same MRI scanner is used. The same reasoning was applied to the RDC coefficient to assess the R&R limit above which a clinically significant change in ADC can be detected when different MRI scanners are used.

The following statistical tests were finally performed. The significance level of these tests was adjusted according to the number of comparisons that were performed (Bonferroni correction). First, a Wilcoxon signed-rank test was also performed to assess phantom measurement bias between centers ($p < 0.0167$).

From Bland-Altman plots, a one-sample two-sided t test was performed to assess whether the mean bias was statistically different from 0 ($p < 0.0031$ for all organs except for the prostate, where $p < 0.0125$). A two-sided t test based on the null hypothesis that the slope of the regression line of paired differences was equal to 0 was performed ($p^{slope} < 0.0125$).

A two-way ANOVA with random effects was performed on a per-organ basis to assess potential differences in CoV

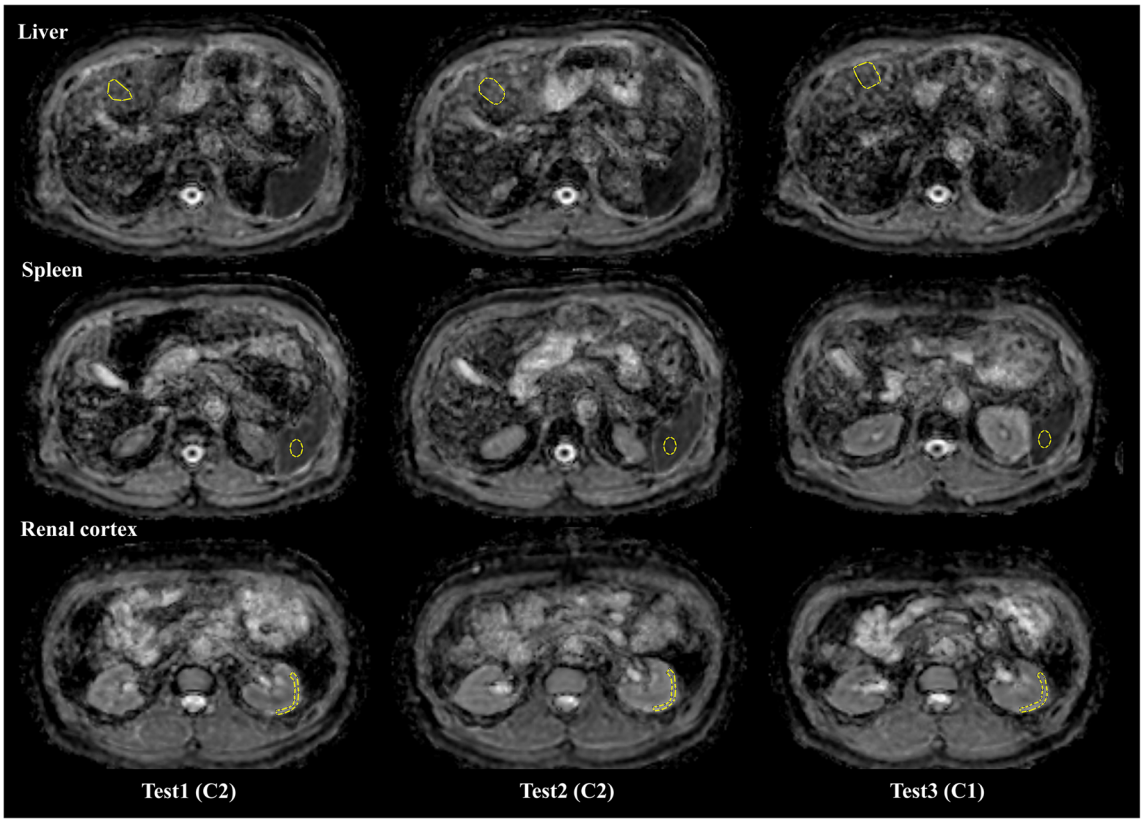

Fig. 1 MR images of the liver, spleen, and renal cortex in a 57-year-old healthy man. Apparent diffusion coefficient (ADC) maps illustrate the regions of interest (ROIs), as measured by reader 1. Tests 1 and 2 were obtained at center C2, and test 3 was obtained at center C1. Liver ADC values were 1277 × 10^{-6} mm^2 s^{-1} (test 1), 1345 × 10^{-6} mm^2 s^{-1} (test 2), and 1091 × 10^{-6} mm^2 s^{-1} (test 3). Spleen ADC values were 800 × 10^{-6} mm^2 s^{-1} (test 1), 755 × 10^{-6} mm^2 s^{-1} (test 2), and 746 × 10^{-6} mm^2 s^{-1} (test 3). Renal cortex ADC values were 1815 × 10^{-6} mm^2 s^{-1} (test1), 2260 × 10^{-6} mm^2 s^{-1} (test 2), and 1944 × 10^{-6} mm^2 s^{-1} (test 3)

between centers (fixed factor) and readers (random factor) for overall repeatability ($p < 0.0045$).

Results

Phantom measurements

Results from phantom measurements are provided in Supplemental Materials SI. CoV associated with four measurements of ADC on the same scanner ranged from 0.18% (C1), 0.54% (C2) to 1.20% (C3). The bias ranged from − 2.90% (C3) to + 4.40% (C2) (+ 3.70% for C1), with no significant difference between centers (all p values = 0.1250).

Healthy volunteer measurements

MR images from 3 illustrative cases are shown in Figs. 1, 2, and 3. The mean value of ADC (for all volunteers) is reported on a per-reader, per-center, and per-organ basis in Table 2. Data from reference studies were included for comparison.

Bland-Altman plots

A total of 168 Bland-Altman plots (4 repeatability + 4 reproducibility plots per-region, only 1 repeatability + 1 reproducibility plots for the central and peripheral zones of the prostate) and 2 readers were constructed. Figures 4 and 5 show examples of these plots for assessment of the overall R&R by reader 1.

Bland-Altman plots: limits of agreement

LoA were narrower in soft tissues compared to bone marrow regions (Tables 3 and 4). Equivalently, CoV, RC, and RDC were lower in soft tissues. The higher coefficient values obtained in bone marrow are associated with the inherently lower absolute values of ADC in these regions (Table 2). Cerebrum white matter had the lowest RC (reader 1: 7.5%; reader 2: 7.6%) and RDC (reader 1: 8.9%; reader 2: 8.4%) values, whereas acetabulum had the highest RC (reader 1: 59%; reader 2: 52%) and RDC (reader 1: 73%; reader 2: 58%) values.

The highest RC and RDC coefficients observed from both readers, and the upper limit of their 95% CI, are reported for each organ in Supplemental Materials SIII. R&R data from

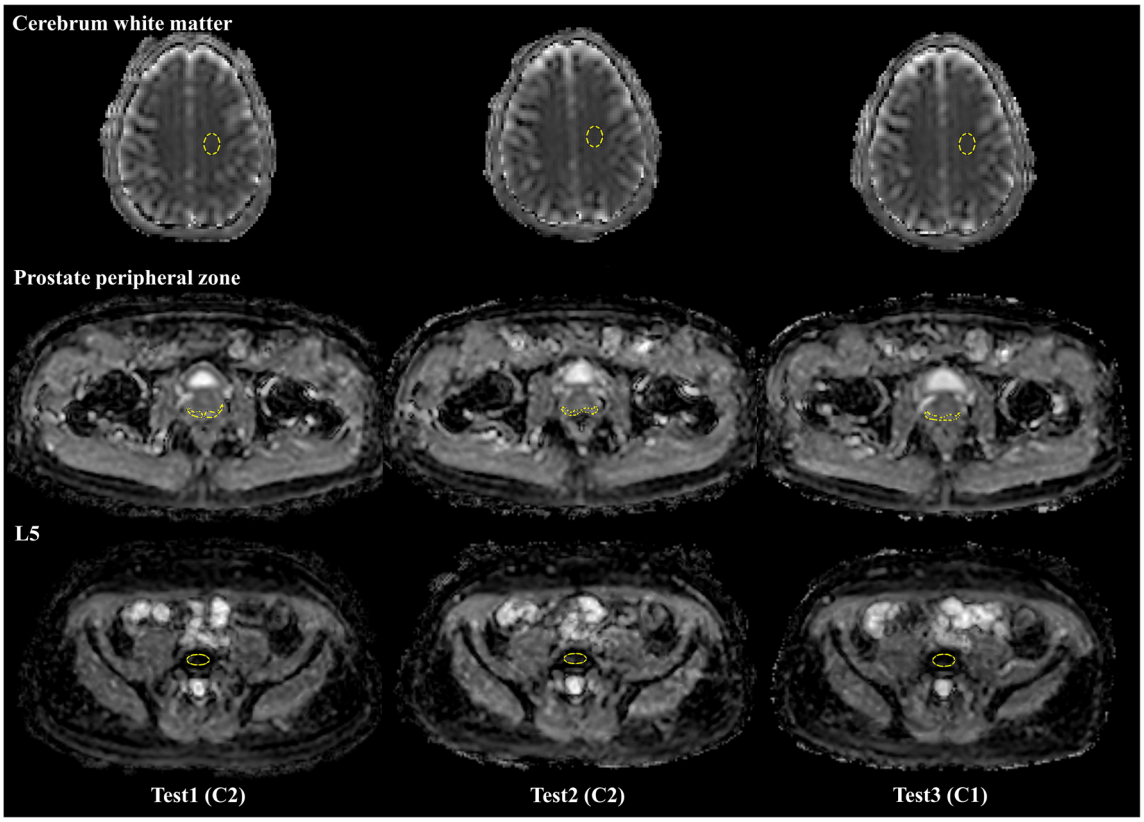

Fig. 2 MR images of cerebrum white matter, prostate peripheral zone, and L5 vertebra in a 57-year-old healthy man. Apparent diffusion coefficient (ADC) maps illustrate the regions of interest (ROIs), as measured by reader 1. Tests 1 and 2 were obtained at center C2, and test 3 was obtained at center C1. Cerebrum white matter ADC values were 708×10^{-6} mm^2 s^{-1} (test 1), 719×10^{-6} mm^2 s^{-1} (test 2), and 710×10^{-6} mm^2 s^{-1} (test 3). Prostate peripheral zone ADC values were 1480×10^{-6} mm^2 s^{-1} (test 1), 1690×10^{-6} mm^2 s^{-1} (test 2), and 1374×10^{-6} mm^2 s^{-1} (test 3). L5 vertebra ADC values were 204×10^{-6} mm^2 s^{-1} (test1), 226×10^{-6} mm^2 s^{-1} (test 2), and 249×10^{-6} mm^2 s^{-1} (test 3)

reference studies were added for comparison. In the repeatability study (the same MRI scanner is used), a true change in ADC must be superior to 9.7% (cerebrum white matter), 22% (renal cortex), 23% (prostate CZ), 26% (prostate PZ), 29% (renal medulla), 32% (liver), 45% (spleen), 44% (L5 vertebra), 46% (ilium), 55% (femur), and 76% (acetabulum) to be detected with a 95% confidence level by the current WB-MRI protocol.

In the reproducibility study (different MRI scanners are used), a true change in ADC must be superior to 12% (cerebrum white matter), 22% (renal cortex), 23% (renal medulla), 23% (prostate PZ), 26% (prostate CZ), 35% (liver), 44% (spleen), 50% (ilium), 66% (L5 vertebra), 68% (femur), and 94% (acetabulum) to be detected with confidence.

Bland-Altman plots: mean bias

Mean bias was not significantly different from 0, regardless of the study (repeatability or reproducibility), reader, or center (all p values > 0.0031) (Tables 3 and 4). A single statistical difference from reader 2 in the overall reproducibility of ADC for femur corresponds to a significant overestimation of 15.4% in test 3 compared to test 1 ($p = 0.0008$).

Bland-Altman plots: proportional differences

In the overall assessment of R&R (Figs. 4 and 5), no association between paired differences between readings and the mean of the readings was observed with a few exceptions. Review of repeatability data from reader 1 revealed an overestimation of ADC during test 2 compared to test 1 for small values of ADC becoming an underestimation for large values of ADC in the acetabulum ($p^{slope} = 0.0004$) (Fig. 4). Although the size of the ROIs in this region was 9.2% smaller during test 2 (two-sided paired t test, $p = 0.0081$), no significant correlation between the change in size and the change in ADC was found (Pearson correlation, $p = 0.5461$). From reader 2 (Bland-Altman plots not shown), an underestimation of ADC during test 2 compared to test 1 for small values of ADC becoming an overestimation for large values of ADC was observed in the kidneys ($p^{slope} < 0.0001$). The size of the ROIs in this region was not statistically different between the two readings ($p > 0.5000$). Review of reproducibility data

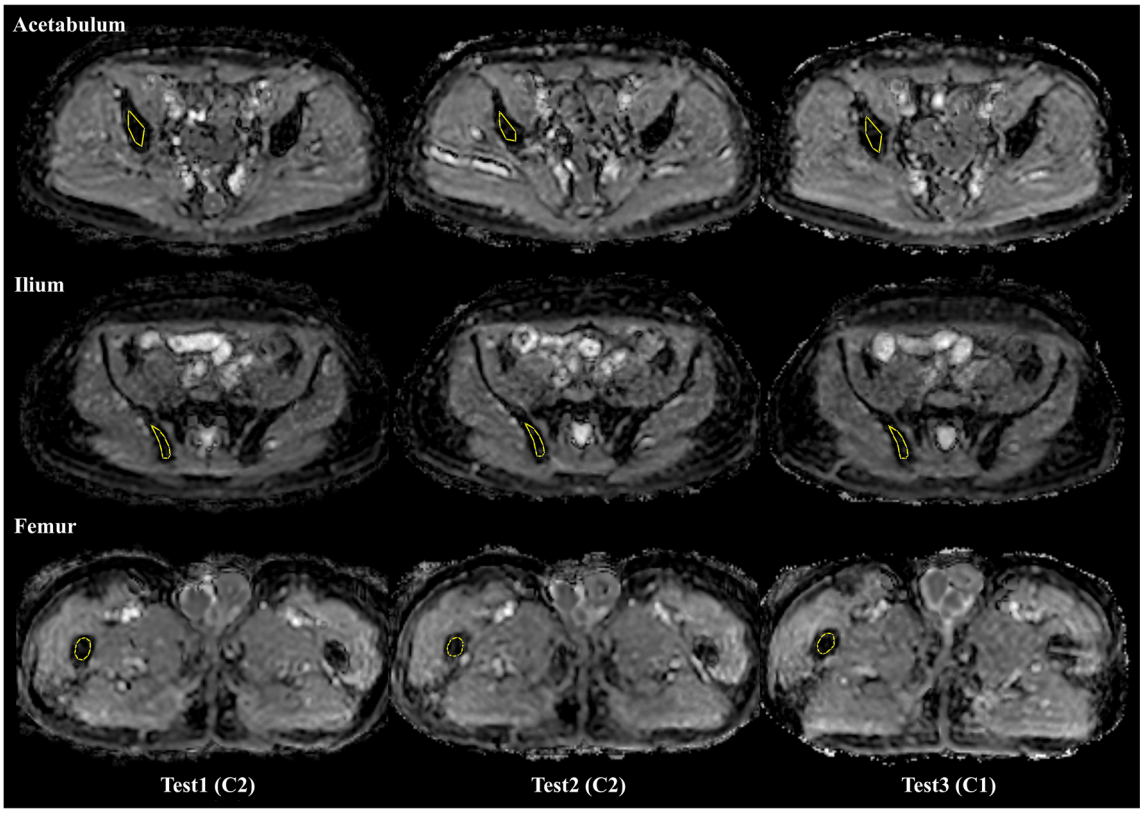

Fig. 3 MR images of the acetabulum, ilium, and femur in a 57-year-old healthy man. Apparent diffusion coefficient (ADC) maps illustrate the regions of interest (ROIs), as measured by reader 1. Tests 1 and 2 were obtained at center C2, and test 3 was obtained at center C1. Acetabulum ADC values were 117×10^{-6} mm^2 s^{-1} (test 1), 137×10^{-6} mm^2 s^{-1} (test 2), and 173×10^{-6} mm^2 s^{-1} (test 3). Ilium ADC values were 210×10^{-6} mm^2 s^{-1} (test 1), 212×10^{-6} mm^2 s^{-1} (test 2), and 245×10^{-6} mm^2 s^{-1} (test 3). Femur ADC values were 225×10^{-6} mm^2 s^{-1} (test 1), 266×10^{-6} mm^2 s^{-1} (test 2), and 192×10^{-6} mm^2 s^{-1} (test 3)

from reader 2 revealed an underestimation of ADC during test 3 compared to test 1 for small values of ADC becoming an overestimation for large values of ADC in muscle (p^{slope} = 0.0007). The size of the ROIs was not statistically different between the two readings ($p > 0.9999$).

Reader and center agreement

In the overall repeatability study, regardless of organ, CoV of ADC was not found to be influenced by the center or reader. All p values from the two-way mixed ANOVA were statistically non-significant (all p values > 0.0045) (Supplemental Materials SII).

Discussion

DWI and ADC measurements have been repeatedly presented as a key biomarker derived from MRI studies obtained at the level of individual organs and more recently at the level of the whole body [39, 40]. However, the number of multicentric clinical trials currently incorporating DWI sequences and ADC measurements, except in the brain, remains limited [41]. This is mainly due to a variety of acquisition and analysis methodologies employed at different centers, which make ADC values non-comparable between them. Our study addressed this need and from there should facilitate a step in this direction. In using the same imaging method on the same scanner, with the same methodology analysis (following [36]), the variability induced by the hardware and software was limited and thus, the best achievable (i.e., minimal) R&R of ADC using clinical WB-DWI was then assessable. This minimal variability achievable allows discussing (i) the applications for which ADC is a reliable quantitative endpoint, and (ii) which level of standardization may be mandatory to confidently detect true variations in ADC measurements at the level of multiple organs beyond the brain.

Firstly, the study assessed the accuracy and precision of WB-DWI. Accurate and repeatable values of ADC in the ice-water phantom were yielded showing the level of repeatability that may be achieved after protocol standardization over the three sites. Previous studies have demonstrated a repeatability of 2.3–5.0%, which is consistent with our results

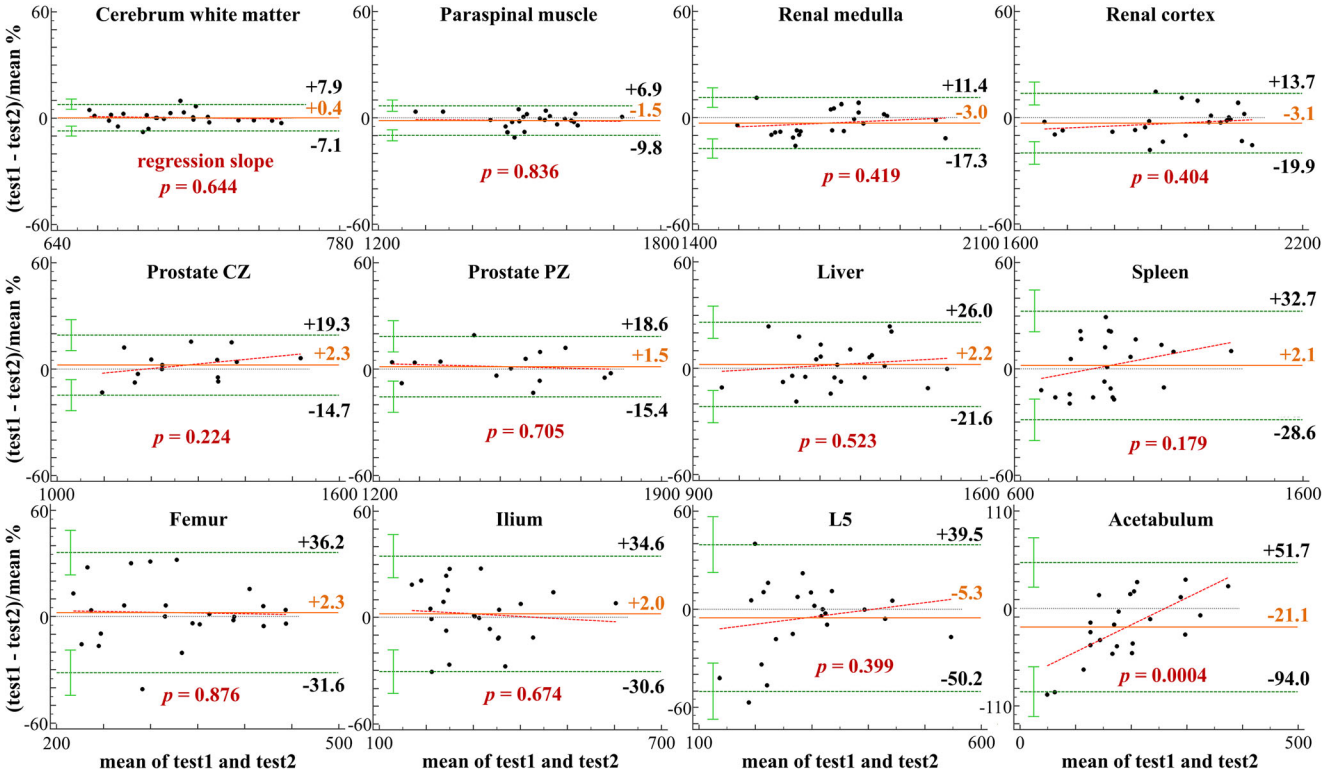

Fig 4 Bland-Altman plot for organs and bones reconstructed to assess the overall repeatability of ADC (reader 1, all centers combined). No association between difference and mean is observed, except for acetabulum (p^{slope} = 0.0004). LoA defines the interval within which 95% of differences between two measurements (when reader and MRI scanner are constant) are expected to lie. The narrower this interval, the better the repeatability. LoA, green dotted line; 95% CI on LoA, green whiskers; mean bias, orange solid line; and regression line differences, red dotted line

(< 5%) [22–24, 42]. These reference studies identified image and measurement variabilities associated with inconsistent equipment and protocols [30, 43], whereas our study used the same MRI scanners, protocol, and software. Notably, the systematic error (bias) had the largest variations, which may result from variability in preparation of the phantom.

Secondly, this study demonstrated that (i) ADC values in the organs of interest fall within the range of values reported in normal soft tissues [29], and (ii) R&R of ADC in WB-MRI agrees with published values. Even within a mobile organ such as the liver, where ADC variability is influenced by the effect of cardiac movement on DWI signal intensity [44, 45], similar R&R was found. Previous studies were performed at 1.5 or 3.0 T, used MRI scanners and coils from different vendors, used various processing algorithms to compute ADC, did not use a WB-MRI protocol, and often relied on multiple ROIs in the same organ to average the diffusion by "smoothing" the variations associated to repeated measurements [46]. In contrast, we used the same 3.0-T MRI scanners, same algorithms, a WB-MRI protocol, and a single ROI per organ, which is the routine clinical method for assessing a focal lesion.

Thirdly, results from the ANOVA showed that R&R of ADC was not influenced by the site of acquisition of DW images or by the choice of the MRI reader. Given that the image post-processing was centralized, the potential sources of variation in ADC were as follows: (1) capability of the scanner to measure the diffusion in vivo accurately; (2) imperfect anatomical matching of MRI volumes owing to lack of co-registration algorithm; (3) manual definition of the ROI owing to lack of automated segmentation; and (4) possible effects of physiologic changes during the time between scans. Variations from source (1) have been demonstrated as limited. No variation from source (4) has been observed. Variations from sources (2) and (3) may have been further reduced through more advanced post-processing of the data. Although promising solutions are studied [47, 48], the analysis of WB-DWI in clinical routine does not benefit either from a spatial co-registration of MRI examinations that may ensure a better anatomical matching of the studied regions or from an automated segmentation of these regions. Solving these two problems would allow targeting smaller changes in ADC measurement. We are currently working to solve this issue, in parallel with the development of a semi-automated segmentation of the total disease volume in bone tissue [49].

The assessment of repeatability and reproducibility coefficients with their 95% CI (instead of the CoV ensuring a level of confidence of 68% only) in imaging using quantitative

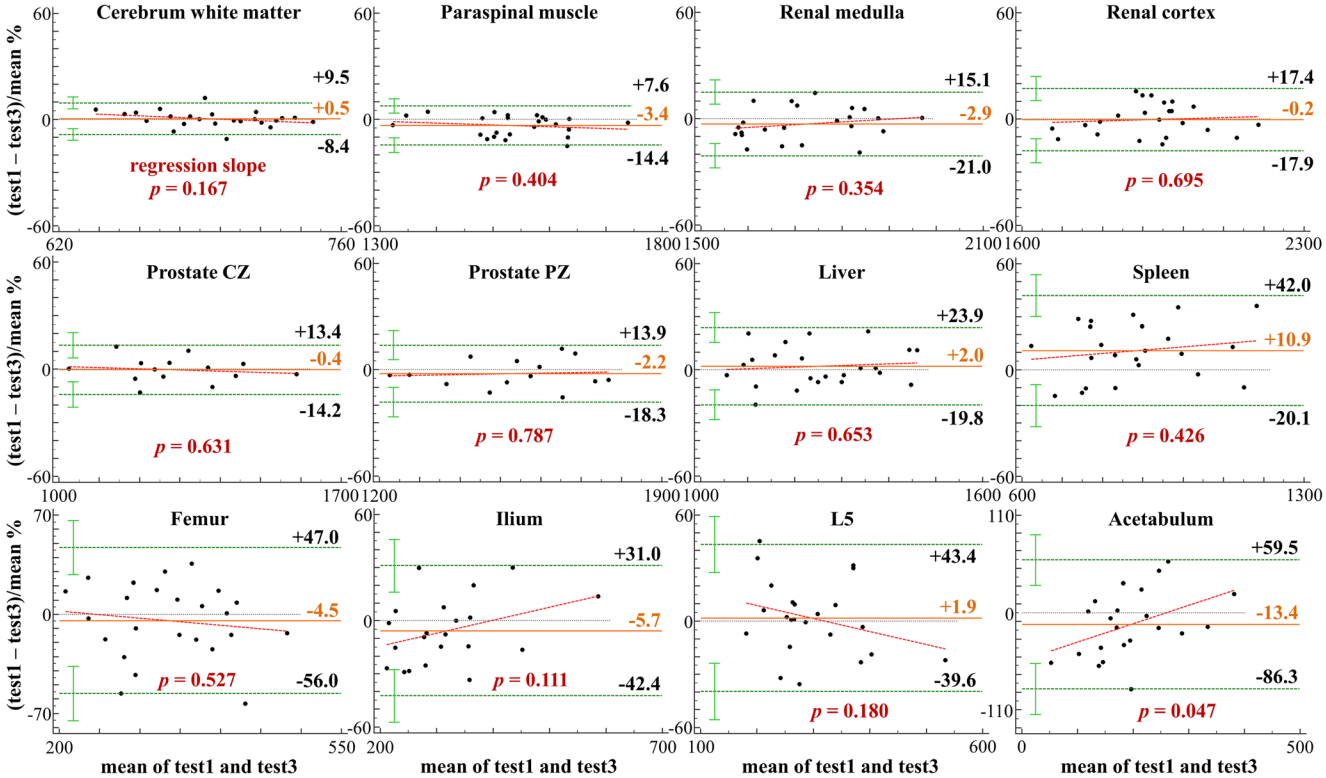

Fig. 5. Bland-Altman plot for organs and bones reconstructed to assess the overall reproducibility of ADC (reader 1, for all pairs of centers: C1 vs. C2, C1 vs. C3, and C2 vs. C3). No association between difference and mean is observed. LoA defines the interval within which 95% of differences between two measurements (when reader is constant but MRI scanner is not) are expected to lie. The narrower this interval, the better the reproducibility. LoA, green dotted line; 95% CI on LoA, green whiskers; mean bias, orange solid line; and regression line differences, red dotted line

endpoints is recommended in reference papers [36, 38]. The upper limits of the 95% CI associated with RC and RDC can be used with confidence as initial references for detecting microstructure changes when a patient undergoes a standardized WB-DWI protocol. However, two important considerations should be kept in mind. Firstly, a R&R study remains mandatory to assess the exact variability in diffusion measurements specific to a protocol in situ, and to be sure that the qualifications for a confident interpretation of QIB are met [36]. The variability observed may be interestingly compared to that observed in reference papers for validation purpose, and to assess the contribution of (hardware/software) factors influencing the ADC. Secondly, determining whether measurements from a protocol are repeatable and reproducible depends on the context in which the measurements are used. Large ADC variations observed in bones are associated with small absolute values of the diffusion coefficient, which may be usable in clinical imaging if disease progression and/or the effects of treatment are expected to yield even larger variations. In this study, the reproducibility of diffusion measurements in bone was noticeably lower than that in other tissues. This is because the fat suppression technique eliminates most of the signal from the normal fatty marrow and because the presence of bone trabeculae responsible for susceptibility artifacts further decreases the signal intensity. The reported reproducibility of measurements in bone metastases is higher, because destruction of trabeculae and replacement of fatty marrow by solid tumoral tissue increase T1 and T2 relaxation times, resulting in higher signal-to-noise ratio and ADC value [50, 51].

The main limitation of our study is definitely that the ADC variability was assessed in healthy volunteers. Biological variability may result in differences in R&R evaluated in healthy volunteers as compared with individual patients (who, in addition, may demonstrate biological variability over time due to either the progression of the disease or its response to therapy [40, 52]). Assuming that an exact anatomical matching of ROIs between readings is achieved, R&R of ADC in tumors is expected to be in the same range or lower than that in normal soft tissues. ADC repeatability in liver metastasis larger than 2 cm was reported to be around 21% [53] and 22% [54], which is slightly lower than the RC = 24–25% observed in our study. In bone marrow, the lower R&R of ADC observed in volunteers may result from the low ADC values; smaller variations in ADC, higher R&R, and lower threshold variations are likely to be observed in tumors replacing the bone marrow, as suggested in the literature [50, 51], allowing

Table 3 Summary of parameters from the Bland-Altman plots for the R&R study of ADC from reader 1

Repeatability												
	Center	**Cerebrum WM**	**Muscle**	**Renal medulla**	**Renal cortex**	**Prostate CZ**	**Prostate PZ**	**Liver**	**Spleen**	**Femur**	**Ilium**	**L5**	**Acetabulum**
CoV	C1	2.2	3.4	9.1	9.3	—	—	14	17	8.2	17	20	27
	C3	3.6	5.7	5.1	3.4	—	—	7.9	15	15	14	20	44
	C2	5.0	2.1	8.1	10.8	—	—	11	14	19	15	9.0	20
	All	3.8	4.2	7.3	8.6	8.7	8.8	12	15	16	16	17	30
Bias	C1	+ 0.6	+ 0.2	− 1.5	+ 3.6	—	—	+ 4.5	− 1.2	+ 1.4	− 5.7	+ 4.2	− 4.4
	C3	+ 1.7	− 3.9	− 5.0	− 4.5	—	—	− 3.2	+ 3.4	11	+ 7.6	16	− 44
	C2	− 1.2	− 0.6	− 2.3	− 5.3	—	—	+ 1.6	− 1.8	− 6.1	+ 4.2	− 3.4	− 18
	All	+ 0.4	− 1.5	− 3.0	− 3.1	+ 2.3	+ 1.5	+ 2.2	+ 2.1	+ 2.3	+ 2.0	+ 5.3	− 21
LoA	C1	[− 3.7; + 5.0]	[− 6.4; + 6.7]	[− 19; + 16]	[− 18; + 25]	—	—	[− 32; + 41]	[− 48; + 45]	[− 16; + 19]	[− 44; + 32]	[− 48; + 56]	[− 60; + 51]
	C3	[− 5.3; + 8.8]	[− 15; + 7.3]	[− 15; + 4.9]	[− 12; + 2.7]	—	—	[− 18; + 12]	[− 29; + 35]	[− 21; + 44]	[− 19; + 34]	[− 37; + 70]	[− 146; + 58]
	C2	[− 11; + 8.6]	[− 5.0; + 3.7]	[− 18; + 14]	[− 26; + 15]	—	—	[− 21; + 24]	[− 30; + 26]	[− 47; + 35]	[− 26; + 35]	[− 22; + 15]	[− 55; + 20]
	All	[− 7.1; + 7.9]	[− 9.8; + 6.9]	[− 17; 11]	[− 20; + 14]	[− 15; + 19]	[− 15; + 19]	[− 22; + 26]	[− 29; + 33]	[− 32; + 36]	[− 31; + 35]	[−50; + 40]	[− 94; + 52]
RC	All	7.5 [5.3; 9.7]	8.2 [5.8; 11]	14 [10; 19]	17 [12; 22]	17 [11; 23]	17 [11; 24]	24 [17; 30]	30 [21; 39]	31 [22; 40]	31 [22; 40]	34 [24; 44]	59 [42; 76]
Reproducibility													
	Center	**Cerebrum WM**	**Muscle**	**Prostate CZ**	**Renal cortex**	**Prostate PZ**	**Renal medulla**	**Liver**	**Spleen**	**Ilium**	**L5**	**Femur**	**Acetabulum**
CoV	C1 vs. C3	4.8	7.1	—	10	—	12	12	12	13	21	32	27
	C1 vs. C2	2.0	4.1	—	8.9	—	8.5	9.6	16	18	22	20	43
	C3 vs. C2	6.2	6.2	—	8.8	—	8.2	12	16	17	22	26	45
	All	4.6	5.7	7.1	8.9	8.5	9.2	11	15	20	21	27	37
Bias	C1 vs. C3	+ 1.7	− 3.0	—	− 1.3	—	− 3.8	− 1.3	− 2.2	− 25	+ 5.9	− 14	− 25
	C1 vs. C2	+ 0.0	− 2.6	—	− 0.7	—	− 2.8	+ 3.0	12	15	− 3.4	+ 2.9	− 0.5
	C3 vs. C2	− 0.2	− 4.6	—	− 1.4	—	− 2.4	+ 0.4	17	− 5.5	− 4.4	− 2.3	− 39
	All	+ 0.5	− 3.4	− 0.4	− 0.2	− 2.2	− 2.9	+ 2.0	11	− 5.7	− 1.9	− 4.5	− 13
LoA	C1 vs. C3	[− 7.7; + 11]	[− 16; + 10]	—	[− 26; + 23]	—	[− 28; + 20]	[− 25; + 22]	[− 39; + 35]	[− 62; + 11]	[− 43; + 55]	[− 77; + 49]	[− 77; + 25]
	C1 vs. C2	[− 4.0; + 4.0]	[− 11; + 5.4]	—	[− 18; + 17]	—	[− 19; + 14]	[− 16; + 22]	[− 21; + 45]	[− 32; + 62]	[− 41; + 34]	[− 38; + 44]	[− 84; + 83]
	C3 vs. C2	[− 12; + 12]	[− 17; + 7.8]	—	[− 19; + 16]	—	[− 19; + 14]	[− 32; + 32]	[− 14; + 48]	[− 37; + 26]	[− 66; + 57]	[− 51; + 47]	[− 187; + 109]
	All	[− 8.4; + 9.5]	[− 14; + 7.6]	[− 14; + 13]	[− 18; + 17]	[− 18; + 14]	[− 21; + 15]	[− 20; + 24]	[− 20; + 42]	[− 42; + 31]	[− 40; + 43]	[− 56; + 47]	[− 86; + 60]
RDC	All	8.6 [6.4; 12]	11 [7.9; 15]	14 [8.6; 19]	17 [12; 22]	17 [10; 23]	18 [13; 23]	22 [15; 28]	30 [21; 39]	39 [27; 50]	42 [30; 54]	52 [37; 68]	73 [51; 94]

WM, white matter; *CZ*, central zone; *PZ*, peripheral zone. CoV (coefficient of variation), bias, LoA (limits of agreement), RC (repeatability coefficient), and RDC (reproducibility coefficient) are all expressed as percentages. LoA and 95% CI are given as intervals in brackets. Because of the size of the cohort (only 4 male volunteers per center), R&R in the prostate was assessed for all centers combined. Organs were sorted according to increasing values of the RC and RDC coefficients

Table 4 Summary of parameters from the Bland-Altman plots for the R&R study of ADC from reader 2

Repeatability												
	Center	**Cerebrum WM**	**Muscle**	**Renal cortex**	**Prostate CZ**	**Prostate PZ**	**Renal medulla**	**Liver**	**L5**	**Spleen**	**Ilium**	**Femur**	**Acetabulum**
CoV	C1	2.9	4.1	6.4	-	-	7.5	14	16	21	19	20	25
	C3	4.1	5.8	5.9	-	-	13	11	23	17	12	23	25
	C2	4.6	4.4	4.3	-	-	11	13	7.9	13	20	22	25
	All	3.9	4.8	5.6	6.2	9.6	11	13	17	18	18	22	26
Bias	C1	− 0.1	+ 0.4	+ 1.7	-	-	+ 2.3	+ 4.6	− 5.4	− 9.0	+ 2.5	+ 1.3	+ 9.6
	C3	+ 1.1	− 1.2	− 2.3	-	-	+ 3.3	− 3.4	+ 8.5	− 0.3	+ 2.8	− 6.4	+ 8.8
	C2	− 0.8	− 0.5	− 0.1	-	-	− 0.2	− 3.5	− 3.2	− 14	+ 0.4	− 1.1	− 12
	All	+ 0.1	− 0.4	− 0.3	− 0.0	+ 1.0	+ 2.9	− 0.6	− 0.3	− 7.5	+ 1.9	− 2.1	+ 2.6
LoA	C1	[− 5.9; + 5.6]	[− 7.8; + 8.6]	[− 11; + 14]	-	-	[− 14; + 19]	[− 24; + 34]	[− 44; + 35]	[− 50; + 32]	[− 44; + 49]	[− 44; + 47]	[− 56; + 75]
	C3	[− 7.0; + 9.2]	[− 9.9; + 7.5]	[− 14; + 9.3]	-	-	[− 20; + 26]	[− 29; + 22]	[− 4.2; + 21]	[− 23; + 23]	[− 38; + 43]	[− 51; + 38]	[− 42; + 60]
	C2	[− 9.8; + 8.2]	[− 12; + 11]	[− 8.5; + 8.2]	-	-	[− 27; + 27]	[− 26; + 20]	[− 39; + 32]	[− 50; + 21]	[− 41; + 42]	[− 52; + 50]	[− 64; + 39]
	All	[− 7.5; + 7.6]	[− 9.7; + 8.8]	[− 11; + 11]	[− 13; + 13]	[− 18; + 20]	[− 20; + 25]	[− 27; + 25]	[− 34; + 32]	[− 42; + 27]	[− 40; + 44]	[− 48; + 44]	[− 56; + 61]
RC	All	7.6 [5.4; 9.7]	9.4 [6.6; 12]	11 [7.8; 14]	12 [7.5; 17]	19 [12; 26]	23 [16; 29]	25 [18; 32]	33 [23; 42]	35 [25; 45]	36 [26; 46]	43 [30; 55]	52 [37; 67]
Reproducibility													
	Center	**Cerebrum WM**	**Renal cortex**	**Muscle**	**Prostate PZ**	**Renal medulla**	**Prostate CZ**	**Liver**	**Spleen**	**Ilium**	**L5**	**Femur**	**Acetabulum**
CoV	C1 vs. C3	5.0	8.9	9.1	—	6.2	—	16	17	21	28	25	31
	C1 vs. C2	2.5	3.1	4.2	—	7.1	—	8.6	18	14	30	25	26
	C3 vs. C2	4.8	6.1	5.1	—	8.3	—	14	15	22	17	29	32
	All	4.3	6.3	6.3	6.7	7.1	9.4	14	17	19	26	27	30
Bias	C1 vs. C3	+ 0.9	− 0.5	+ 0.7	—	− 0.3	—	− 5.4	− 9.3	− 3.9	− 3.6	− 23	− 15
	C1 vs. C2	− 1.2	− 0.9	− 0.1	—	− 1.0	—	+ 7.4	+ 2.5	+ 6.2	− 4.4	− 2.5	+ 5.3
	C3 vs. C2	+ 1.6	− 3.4	− 0.0	—	+ 1.9	—	− 4.6	− 8.1	+ 7.9	12	− 22	+ 7.9
	All	+ 0.4	− 1.6	+ 0.2	− 0.4	+ 0.2	− 5.8	− 0.5	− 4.5	+ 3.4	+ 0.4	− 15*	− 0.9
LoA	C1 vs. C3	[− 9.0; + 11]	[18; + 17]	[− 18; + 19]	—	[− 12; + 12]	—	[− 41; + 30]	[− 39; + 20]	[− 56; + 48]	[− 88; + 81]	[− 79; + 33]	[− 91; + 62]
	C1 vs. C2	[− 6.1; + 3.6]	[− 6.8; + 5.0]	[− 8.6; + 8.5]	—	[− 15; + 13]	—	[− 8.2; + 23]	[− 29; + 34]	[− 30; + 42]	[− 69; + 60]	[− 42; + 37]	[− 44; + 54]
	C3 vs. C2	[− 7.9; + 11]	[− 15; + 8.3]	[− 10; + 10]	—	[− 14; + 18]	—	[− 30; + 21]	[− 45; + 29]	[− 36; + 52]	[− 10; + 34]	[− 91; + 47]	[− 71; + 87]
	All	[− 8.0; + 8.8]	[− 14; + 11]	[− 13; + 13]	[− 14; + 13]	[− 14; + 14]	[− 23; + 11]	[− 29; + 28]	[− 37; + 28]	[− 41; + 48]	[− 64; + 64]	[− 73; + 42]	[− 71; + 69]
RDC	All	8.4 [6.0; 11]	12 [8.7; 16]	12 [8.8; 16]	13 [8.1; 18]	14 [10; 18]	19 [12; 26]	27 [19; 35]	34 [24; 44]	37 [26; 48]	51 [36; 66]	52 [37; 67]	58 [41; 75]

WM, white matter; *CZ*, central zone; *PZ*, peripheral zone. CoV (coefficient of variation), bias, LoA (limits of agreement), RC (repeatability coefficient), and RDC (reproducibility coefficient) are all expressed as percentages. LoA and 95% CI are given as intervals in brackets. Because of the size of the cohort (only 4 male volunteers per center), R&R in the prostate was assessed for all centers combined. Organs were sorted according to increasing values of the RC and RDC coefficients.

*Bias significantly different from 0 ($p < 0.0031$)

for clinical detection of significant longitudinal changes in patients. Of note, a similar trend with a lower R&R in volunteers (smallest detectable change in ADC in sacroiliac joint = 48%) compared to that in patients with axial spondyloarthritis (smallest detectable change = 22%) has been reported [55]. Another limitation comes from the fact that, due to practical reasons related to MRI examination scheduling, the delay between test 2 and test 3 ranged between the same day and 7 days, while test 1 and test 2 were performed on the same day. This delay between imaging sessions remains short and is in a range comparable with other repeatability studies, and RDC was not systematically superior to RC. However, we cannot rule out the possibility of a change in ADC in individual healthy volunteers during this delay.

In conclusion, we assessed the per-organ R&R limits above which a true change in ADC can be detected with confidence. The close agreement observed on ADC values and R&R values (in phantom and in volunteers) between this study and reference studies performed on other MRI scanners and the choice of a 95% confidence level to assess changes in ADC ensure the applicability of the results by promoters of multicenter WB-DWI studies. These R&R limits may be used as initial guests in oncological trial to detect disease or treatment-induced changes in tissue microstructure. This study also allows adopting a less conservative criterion for detecting a true change in ADC, based on R&R values estimated at a 68% confidence level which can be easily derived from the CoV coefficients provided here. These different criteria may be tested in clinical setting to determine which one is the more suitable in oncological trial. Overall, due to the intrinsic dependence of ADC values on biophysical tissue properties, further investigations in patients are recommended to refine the estimates of the R&R limits and provide a confident interpretation of longitudinal changes measured in WB-DWI.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-020-07522-0.

Funding This research was supported by Innoviris (Institut pour l'encouragement de la recherche scientifique et de l'innovation de la région Bruxelles-Capitale, Brussels, Belgium; Grant 2013-PFS-EH-7).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Prof. Frédéric Lecouvet (Université Catholique de Louvain, Radiology Department, Brussels, Belgium).

Conflict of interest The 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 Two of the authors have significant statistical expertise (Laurence Collette, PhD, EORTC; Nicolas Michoux, PhD, UCL).

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

Ethical approval Institutional Review Board approval was obtained.

Methodology
- prospective
- observational
- multicenter study

References

1. Vilanova JC, García-Figueiras R, Luna A, Baleato-González S, Tomás X, Narváez JA (2019) Update on whole-body MRI in musculoskeletal applications. Semin Musculoskelet Radiol 23:312–323
2. Kalus S, Saifuddin A (2019) Whole-body MRI vs bone scintigraphy in the staging of Ewing sarcoma of bone: a 12-year single-institution review. Eur Radiol 29:5700–5570
3. Lecouvet FE, Van Nieuwenhove S, Jamar F, Lhommel R, Guermazi A, Pasoglou VP (2018) Whole-Body MR imaging: the novel, "intrinsically hybrid," approach to metastases, myeloma, lymphoma, in bones and beyond. PET Clin 13:505–522
4. Pasoglou V, Michoux N, Larbi A, Van Nieuwenhove S, Lecouvet F (2018) Whole body MRI and oncology: recent major advances. Br J Radiol 91:20170664
5. Park HY, Kim KW, Yoon MA et al (2020) Role of whole-body MRI for treatment response assessment in multiple myeloma: comparison between clinical response and imaging response. Cancer Imaging 20:14
6. Latifoltojar A, Punwani S, Lopes A et al (2019) Whole-body MRI for staging and interim response monitoring in paediatric and adolescent Hodgkin's lymphoma: a comparison with multi-modality reference standard including 18F-FDG-PET-CT. Eur Radiol 29:202–212
7. Winfield JM, Poillucci G, Blackledge MD et al (2018) Apparent diffusion coefficient of vertebral haemangiomas allows differentiation from malignant focal deposits in whole-body diffusion-weighted MRI. Eur Radiol 28:1687–1691
8. Machado Medeiros T, Altmayer S, Watte G (2020) 18F-FDG PET/CT and whole-body MRI diagnostic performance in M staging for non-small cell lung cancer: a systematic review and meta-analysis. Eur Radiol. https://doi.org/10.1007/s00330-020-06703-1
9. Han SN, Amant F, Michielsen K (2018) Feasibility of whole-body diffusion-weighted MRI for detection of primary tumour, nodal and distant metastases in women with cancer during pregnancy: a pilot study. Eur Radiol 28:1862–1874
10. Tordjman M, Mali R, Madelin G et al (2020) Diagnostic test accuracy of ADC values for identification of clear cell renal cell carcinoma: systematic review and meta-analysis. Eur Radiol. https://doi.org/10.1007/s00330-020-06740-w
11. Johnston EW, Latifoltojar A, Sidhu HS et al (2019) Multiparametric whole-body 3.0-T MRI in newly diagnosed intermediate- and high-risk prostate cancer: diagnostic accuracy and interobserver agreement for nodal and metastatic staging. Eur Radiol 29:3159–3169

12. Larbi A, Omoumi P, Pasoglou V et al (2019) Whole-body MRI to assess bone involvement in prostate cancer and multiple myeloma: comparison of the diagnostic accuracies of the T1, short tau inversion recovery (STIR), and high B-values diffusion-weighted imaging (DWI) sequences. Eur Radiol 29:4503–4513
13. Lecouvet F, Vander Maren N, Collette L et al (2019) Whole body MRI in spondyloarthritis (SpA): preliminary results suggest that DWI outperforms STIR for lesion detection. Eur Radiol 28:4163–4173
14. Medeiros TM, Altmayer S, Guilherme Watte G et al (2020) 18F-FDG PET/CT and Whole-body MRI diagnostic performance in M staging for non-small cell lung cancer: a systematic review and meta-analysis. Eur Radiol 30:3641–3649
15. Kharuzhyk S, Zhavrid E, Dziuban A, Sukolinskaja E, Kalenik O (2020) Comparison of whole-body MRI with diffusion-weighted imaging and PET/CT in lymphoma staging. Eur Radiol 30:3915–3923
16. Lai AYT, Angela Riddell A, Tara Barwick T et al (2020) Interobserver agreement of whole-body magnetic resonance imaging is superior to whole-body computed tomography for assessing disease burden in patients with multiple myeloma. Eur Radiol 30:320–327
17. Donners R, Blackledge M, Tunariu N, Messiou C, Merkle EM, Koh DM (2018) Quantitative whole-body diffusion-weighted MR imaging. Magn Reson Imaging Clin N Am 26:479–494
18. Schmeel FC (2019) Variability in quantitative diffusion-weighted MR imaging (DWI) across different scanners and imaging sites: is there a potential consensus that can help reducing the limits of expected bias? Eur Radiol 29:2243–2245
19. Padhani AR, Makris A, Gall P, Collins DJ, Tunariu N, de Bono JS (2014) Therapy monitoring of skeletal metastases with whole-body diffusion MRI. J Magn Reson Imaging 39:1049–1078
20. Petralia G, Padhani AR, Pricolo P et al (2019) Whole-body magnetic resonance imaging (WB-MRI) in oncology: recommendations and key uses. Radiol Med 124:218–233
21. Sasaki M, Yamada K, Watanabe Y et al (2008) Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology 249:624–630
22. Chenevert TL, Galban 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
23. 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
24. Doblas S, Almeida GS, Ble FX et al (2015) Apparent diffusion coefficient is highly reproducible on preclinical imaging systems: evidence from a seven-center multivendor study. J Magn Reson Imaging 42:1759–1764
25. Winfield JM, Tunariu N, Rata M et al (2017) Extracranial soft-tissue tumors: repeatability of apparent diffusion coefficient estimates from diffusion-weighted MR imaging. Radiology 284:88–99
26. Donati OF, Chong D, Nanz D et al (2014) Diffusion-weighted MR imaging of upper abdominal organs: field strength and intervendor variability of apparent diffusion coefficients. Radiology 270:454–463
27. Fedeli L, Belli G, Ciccarone A et al (2018) Dependence of apparent diffusion coefficient measurement on diffusion gradient direction and spatial position - a quality assurance intercomparison study of forty-four scanners for quantitative diffusion-weighted imaging. Phys Med 55:135–141
28. Barbieri S, Donati OF, Froehlich JM, Thoeny HC (2016) Comparison of intravoxel incoherent motion parameters across MR imagers and field strengths: evaluation in upper abdominal organs. Radiology 279:784–794
29. Jafar MM, Parsai A, Miquel ME (2016) Diffusion-weighted magnetic resonance imaging in cancer: reported apparent diffusion coefficients, in-vitro and in-vivo reproducibility. World J Radiol 8:21–49
30. Malyarenko D, Fedorov A, Bell L et al (2018) Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging (Bellingham) 5:011006
31. Ghosh A, Singh T, Singla V, Bagga R, Khandelwal N (2017) Comparison of absolute apparent diffusion coefficient (ADC) values in ADC maps generated across different postprocessing software: reproducibility in endometrial carcinoma. AJR Am J Roentgenol 209:1312–1320
32. Zeilinger MG, Lell M, Baltzer PA, Dorfler A, Uder M, Dietzel M (2017) Impact of post-processing methods on apparent diffusion coefficient values. Eur Radiol 27:946–955
33. Barnes A, Alonzi R, Blackledge M et al (2018) UK quantitative WB-DWI technical workgroup: consensus meeting recommendations on optimisation, quality control, processing and analysis of quantitative whole-body diffusion-weighted imaging for cancer. Br J Radiol 91:20170577
34. deSouza NM, Winfield JM, Waterton JC et al (2018) Implementing diffusion-weighted MRI for body imaging in prospective multicentre trials: current considerations and future perspectives. Eur Radiol 28:1118–1131
35. Raunig DL, McShane LM, Pennello G et al (2015) Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 24:27–67
36. QIBA Profile: diffusion-weighted magnetic resonance imaging (DWI) (2017) Available via https://qibawiki.rsna.org/index.php/DWI_Profile_Development_Archive
37. Pasoglou V, Michoux N, Peeters F et al (2015) Whole-body 3D T1-weighted MR imaging in patients with prostate cancer: feasibility and evaluation in screening for metastatic disease. Radiology 275:155–166
38. Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 31:466–475
39. Messiou C, Hillengass J, Delorme S et al (2019) Guidelines for acquisition, interpretation, and reporting of whole-body MRI in myeloma: Myeloma Response Assessment and Diagnosis System (MY-RADS). Radiology 291:5–13
40. Padhani AR, Lecouvet FE, Tunariu N et al (2017) METastasis Reporting and Data System for Prostate Cancer: practical guidelines for acquisition, interpretation, and reporting of whole-body magnetic resonance imaging-based evaluations of multiorgan involvement in advanced prostate cancer. Eur Urol 71(1):81–92
41. Ellingson BM, Bendszus M, Boxerman J et al (2015) Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. Neuro Oncol 17:1188–1198
42. Malyarenko D, Galban 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
43. Chenevert TL, Malyarenko DI, Newitt D et al (2014) Errors in quantitative image analysis due to platform-dependent image scaling. Transl Oncol 7:65–71
44. Braithwaite AC, Dale BM, Boll DT, Merkle EM (2009) Short- and midterm reproducibility of apparent diffusion coefficient measure-

ments at 3.0-T diffusion-weighted imaging of the abdomen. Radiology 250:459–465
45. Metens T, Absil J, Denolin V, Bali MA, Matos C (2016) Liver apparent diffusion coefficient repeatability with individually predetermined optimal cardiac timing and artifact elimination by signal filtering. J Magn Reson Imaging 43:1100–1110
46. Colagrande S, Pasquinelli F, Mazzoni LN, Belli G, Virgili G (2010) MR-diffusion weighted imaging of healthy liver parenchyma: repeatability and reproducibility of apparent diffusion coefficient measurement. J Magn Reson Imaging 31:912–920
47. Blackledge MD, Collins DJ, Tunariu N et al (2014) 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 9:e91779
48. Dzyubachyk O, Lelieveldt BP, Blaas J, Reijnierse M, Webb A, van der Geest RJ (2013) Automated algorithm for reconstruction of the complete spine from multistation 7T MR data. Magn Reson Med 69:1777–1786
49. Ceranka J, Polfliet M, Lecouvet F, Michoux N, de Mey J, Vandemeulebroucke J (2018) Registration strategies for multi-modal whole-body MRI mosaicing. Magn Reson Med 79:1684–1695
50. Blackledge MD, Tunariu N, Orton MR et al (2016) Inter- and intra-observer repeatability of quantitative whole-body, diffusion-weighted imaging (WBDWI) in metastatic bone disease. PLoS One 11:e0153840
51. Padhani AR, van Ree K, Collins DJ, D'Sa S, Makris A (2013) Assessing the relation between bone marrow signal intensity and apparent diffusion coefficient in diffusion-weighted MRI. AJR Am J Roentgenol 200:163–170
52. García-Figueiras R, Baleato-González S, Padhani AR et al (2019) How clinical imaging can assess cancer biology. Insights Imaging 10:28
53. Pathak R, Ragheb H, Thacker NA (2017) A data-driven statistical model that estimates measurement uncertainty improves interpretation of ADC reproducibility: a multi-site study of liver metastases. Sci Rep 7:14084
54. Deckers F, De Foer B, Van Mieghem F (2014) Apparent diffusion coefficient measurements as very early predictive markers of response to chemotherapy in hepatic metastasis: a preliminary investigation of reproducibility and diagnostic value. J Magn Reson Imaging 40:448–456
55. Møller JM, Østergaard M, Thomsen HS, Sørensen IJ, Madsen OR, Pedersen SJ (2020) Test-retest repeatability of the apparent diffusion coefficient in sacroiliac joint MRI in patients with axial spondyloarthritis and healthy individuals. Acta Radiol Open 9: 2058460120906015
56. Thoeny HC, De Keyzer F, Oyen RH, Peeters RR (2005) Diffusion-weighted MR imaging of kidneys in healthy volunteers and patients with parenchymal diseases: initial experience. Radiology 235:911–917
57. Gibbs P, Pickles MD, Turnbull LW (2007) Repeatability of echo-planar-based diffusion measurements of the human prostate at 3 T. Magn Reson Imaging 25:1423–1429
58. Jacobs MA, Macura KJ, Zaheer A et al (2018) Multiparametric whole-body MRI with diffusion-weighted imaging and ADC mapping for the identification of visceral and osseous metastases from solid tumors. Acad Radiol 25:1405–1414
59. Lavdas I, Rockall AG, Castelli F et al (2015) Apparent diffusion coefficient of normal abdominal organs and bone marrow from whole-body DWI at 1.5 T: the effect of sex and age. AJR Am J Roentgenol 205:242–250
60. Messiou C, Collins DJ, Morgan VA, Desouza NM (2011) Optimising diffusion weighted MRI for imaging metastatic and myeloma bone disease and assessing reproducibility. Eur Radiol 21:1713–1718
61. Grech-Sollars M, Hales PW, Miyazaki K et al (2015) Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. NMR Biomed 28:468–485
62. Bilgili MY (2012) Reproducibility of apparent diffusion coefficients measurements in diffusion-weighted MRI of the abdomen with different b values. Eur J Radiol 81:2066–2068
63. Miquel ME, Scott AD, Macdougall ND, Boubertakh R, Bharwani N, Rockall AG (2012) In vitro and in vivo repeatability of abdominal diffusion-weighted MRI. Br J Radiol 85:1507–1512