Title

Extracranial Soft-Tissue Tumors: Repeatability of Apparent Diffusion Coefficient Estimates from Diffusion-weighted MR Imaging.

Permalink

https://escholarship.org/uc/item/80n5m02t

Journal

Radiology, 284(1)

Authors

Winfield, Jessica
Tunariu, Nina
Rata, Mihaela
et al.

Publication Date

2017-07-01

DOI

10.1148/radiol.2017161965

Peer reviewed

Published in final edited form as:
Radiology. 2017 July ; 284(1): 88–99. doi:10.1148/radiol.2017161965.

Extracranial Soft-Tissue Tumors: Repeatability of Apparent Diffusion Coefficient Estimates from Diffusion-weighted MR Imaging

Jessica M Winfield, PhD[1,2], **Nina Tunariu, MD**[1,2], **Mihaela Rata, PhD**[1,2], **Keiko Miyazaki, PhD**[1,2], **Neil P Jerome, PhD**[1,2], **Michael Germuska, PhD**[1,2,a], **Matthew D Blackledge, PhD**[1,2], **David J Collins, BA**[1,2], **Johann S de Bono, MD, PhD**[3,4], **Timothy A Yap, MD, PhD**[3,4], **Nandita M deSouza, MD**[1,2], **Simon J Doran, PhD**[1,2], **Dow-Mu Koh, MD**[1,2], **Martin O Leach, PhD**[1,2], **Christina Messiou, MD**[1,2], and **Matthew R Orton, PhD**[1,2]

[1]Cancer Research UK Cancer Imaging Centre, Division of Radiotherapy and Imaging, The Institute of Cancer Research and Royal Marsden Hospital, 123 Old Brompton Road, London. SW7 3RP. UK

[2]MRI Unit, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey. SM2 5PT. UK

[3]Drug Development Unit, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey. SM2 5PT. UK

[4]Division of Clinical Studies, The Institute of Cancer Research, 123 Old Brompton Road, London. SW7 3RP. UK

Abstract

Purpose—To assess repeatability of apparent diffusion coefficient (ADC) estimates in extracranial soft-tissue diffusion-weighted magnetic resonance imaging across a wide range of imaging protocols and patient populations.

Materials and methods—Nine prospective patient studies and one prospective volunteer study, conducted between 2006 and 2016, with research ethics committee approval and written informed consent from each subject, were included in this single-institution study. A total of 141 tumors/healthy organs were imaged twice (interval between repeated examinations ranged from 45 minutes to 10 days, depending on study) to assess repeatability of median and mean ADC estimates. Levene's test was used to determine whether ADC repeatability differed between studies. Pearson's linear correlation coefficient was used to assess correlation between coefficient of variation (CoV) and the year the study started, study size, and volumes of tumors/healthy organs. Repeatability of small, medium, and large tumors/healthy organs was assessed irrespective

Address for correspondence: Professor Martin Leach, MRI Unit, Royal Marsden Hospital, Downs Road, Sutton, Surrey. SM2 5PT, UK, martin.leach@icr.ac.uk, Telephone: +44 208 661 3338, Fax: +44 20 8661 0846.
[a]Current address: Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Maindy Road, Cardiff. CF24 4HQ. UK

This paper has not been presented at an RSNA meeting nor been accepted for presentation at a future meeting.

of study and Levene's test used to determine whether ADC repeatability differed between these groups.

Results—CoV aggregated across all studies was 4.1% (range for each study 1.7% to 6.5%). No correlation was observed between CoV and the year the study started or study size. CoV was weakly correlated with volume (r=-0.5, p=0.1). Repeatability was significantly different between small, medium and large tumors (p<0.05), with the lowest CoV (2.6%) for large tumors. There was a significant difference in repeatability between studies, which did not persist after excluding the study with the largest tumors.

Conclusion—ADC is a robust imaging metric with excellent repeatability in extra-cranial soft-tissues across a wide range of tumor sites, sizes, patient populations, and imaging protocol variations.

Introduction

Body diffusion-weighted magnetic resonance imaging (DW-MRI) is well-established as a qualitative and quantitative technique in oncology (1). The simplest quantitative metric derived from DW-MRI is the apparent diffusion coefficient (ADC), which is estimated by fitting a mono-exponential curve to the measured signal at two or more diffusion weightings (b-values). Baseline ADC estimates or post-treatment changes in ADC have been shown to be indicative of response to chemotherapy and/or chemoradiation therapy in many tumor types, including rectal adenocarcinoma (2), hepatic metastases of colorectal (3) and gastric cancers (4), cervical cancer (5), breast cancer (6), head-and-neck squamous cell carcinoma (7), ovarian cancer (8), and non-small cell lung cancer (9).

As with all quantitative metrics, the repeatability of ADC estimates determines the ability of the technique to detect treatment-induced changes, thereby influencing the number of patients required for clinical trials and determining the size of post-treatment changes that can be detected in individual patients. Repeatability is usefully defined as "closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement" (10) where, in imaging studies, repeatability conditions include use of the same scanner, imaging protocol, observers, and repetition after a short interval (typically 1 hour to 7 days). In DW-MRI studies that report ADC estimates, the "measurand" is usually the mean or median of ADC estimates from voxels in a tumor. On the other hand, *reproducibility* may be defined as "closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement" (10) e.g. using a different MR scanner. The inter-scanner reproducibility of ADC estimates is particularly important in multi-center studies where it has been shown that good quality diffusion-weighted images with reproducible ADC estimates across platforms can be obtained following careful optimization of imaging protocols (11).

Exploratory DW-MRI studies in clinical trials often incorporate ADC repeatability estimates, usually by obtaining two baseline examinations with the second examination during the same visit (so-called 'coffee-break' repeatability study) or at a second visit one or more days later. The requirement for two baseline examinations increases the burden on patients, which may reduce recruitment or retention rates, and requires additional scanner time and resources, which may be difficult to accommodate in busy radiology departments. It would be advantageous to estimate ADC repeatability from previous studies, but this would only be feasible if repeatability was broadly the same across studies, despite variations in imaging protocol, tumor type or patient cohort; large differences in repeatability would argue strongly for study-specific repeatability estimates. The variety of repeatability metrics reported in the literature hinders comparison between studies and a framework for assessment of the technical performance of quantitative imaging biomarkers has been proposed by the Radiological Society of North America (RSNA) Quantitative Imaging Biomarkers Alliance (QIBA) (12,13). The QIBA framework recommends reporting repeatability using the within-subject standard deviation, limits of agreement, repeatability coefficient, intraclass correlation coefficient, and within-subject coefficient of variance; QIBA also emphasise the importance of reporting measurement conditions. A detailed investigation of ADC repeatability across a wide range of studies using the QIBA framework is therefore desirable.

The aim of this study was to assess ADC repeatability using the framework proposed by QIBA in extra-cranial soft-tissue DW-MRI studies to investigate whether ADC repeatability differs between studies carried out using different imaging protocols and patient populations over a period of 10 years at a single institution.

Materials and Methods

Study population

Nine patient studies and one healthy volunteer study were included in this analysis. All studies were approved by relevant National Research Ethics Committees. All patients and volunteers gave their written consent to participate in the studies. Only repeatability data from double-baseline examinations are reported here; post-treatment changes were outside the scope of this study but have been reported in the literature for some studies (14–18).

Tables 1 and 2 describe the subjects and DW-MRI protocols for each study (labelled A to K); further information is available in the references given. All studies were carried out at 1.5T using Siemens MAGNETOM Avanto or Aera MR scanners (Table 2). In studies where the imaging study or ADC repeatability study formed a subset of the total cohort, only patients contributing to the ADC repeatability results are reported (studies C and G). In multi-center studies, only data from our center are reported (studies D, E, and K). In studies including intra-cranial and extra-cranial tumors, only extra-cranial data are reported (studies A and F). One result (coefficient of variation of ADC_{median} in study K) has been reported previously (11) but other results from study K were not reported previously. No other results presented here have been reported previously, as publications from the original studies included data from intra-cranial tumors (14,15,19) or data from other centers (17), which are excluded from this analysis.

Image and data analysis

A total of 141 tumors/healthy organs were included in this analysis. All DW-MRI data were fitted using in-house software (Adept, The Institute of Cancer Research, London; or Matlab, Mathworks, Natick, MA). Regions of interest (ROIs) were drawn as described in Table 1. Software, methodology, and observers were fixed within each study; differences between studies reflect changes in technology and personnel (Table 1).

For each tumor/healthy organ, all fitted pixels in the ROIs were combined to create a volume of interest (VOI). Median and mean ADC (ADC$_{\text{median}}$ and ADC$_{\text{mean}}$) were estimated for each VOI. Bland-Altman plots of untransformed data show a tendency for differences between pairs of baseline measurements to scale with their ADC value (see Supplemental Material, Figure 5 [online]), in which case it is recommended (13,20) that repeatability (and changes due to treatment) be quantified using a proportional, i.e. ratio-based, measure so that the same measure applies across the range of ADCs encountered. This can most easily be achieved by using the natural logarithm of the data (12, 13, 20–22), and this was done for all statistical analyses in this study. A paired t-test was used to assess whether there was a significant difference between the first and second baseline measurements in each study.

Repeatability was assessed using the methods recommended by QIBA (13). The within-subject standard deviation (s_W) of the log-transformed ADC estimates was estimated according to Eq. 1, where d_{i} is the difference between two baseline estimates of log(ADC$_{\text{median}}$) or log(ADC$_{\text{mean}}$) for the i^{th} VOI, and N is the number of VOIs.

$$s_W = \sqrt{\frac{1}{2N}\sum_{i=1}^{N} d_i^2} \quad \text{Equation 1}$$

The within-subject coefficient of variation (CoV) (23), 95% limits of agreement (LoA), and repeatability coefficient (RC), which depend only on s_W, were estimated according to Eqs. 2, 3, and 4 respectively.

$$\text{CoV} = 100\,\% \times \sqrt{\exp\left(s_W^2\right) - 1} \quad \text{Equation 2}$$

$$\text{LoA} = 100\,\% \times \left[\exp\left(\pm 1.96\sqrt{2}s_W\right) - 1\right] \quad \text{Equation 3}$$

$$\text{RC} = 1.96\sqrt{2}s_W \quad \text{Equation 4}$$

The intra-class correlation coefficient (ICC) was estimated according to Eq. 5, where s_B is the between-subject standard deviation.

$$\mathrm{ICC} = \frac{s_B^2}{s_B^2 + s_W^2} \quad \text{Equation 5}$$

s_B was estimated as $s_B = \sqrt{(\mathrm{BMS{-}WMS})/K}$ where $\mathrm{BMS} = K\sum_{i=1}^{N}\left(\bar{Y}_i - \bar{Y}\right)^2/N$ is the between-subject mean squares, $\mathrm{WMS} = \sum_{i=1}^{N}\sum_{k=1}^{K}\left(Y_{ik} - \bar{Y}_i\right)^2/N(K-1)$ is the within-subject mean squares, K is the number of replications (K=2 for all studies in this analysis), Y_{ik} is the observed value of log(ADC$_{\mathrm{median}}$) or log(ADC$_{\mathrm{mean}}$) for the i^{th} VOI at the k^{th} replication, $\bar{Y}_i$ is the average over replications for the i^{th} VOI, and $\bar{Y}$ is the grand mean of log(ADC$_{\mathrm{median}}$) or log(ADC$_{\mathrm{mean}}$) over all observations (24).

The 95% confidence intervals (CI) for s_W were estimated as $\left(\sqrt{\mathrm{WMS} \times N/\mathrm{Inv} - \chi_N^2(0.975)}, \sqrt{\mathrm{WMS} \times N/\mathrm{Inv} - \chi_N^2(0.025)}\right)$, where $\mathrm{Inv}{-}\chi_N^2(p)$ is the p^{th} centile of the χ^2 distribution with N degrees of freedom (24).

95% CI for ICC were estimated as $\left(\frac{F_L - 1}{F_L + 1}, \frac{F_U - 1}{F_U + 1}\right)$, where $F_U = F_0$.Inv-$F_{N,N-1}$(0.975) and $F_L = F_0$/Inv-$F_{N-1,N}$(0.975), with F_0 = BMS/WMS, and Inv-$F_{d_1,d_2}(p)$ is the p^{th} centile of the F distribution with d_1 and d_2 degrees of freedom (25).

In addition to analysis of each study individually, VOIs were grouped into small, medium, and large, regardless of study (i.e. smallest 1/3, middle 1/3 and largest 1/3 of VOIs) and repeatability assessed for the three groups (47 VOIs per group). Finally, VOIs were aggregated from all studies and repeatability assessed for 141 VOIs together.

Levene’s test for homoscedasticity (LeveneAbsolute, vartestn, Matlab 2016a) was used to assess whether repeatability differed between studies (24). Baseline differences were calculated for each VOI for log(ADC$_{\mathrm{mean}}$) and log(ADC$_{\mathrm{median}}$) and Levene’s test used to assess whether the variance of the differences was the same for all studies; Levene’s test was also used to assess whether repeatability differed between small, medium, and large VOIs.

Pearson’s linear correlation coefficient (Matlab 2016a) was used to assess correlation between CoV and the year the study started, the number of VOIs in the study, and the median volume of VOIs in the study.

Results

The repeatability of ADC$_{\mathrm{mean}}$ was similar to the repeatability of ADC$_{\mathrm{median}}$ in all studies (Tables 3 and 4 [online]); for clarity, only ADC$_{\mathrm{median}}$ is shown in Figures 1 to 4. Bland-Altman plots showed no relationship between differences between pairs of baseline measurements and their means (Figure 1). None of the studies showed a significant difference between pairs of baseline measurements (paired t-test, p>0.05). The repeatability of ADC$_{\mathrm{median}}$ (Table 3) and ADC$_{\mathrm{mean}}$ (Table 4 [online]) was good with CoVs between 1.7% and 6.3% for ADC$_{\mathrm{median}}$ and between 1.7% and 6.5% for ADC$_{\mathrm{mean}}$ for all studies (Figure

2). Aggregating VOIs from all studies, CoV was 4.1% for $\mathrm{ADC}_{\mathrm{median}}$ and 3.9% for $\mathrm{ADC}_{\mathrm{mean}}$, with upper and lower 95% LoA of 12.1% and -10.8% respectively for $\mathrm{ADC}_{\mathrm{median}}$ and 11.5% and -10.3% for $\mathrm{ADC}_{\mathrm{mean}}$. Levene's test showed a significant difference between studies (p=0.01 for $\mathrm{ADC}_{\mathrm{median}}$ and $\mathrm{ADC}_{\mathrm{mean}}$), which did not persist after excluding the study with the lowest CoV (study B, which included some of the largest VOIs).

There was no correlation between the CoV and the year the studies started (Figure 3a, r=-0.4, p=0.2 for $\mathrm{ADC}_{\mathrm{median}}$; r=-0.3, p=0.3 for $\mathrm{ADC}_{\mathrm{mean}}$) nor between the CoV and the number of VOIs in each study (Figure 3b, r=-0.3, p=0.3 for $\mathrm{ADC}_{\mathrm{median}}$; r=-0.4, p=0.2 for $\mathrm{ADC}_{\mathrm{mean}}$). Only weak correlation was demonstrated between the CoV and the median VOI volume in each study (Figure 3c, r=-0.5, p=0.1 for $\mathrm{ADC}_{\mathrm{median}}$ and $\mathrm{ADC}_{\mathrm{mean}}$), although the CoV is noticeably lower in one study with very large tumors (study B) compared with other studies. Grouping into small, medium, and large VOIs showed a significant difference in ADC repeatability between sizes (Levene's test, p=0.02 for $\mathrm{ADC}_{\mathrm{median}}$; p=0.04 for $\mathrm{ADC}_{\mathrm{mean}}$) with the lowest CoV for large VOIs (Figure 4). Although 19 VOIs in the 'large' group were from study B, the majority (28 VOIs) were from other studies.

Discussion

The excellent repeatability of $\mathrm{ADC}_{\mathrm{median}}$ and $\mathrm{ADC}_{\mathrm{mean}}$ (CoV between 1.7% and 6.5% in all studies) demonstrates that ADC is a robust metric in clinical practice in oncology. The results reported in this analysis are comparable to results from similar test-retest repeatability studies although comparison with the literature is hindered by the variety of metrics that have been reported. From the published literature, a study of malignant hepatic tumors reported ICCs in the range 0.898 to 0.933 and LoA in the range 18.8% to 24.0% for $\mathrm{ADC}_{\mathrm{mean}}$ (26). A study in head-and-neck squamous cell carcinoma reported a RC of 15% for $\mathrm{ADC}_{\mathrm{mean}}$ (27). A study of hepatocellular carcinoma reported a CoV of 8.3% and lower and upper LoA of -41.1% and 18.6% respectively for $\mathrm{ADC}_{\mathrm{mean}}$ (28). In healthy volunteers, a study in abdominal organs reported RCs between 6.4% and 9.6% for $\mathrm{ADC}_{\mathrm{mean}}$ (29). A study of normal thyroid glands in healthy volunteers, which also followed the QIBA framework, reported $s_w^2=0.0147\times10^{-3}\mathrm{mm}^2\mathrm{s}^{-1}$, $\mathrm{RC}=0.3355\times10^{-3}\mathrm{mm}^2\mathrm{s}^{-1}$, ICC=0.9273, and CoV=9.88% using reduced-field-of-view DW-MRI (30). Comparison between published studies is not straightforward since they report different repeatability metrics but each result is similar to the present analysis for their respective metrics; however, most studies do not report CIs, which further hinders comparison.

The CoV and LoA, expressed as percentages, may be more intuitive for investigators to understand, compared with s_W or RC expressed on a log scale. Although the ICC is listed in the QIBA framework for reporting repeatability, ICC may not be an appropriate metric for comparison between studies as results are scaled to the inter-subject variability of the study cohort via s_B; a low ICC may therefore reflect a homogeneous cohort rather than poor repeatability (13). This is exemplified in study K where ICCs are low (ICC 0.126 to 0.677 in studies K1, K2, and K3) despite CoVs being comparable to other studies. Values of s_B are an order-of-magnitude lower than in studies A to J, reflecting the narrow range of ADC estimates in healthy organs in the tightly-controlled volunteer cohort. These results strongly suggest that the ICC should not be used to compare ADC repeatability between studies.

Knowledge of ADC repeatability is essential for assessment of post-treatment changes in an individual patient (as opposed to cohort changes, which can be assessed using a t-test, or similar); knowledge of measurement repeatability is also essential in power calculations to estimate the sample size necessary to detect a treatment effect in prospective cohort studies. Considering changes in ADC post-treatment, an increase of 12% or more in ADC_{median} or ADC_{mean} would be outside the 95% LoA for all VOIs analysed together – even considering the studies with the poorest repeatability (i.e. 'worst-case' studies), an increase of 20% would have been outside the 95% LoA in all studies. A tumor exhibiting such a change in ADC after treatment would therefore be assessed as exhibiting a post-treatment effect outside the expected variation of repeated measurements, with 95% confidence, when measured on the same scanner using the same imaging protocol, operator, and reader i.e. under repeatability conditions. This can be compared with post-treatment changes reported elsewhere: 23% and 24% increases in ADC_{mean} in responding patients with hepatic metastases of colorectal (3) and gastric cancers (4), respectively; and increases of 20% (ADC_{mean}) and 22% (ADC_{median}) in responding ovarian cancer patients treated with platinum-based chemotherapy (8). In studies reporting ADC changes in individual patients, as opposed to cohort changes, post-treatment increases in ADC_{mean} up to 100% were reported in cervical cancer patients following chemoradiotherapy (5) and increases in ADC_{mean} up to 50% were reported in patients with non-small cell lung cancer (9), thus the excellent repeatability demonstrated in the present analysis shows that ADC is sensitive to changes that are observed in clinical studies.

The significant difference between small, medium, and large VOIs shows that volume is an important factor in ADC repeatability. The weak correlation between the CoV and the median VOI volume in each study may reflect the range of tumor sizes within each study. The low CoV of 1.7% in study B may relate to the large tumors in this study. For future studies, the assumption of a CoV of 6.5% would be a conservative choice. It is worthwhile noting that the VOIs did not always encompass the whole tumor: ROIs were drawn around the whole area of the tumor/healthy organ on at least three slices in all studies, but studies A, B, and E included considerably more slices. Larger VOIs may provide more robust estimates of ADC_{median} and ADC_{mean} due to larger sample sizes. Furthermore, larger tumors may be less affected by motion or partial volume effects, which may lead to better ADC repeatability. ADC repeatability in paediatric patients (study F) was not worse than other studies, despite the additional challenges associated with patient compliance in this group.

The apparent absence of a relationship between the CoV and the year the study commenced may suggest that ADC repeatability has not changed markedly over 10 years despite advances in scanner technology and imaging protocol methodology during that time. This suggests that ADC repeatability assessments from older studies may inform future studies, although this may not apply across substantial changes in hardware/methodology, such as a change in field strength. Whilst this analysis only considered ADC repeatability, imaging protocol variations may also affect overall image quality, qualitative interpretation, and absolute values of ADC estimates, but these effects are outside the scope of this analysis. Reasons for variations in imaging protocols include changes in hardware and software capabilities; advances in knowledge; requirements for imaging particular patient cohorts, such as size of field-of-view or orientation of imaging plane; requirements of study sponsors; and requirements to match protocols in multi-center studies.

The apparent absence of a relationship between the CoV and the number of VOIs in the study (over the range 6 to 26 VOIs) may suggest that an informative estimate of repeatability may be obtained from as few as 6 patients, indicating that double-baseline examinations from relatively small subsets of patients may be used to efficiently estimate repeatability for larger studies. Repeatability studies may thus be easily conducted if a center wishes to assess its DW-MRI protocols. Inclusion of larger numbers of subjects, however, allows narrower CIs to be placed on estimated quantities and is advocated in clinical trials.

Repeatability estimates for ADC_{median} and ADC_{mean} do not apply to all summary statistics, for example other ADC histogram centiles may exhibit poorer repeatability (31). Alternative acquisition techniques, e.g. motion compensation, would also require new repeatability studies. Furthermore, it is common practice to use data from previous imaging studies to develop novel analysis methods, which require assessment of repeatability of resulting metrics in order to evaluate their potential value in clinical practice. Double-baseline studies therefore provide an invaluable resource for future developments of analysis methods.

There are limitations to our analysis. First, all studies were carried out at a single expert center and senior members of staff with extensive experience of extra-cranial DW-MRI were involved in development of imaging protocols for all studies. Second, all but one of the studies were carried out on the same scanner, with the remaining study carried out on a scanner from the same manufacturer; the generality of our conclusions for test-retest measurements across scanners from other manufacturers remains to be tested. Third, only one healthy volunteer study was included. Fourth, many of the studies were sub-studies that formed part of a larger clinical trial and there may be selection bias due to inclusion/exclusion criteria for these trials (e.g. including patients with lesions larger than 2cm, or excluding patients who had difficulty lying still). Generalization to routine clinical practice remains to be tested but the repeatability of ADC estimates in less controlled situations might be expected to be worse than the repeatability reported here.

In conclusion, **ADC is a robust imaging metric which demonstrates excellent repeatability in extra-cranial soft-tissue DW-MRI studies across a wide range of tumor sites, sizes, patient populations, and imaging protocol variations.** Estimates of ADC repeatability obtained from similar data can inform studies where double-baseline measurements are not possible, but a double-baseline format remains critical for future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr Thorsten Feiweier and Dr Berthold Kiefer at Siemens Healthcare for providing DWI prototype packages.

Funding information

We acknowledge CRUK and EPSRC support to the Cancer Imaging Centre at ICR and RMH in association with MRC and Department of Health C1060/A10334, C1060/A16464 and NHS funding to the NIHR Biomedical Research Centre and the Clinical Research Facility in Imaging. MOL is an Emeritus NIHR Senior Investigator.

For specific studies included in this work, we acknowledge funding from CRUK BIDD grants C7273/A12064 and C1353/A12762; CRUK and EPSRC Cancer Imaging Programme at the Children's Cancer and Leukaemia Group (CCLG) in association with the MRC and Department of Health (England) (C7809/A10342); an Experimental Cancer Medicine Centre Network award (joint initiative, CRUK and UK Department of Health) grants C51/A7401 and C12540/A15573; EPSRC Platform Grant EP/H046526/1; Experimental Cancer Medicine Centre (ECMC) Network funding for support to early clinical trials; and the support of the National Institute for Health Research through the Cancer Research Network. Some of these studies were supported by AstraZeneca, Merck, Basilea, ArQule, and Genentech. MRO was funded by AstraZeneca.

References

1. Taouli B, Beer AJ, Chenevert T, et al. Diffusion-weighted imaging outside the brain: consensus statement from an ISMRM-sponsored workshop. J Magn Reson Imag. 2016; doi: 10.1002/jmri.25196
2. Dzik-Jurasz A, Domenig C, George M, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. Lancet. 2002; 360:307–308. [PubMed: 12147376]
3. Koh D-M, Scurr E, Collins D, et al. Predicting response of colorectal hepatic metastasis: value of pretreatment apparent diffusion coefficients. AJR Am J Roentgenol. 2007; 188:1001–1008. [PubMed: 17377036]
4. Cui Y, Zhang X-P, Sun Y-S, Tang L, Shen L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. Radiology. 2008; 248(3):894–900. [PubMed: 18710982]
5. Harry VN, Semple SI, Gilbert FJ, Parkin DE. Diffusion-weighted magnetic resonance imaging in the early detection of response to chemoradiation in cervical cancer. Gynecol Oncol. 2008; 111:213–220. [PubMed: 18774597]
6. Sharma U, Danishad KKA, Seenu V, Jagannathan NR. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. NMR Biomed. 2009; 22:104–113. [PubMed: 18384182]
7. Kim S, Loevner L, Quon H, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res. 2009; 15(3):986–994. [PubMed: 19188170]
8. Kyriazi S, Collins DJ, Messiou C, et al. Metastatic ovarian and primary peritoneal cancer: assessing chemotherapy response with diffusion-weighted imaging - value of histogram analysis of apparent diffusion coefficients. Radiology. 2011; 261(1):182–192. [PubMed: 21828186]
9. Yabuuchi H, Hatakenaka M, Takayama K, et al. Non-small cell lung cancer: detection of early response to chemotherapy by using contrast-enhanced dynamic and diffusion-weighted MR imaging. Radiology. 2011; 261(2):598–604. [PubMed: 21852569]
10. [Accessed June 15, 2016] Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST technical note 1297Published 1994. http://www.nist.gov/pml/pubs/tn1297/index.cfm.
11. Winfield JM, Collins DJ, Priest AN, et al. A framework for optimization of diffusion-weighted MRI protocols for large field-of-view abdominal-pelvic imaging in multicenter studies. Med Phys. 2016; 43(1):95–110. [PubMed: 26745903]
12. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015; 24(1):27–67. [PubMed: 24919831]
13. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standard for quantitative imaging biomarkers. Radiology. 2015; 277(3):813–825. [PubMed: 26267831]

14. Messiou C, Orton M, Ang JE, et al. Advanced solid tumors treated with cediranib: comparison of dynamic contrast-enhanced MR imaging and CT as markers of vascular activity. Radiology. 2012; 265(2):426–436. [PubMed: 22891356]

15. Orton MR, Messiou C, Collins D, et al. Diffusion-weighted MR imaging of metastatic abdominal and pelvic tumours is sensitive to early changes induced by a VEGF inhibitor using alternative diffusion attenuation models. Eur Radiol. 2016; 26(5):1412–1419. [PubMed: 26253255]

16. Yap TA, Yan L, Patnaik A, et al. Interrogating two schedules of the AKT inhibitor MK-2206 in patients with advanced solid tumors incorporating novel pharmacodynamic and functional imaging biomarkers. Clin Cancer Res. 2014; 20(22):5672–5685. [PubMed: 25239610]

17. Koh D-M, Blackledge M, Collins DJ, et al. Reproducibility and changes in the apparent diffusion coefficients of solid tumours treated with combretastatin A4 phosphate and bevacizumab in a two-centre phase I clinical trial. Eur Radiol. 2009; 19:2728–2738. [PubMed: 19547986]

18. Yap TA, Olmos D, Brunetto AT, et al. Phase I trial of a selective c-MET inhibitor ARQ 197 incorporating proof of mechanism pharmacodynamic studies. J Clin Oncol. 2011; 29(10):1271–1279. [PubMed: 21383285]

19. Miyazaki K, Jerome NP, Collins DJ, et al. Demonstration of the reproducibility of free-breathing diffusion-weighted MRI and dynamic contrast enhanced MRI in children with solid tumours: a pilot study. Eur Radiol. 2015; 25:2641–2650. [PubMed: 25773937]

20. Bland MJ, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 327(8476):307–310.

21. Keene ON. The log transform is special. Stat Med. 1995; 14:811–819. [PubMed: 7644861]

22. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. BioScience. 2001; 51(5):341–352.

23. He X, Oyadiji SO. Application of coefficient of variation in reliability-based mechanical design and manufacture. J Mater Process Technol. 2001; 119:374–378.

24. Barnhart HX, Barboriak DP. Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets. Trans Oncol. 2009; 2(4):231–235.

25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86(2):420–428. [PubMed: 18839484]

26. Kim SY, Lee SS, Park B, et al. Reproducibility of measurement of apparent diffusion coefficients of malignant hepatic tumors: effect of DWI techniques and calculation methods. J Magn Reson Imag. 2012; 36(5):1131–1138.

27. Hoang JK, Choudhury KR, Chang J, Craciunescu OI, Yoo DS, Brizel DM. Diffusion-weighted imaging for head and neck squamous cell carcinoma: quantifying repeatability to understand early treatment-induced change. AJR Am J Roentgenol. 2014; 203:1104–1108. [PubMed: 25341151]

28. Hectors SJ, Wagner M, Besa C, et al. Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI? J Magn Reson Imag. 2016; doi: 10.1002/jmri.25194

29. Miquel ME, Scott AD, Macdougall ND, Boubertakh R, Bharwani N, Rockall AG. In vitro and in vivo repeatability of abdominal diffusion-weighted MRI. Br J Radiol. 2012; 85:1507–1512. [PubMed: 22674704]

30. Lu Y, Hatzoglou V, Banerjee S, et al. Repeatability investigation of reduced field-of-view diffusion weighted magnetic resonance imaging on thyroid glands. J Comput Assist Tomogr. 2015; 39(3): 334–339. [PubMed: 25700226]

31. Jerome NP, Miyazaki K, Collins DJ, et al. Repeatability of derived parameters from histograms following non-Gaussian diffusion modelling of diffusion-weighted imaging in a paediatric oncological cohort. Eur Radiol. 2017; 27:345–353. [PubMed: 27003140]

Advances in knowledge

1. Repeated apparent diffusion coefficient (ADC) estimates can be obtained from extra-cranial soft-tissue diffusion-weighted magnetic resonance imaging (DW-MRI) with coefficient of variation (CoV) between 2 and 7%.
2. ADC repeatability does not differ markedly (CoV 2 to 7%) between DW-MRI studies across a wide range of patient cohorts and imaging protocol variations.
3. Better ADC repeatability is observed in large tumors, compared with smaller tumors.

Implications for patient care

DW-MRI can be used to estimate ADC with good repeatability in extra-cranial soft-tissues, allowing a post-treatment increase of 12% or more in ADC to be distinguished.

Summary statement

ADC is a robust imaging metric which demonstrates excellent repeatability in extra-cranial soft-tissue DW-MRI studies across a wide range of tumor sites, sizes, patient populations, and imaging protocol variations.

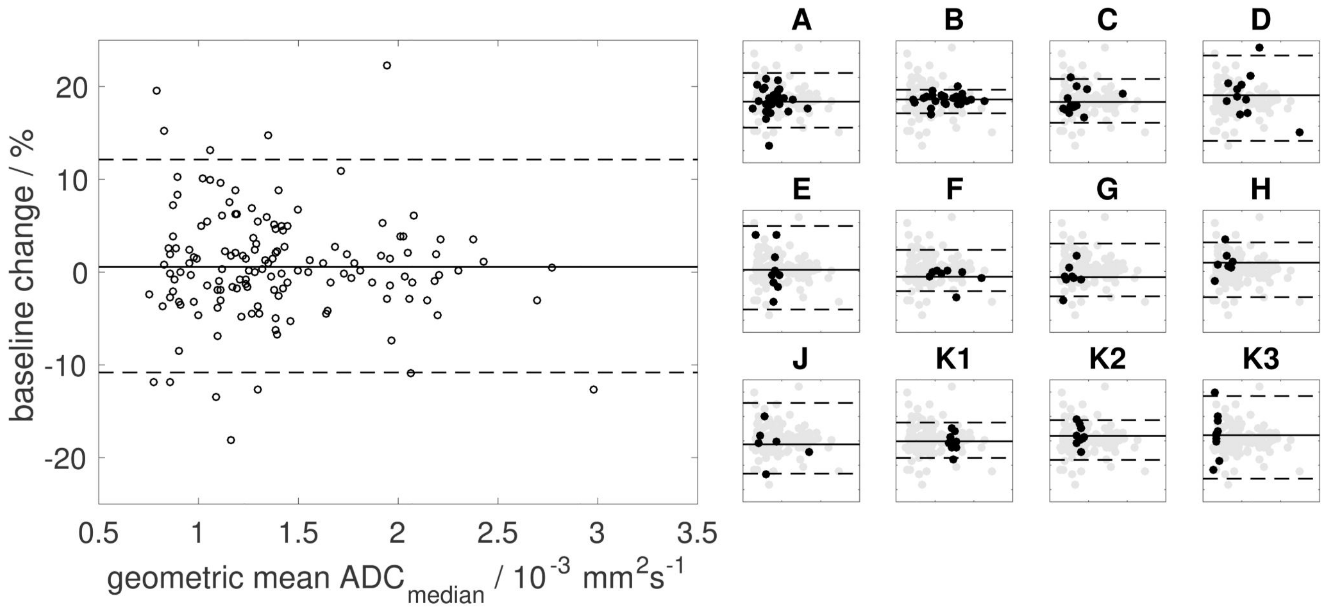

Figure 1.

Bland-Altman plot showing percentage change between two baseline estimates of ADC$_{median}$ versus their geometric mean for all VOIs in all studies. Sub-plots (A to K3) show Bland-Altman plots for each study (black markers) with VOIs from all other studies shown in grey; x- and y-axis limits are the same as main figure. On each plot, solid lines show the mean difference between two baseline examinations for the specified data and dashed lines show the 95 % limits of agreement.

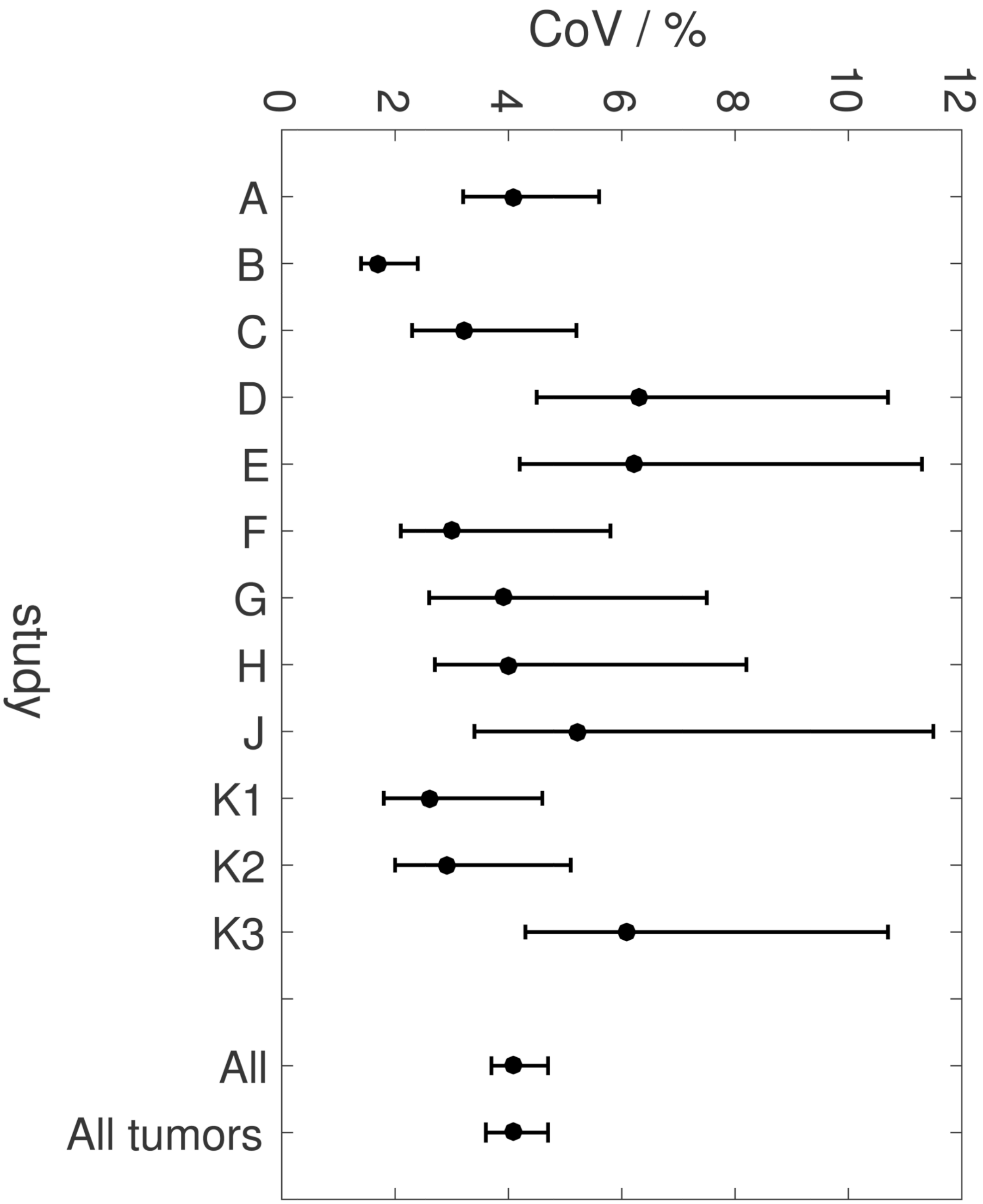

Figure 2.

CoV of ADC_{median} for each study (A to K3); all VOIs analyzed together (All); and all tumor VOIs analyzed together (All tumors). Whiskers represent 95 % confidence intervals for CoV estimates.

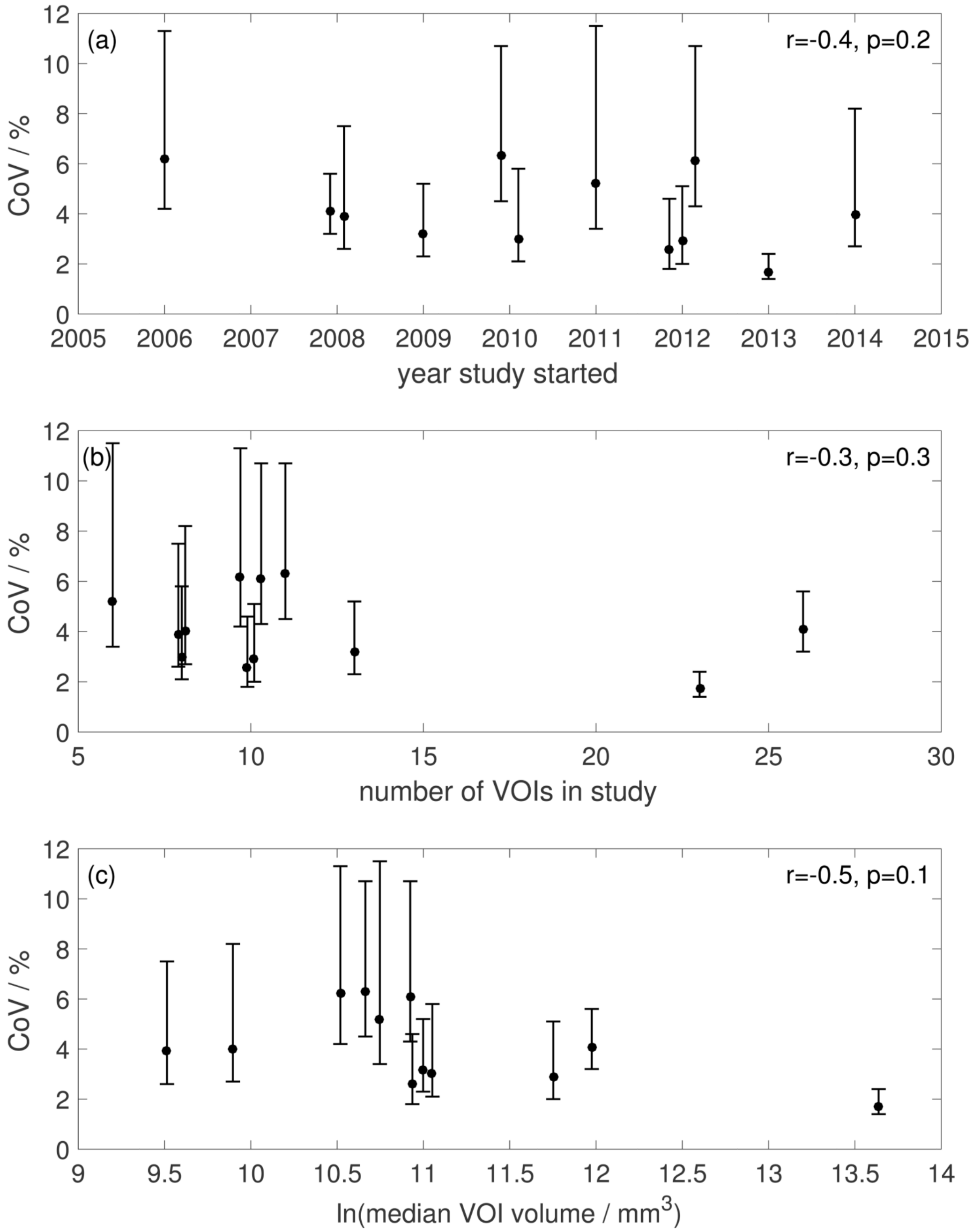

Figure 3.

Plot of CoV of ADC$_{median}$ versus (a) year study started; (b) number of VOIs (subjects or lesions) in the study; (c) natural logarithm of the median volume of the VOIs in the study. Error bars represent 95 % confidence intervals of CoV estimates. In (a) and (b), studies with identical start dates or numbers of VOIs have been offset for clarity.

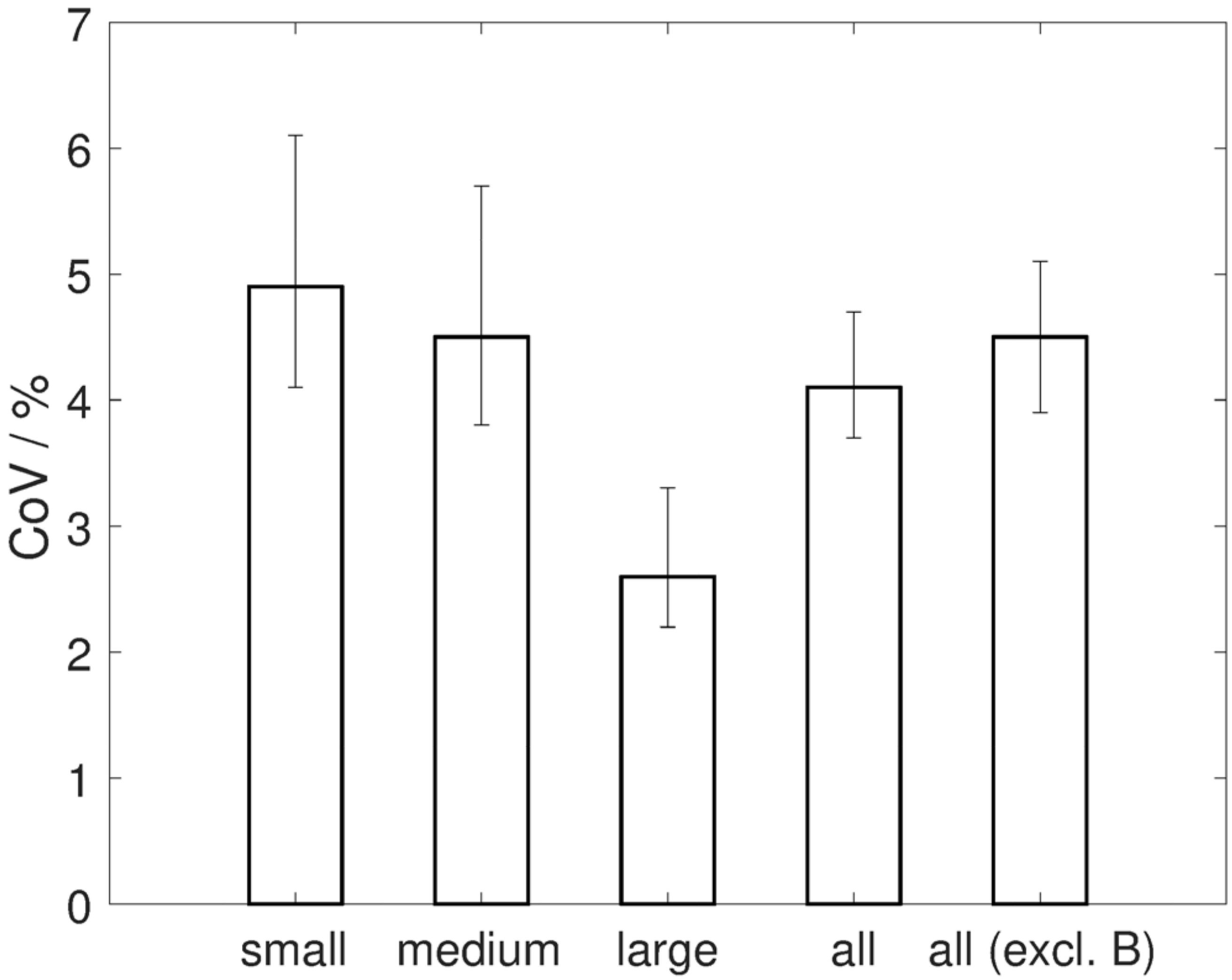

Figure 4.

CoV of $\mathrm{ADC}_{\mathrm{median}}$ for small, medium, and large VOIs, all VOIs together, and all VOIs excluding study B. Error bars represent 95 % confidence intervals of CoV estimates.

Table 1

Subjects and imaging procedures for each study (studies labelled A to K).

Study *	**A(14,15)**	**B**	**C (16)**	**D**	**E (17)**	**F (19)**	**G (18)**	**H**	**J**	**K1, K2, K3 (11)**
Patient / volunteer cohort	Patients (phase 1 trial population; adults)	Patients (adults)	Patients (phase 1 trial population; adults)	Patients (adults)	Patients (phase 1 trial population; adults)	Patients (pediatric)	Patients (phase 1 trial population; adults)	Patients (phase 1 trial population; adults)	Patients (phase 1 trial population; adults)	Healthy volunteers (adults, female)
Tumor site / healthy organ examined	Mixed (17 abdominal, 9 pelvic lesions)	Retroperitoneal soft-tissue masses	Mixed (8 liver lesions, 1 splenic lesion, 1 renal lesion, 1 peritoneal lesion, 1 abdominal wall lesion, 1 pelvic lymph node)	Renal cell carcinoma	Mixed (5 liver lesions, 4 pelvic lesions)	Mixed (extra-cranial solid tumors)	Mixed (4 liver lesions, 2 pelvic lesions, 2 pelvic lymph nodes)	Mixed (5 liver lesions, 1 lung lesion, 1 abdominal lymph node)	Liver lesions	Healthy organs K1: kidneys K2: liver K3: spleen
Number patients / healthy volunteers	26	23	13	11	9	8	8	7	6	10
Start and end dates of study	2008 to 2010	2013 to 2016	2009 to 2010	2010 to 2016	2006 to 2007	2010 to 2014	2008 to 2009	2014 to 2016	2011 to 2012	2012
Interval between two exams	7 days	~ 45 minutes ('coffee-break' repeatability)	4 to 7 days	24 hours	2 to 10 days	24 hours	4 to 10 days	2 to 3 days	5 days	1 to 7 days
Method used to define VOIs †	ROIs drawn by a consultant radiologist (CM) with 7 years of experience, including 3 years experience of extra-cranial DW-MRI. ROIs drawn around whole area of tumor on b=500s	ROIs drawn by a consultant radiologist (CM) with 12 years of experience, including 8 years experience of extra-cranial DW-MRI. ROIs drawn around whole area of tumor on T_2-w images on all slices on which the tumor appeared; ROIs transferred to ADC maps (excluding the most cranial and caudal slices if partial volume effects were visible). 2 stations acquired if necessary to cover very large tumors.	ROIs drawn by a radiologist (NT) with 5 years of experience, including 4 years experience of extra-cranial DW-MRI. ROIs drawn around whole tumor on highest b-value images on three to	ROIs drawn by a consultant radiologist (DMK) with >10 years of experience, including >10 years experience of extra-cranial DW-MRI. ROIs drawn around tumor on 5 central slices on high b-value images with reference to other imaging.	ROIs drawn by a consultant radiologist (DMK) with >10 years of experience, including >5 years experience of extra-cranial DW-MRI. ROIs drawn around whole area of tumor on b=750s	ROIs drawn by a consultant radiologist (DMK) with >10 years of experience, including >10 years experience of extra-cranial DW-MRI. ROIs drawn around tumor on three slices near the center of the imaging volume; matching slices selected for second baseline examination.	ROIs drawn by a radiologist (NT) with 5 years of experience, including 4 years experience of extra-cranial DW-MRI. ROIs drawn around whole tumor on highest b-value images on three to	ROIs drawn by a consultant radiologist (NT) with 9 years of experience, including 8 years experience of extra-cranial DW-MRI. ROIs drawn around tumor on the highest b-value images	ROIs drawn by a consultant radiologist (NT) with 6 years of experience, including 5 years experience of extra-cranial DW-MRI. ROIs drawn around tumor on the highest b-value images	ROIs drawn by a MR physicist (JMW) with 4 years of experience, including 2 years experience of extra-cranial DW-MRI. ROIs drawn by region growing (kidneys and spleen) or freehand (liver). ROIs

Study [*]	A(14,15)	B	C (16)	D	E (17)	F (19)	G (18)	H	J	K1, K2, K3 (11)
	mm^{-2} diffusion-weighted images on all slices on which tumor appeared (up to maximum 20 slices in imaging volume).		six slices near the center of the imaging volume.		mm^{-2} diffusion-weighted images on all slices on which tumor appeared (excluding the most cranial and caudal slices if partial volume effects were visible).		six slices near the center of the imaging volume.	on three slices at the center of the imaging volume; matching slices selected for second baseline examination.	on three slices at the center of the imaging volume; matching slices selected for second baseline examination.	drawn on computed diffusion-weighted images (b=500s mm^{-2} for kidneys, b=800s mm^{-2} for liver, b=1000s mm^{-2} for spleen) encompassing whole area of organ on three contiguous slices.

[*] References provided for previously published studies.

[†] ROIs were drawn by the authors named in each column; years of experience stated for each study reflect experience at the time of the original analysis of the study.

Table 2

Imaging protocols for each study (studies labelled A to K).

Study *	**A (14,15)**	**B**	**C (16)**	**D**	**E (17)**	**F (19)**	**G (18)**	**H**	**J**	**K1, K2, K3 (11)**
MR scanner	Siemens MAGNETOM Avanto	Siemens MAGNETOM Aera	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto	Siemens MAGNETOM Avanto
Orientation of imaging plane	axial	axial	coronal oblique	coronal	axial	coronal	axial	coronal	coronal	axial
Slice thickness / mm	5	5	5	5	5	5	5	5	5	6
Field-of-view (read × phase) / mm	380 × 380	380 × 256	380 × 380	380 × 380	340 × 298	300 × 300	380 × 308	380 × 380	380 × 380	380 × 332
Acquired matrix (read × phase)	128 × 128	160 × 108	128 × 128	128 × 128	128 × 112	128 × 128	128 × 104	128 × 128	128 × 128	128 × 112
Reconstructed matrix (read × phase)	256 × 256	320 × 216	256 × 256	256 × 256	256 × 256	256 × 256	256 × 208	256 × 256	256 × 256	256 × 224
TE / ms	69	65	70	64	72	75	69	75	68	75
TR / ms	3500	9200	2500 to 7000	4000	3500	3500	3500	3500	3500	8000
Fat suppression	SPAIR	SPAIR	SPAIR	SPAIR	chemical fat suppression	SPAIR	SPAIR	SPAIR	SPAIR	SPAIR
b-values used for ADC estimates / s mm^{-2}	0, 50, 100, 250, 500, 750	50, 600, 900	50, 100, 300, 600, 900	0, 20, 40, 60, 80, 100, 250, 500, 750, 1000 †	0, 50, 100, 250, 500, 750	0, 50, 100, 300, 600, 1000	0, 50, 100, 300, 600, 900, 1050	0, 50, 100, 300, 600, 1000	150, 600, 900	100, 500, 900 (b = 0 acquired but not used in ADC estimation)
Diffusion encoding scheme	3-scan trace	3-scan trace	3-scan trace	3-scan trace	orthogonal	3-scan trace	3-scan trace	3-scan trace	3-scan trace	3-scan trace
Receiver bandwidth	1775 Hz/pixel	1954 Hz/pixel	1565 Hz/pixel	1628 Hz/pixel	1445 Hz/pixel	1860 Hz/pixel	1775 Hz/pixel	1954 Hz/pixel	1776 Hz/pixel	1776 Hz/pixel
Phase partial Fourier	6/8	6/8	6/8	6/8	6/8	6/8	6/8	7/8	6/8	not used
NSA	6	NSA = 4 for b = 50 s mm^{-2} and 600 s mm^{-2}; NSA = 5 for b = 900 s mm^{-2}	4	4	5	3	6	5	10	4
Breathing instructions	free breathing	free breathing	respiratory triggering for liver, splenic and renal lesions (n=10 patients), free-breathing for pelvic nodal, abdominal wall and peritoneal lesions (n=3 patients)	free breathing	free breathing	free breathing	free breathing	free breathing	free breathing	free breathing

Study [*]	A (14,15)	B	C (16)	D	E (17)	F (19)	G (18)	H	J	K1, K2, K3 (11)
Acquisition time (mins, secs)	6 mins, 24 secs	6 mins, 28 secs (per station)	variable	11 mins	4 mins	3 mins, 30 secs	7 mins	4 mins, 51 secs	5 mins, 26 secs	5 mins, 44 secs

[*] References provided for previously published studies.

[†] Proprietary DW-MRI prototype packages used.

All studies were carried out at 1.5 T. The following parameters were common to all studies: single-shot echo-planar imaging; parallel imaging using generalized autocalibrating partially parallel acquisitions (GRAPPA), with acceleration factor 2; bipolar diffusion gradient scheme; three diffusion-encoding directions; trace-weighted images.

Table 3

Repeatability of $\mathrm{ADC}_{\mathrm{median}}$

Study	CoV / %	95 % LoA		RC (log scale)	s_W (log scale)	s_B (log scale)	ICC
		upper LoA / %	lower LoA / %				
A	4.1 (3.2, 5.6)	11.9 (9.2, 16.6)	-10.6 (-14.3, -8.5)	0.112 (0.088, 0.154)	0.040 (0.032, 0.055)	0.218	0.967 (0.928, 0.985)
B	1.7 (1.4, 2.4)	4.9 (3.8, 7.0)	-4.7 (-6.5, -3.7)	0.048 (0.037, 0.068)	0.017 (0.014, 0.024)	0.279	0.996 (0.991, 0.998)
C	3.2 (2.3, 5.2)	9.4 (6.7, 15.5)	-8.6 (-13.4, -6.3)	0.090 (0.065, 0.144)	0.032 (0.023, 0.052)	0.256	0.984 (0.951, 0.995)
D	6.3 (4.5, 10.7)	19.0 (13.1, 34.4)	-16.0 (-25.6, -11.6)	0.174 (0.123, 0.296)	0.063 (0.045, 0.107)	0.251	0.941 (0.806, 0.984)
E	6.2 (4.2, 11.3)	18.6 (12.5, 36.6)	-15.7 (-26.8, -11.1)	0.171 (0.118, 0.312)	0.062 (0.042, 0.113)	0.147	0.851 (0.504, 0.964)
F	3.0 (2.1, 5.8)	8.8 (5.9, 17.5)	-8.1 (-14.9, -5.5)	0.084 (0.057, 0.162)	0.030 (0.021, 0.058)	0.217	0.981 (0.915, 0.996)
G	3.9 (2.6, 7.5)	11.4 (7.6, 23.1)	-10.3 (-18.8, -7.1)	0.108 (0.073, 0.208)	0.039 (0.026, 0.075)	0.140	0.928 (0.709, 0.985)
H	4.0 (2.7, 8.2)	11.8 (7.7, 25.6)	-10.6 (-20.4, -7.1)	0.112 (0.074, 0.228)	0.040 (0.027, 0.082)	0.150	0.932 (0.696, 0.988)
J	5.2 (3.4, 11.5)	15.5 (9.7, 37.4)	-13.4 (-27.2, -8.9)	0.144 (0.093, 0.317)	0.052 (0.034, 0.115)	0.302	0.971 (0.839, 0.996)
K1	2.6[†] (1.8, 4.6)	7.5 (5.2, 13.6)	-7.0 (-12.0, -4.9)	0.073 (0.051, 0.127)	0.026 (0.018, 0.046)	0.023	0.427 (-0.205, 0.816)
K2	2.9[†] (2.0, 5.1)	8.4 (5.8, 15.2)	-7.8 (-13.2, -5.5)	0.081 (0.056, 0.142)	0.029 (0.020, 0.051)	0.042	0.677 (0.158. 0.907)
K3	6.1[†] (4.3, 10.7)	18.4 (12.5, 34.5)	-15.6 (-25.7, -11.1)	0.169 (0.118, 0.297)	0.061 (0.043, 0.107)	0.023	0.126 (-0.491, 0.673)
All [*]	4.1 (3.7, 4.7)	12.1 (10.8, 13.8)	-10.8 (-12.2, -9.7)	0.115 (0.103, 0.130)	0.041 (0.037, 0.047)	0.309	0.982 (0.976, 0.987)

Note: Lower and upper 95 % confidence intervals are shown in parentheses.

CoV = coefficient of variation.

LoA = limits of agreement.

RC = repeatability coefficient.

s_W = within-subject standard deviation.

s_B = between-subject standard deviation.

ICC = intraclass correlation coefficient.

[*] Results are shown for each study and for all VOIs analysed together (denoted ‘All’).

[†] CoVs from K1, K2, and K3 reproduced from Winfield et al (11) for completeness.

Note: Estimates of $\mathrm{ADC}_{\mathrm{median}}$ for two baseline examinations for all tumors/organs are tabulated in the Supplemental Material (Table 5).