



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?

Frederic Carsten Schmeel¹

Received: 15 October 2018 / Accepted: 25 October 2018 / Published online: 28 November 2018
© European Society of Radiology 2018

Abstract

Key Points

- Variability of ADC measurements may be substantial across different MRI scanners and imaging sites.
- DWI protocol standardization and increased awareness of frequent sources of error can help reducing the limits of expected bias.
- Focusing on ADC change and normalized ADC values rather than on absolute measurements can facilitate consistent use of ADC in multi-center studies.

Keywords Magnetic resonance imaging · Diffusion · Reproducibility of results · Multi-center studies

Abbreviations

ADC	Apparent diffusion coefficient
DWI	Diffusion-weighted magnetic resonance imaging
EPI	Echo-planar imaging
FOV	Field of view
MRI	Magnetic resonance imaging
ROI	Region of interest
TE	Time to echo
TR	Repetition time

Diffusion-weighted magnetic resonance imaging (DWI) has become an integral part of oncologic imaging studies to aid in the detection and diagnosis of solid tumors. While regions of reduced extracellular space due to increased cell density will result in restricted diffusion of mobile proton species, disruptions in cellular integrity as attributable to apoptosis may lead to an increase in free-water molecule motion. Consequently, the most applied quantitative parameter for describing tissue diffusivity, the apparent diffusion coefficient (ADC), has been

widely investigated as a prognostic biomarker for response assessment to anti-cancer therapy [1].

However, despite its value in diagnosing and monitoring of disease, there is some controversy in the literature regarding the limited reproducibility of ADC estimates across different imaging platforms and imaging sites. This is mainly because variations in acquisition parameters across scanners from different vendors and between models from the same manufacturer can widely affect ADC quantitation. Previous studies investigated ADC estimates in the abdominal organs of DWI data sets from healthy volunteers imaged on scanners from different vendors at 1.5 and 3 T in order to determine the extent of ADC variability. Among others, two of these multi-center studies found poor agreement in the pancreas and kidneys at 3 T and poor agreement particularly in the liver at both field strengths [2, 3]. The DWI parameter settings were kept constant across imagers to the maximum possible extent; as such, the same echo-planar imaging (EPI) sequence, field of view (FOV), matrix size, number of averages, slice thickness, and interslice gap were set identically in all scanners. However, the TE, TR, b-values, and methods of parallel imaging could not be completely standardized across platforms—although the preset contained the same subset of b-values for subsequent acquisition. Another multi-center study focusing on gray and white matter in healthy volunteers also showed significant variations in ADC estimates between scanners from the same vendor and between scanners from different manufacturers; again, TE, TR, and methods of parallel imaging

✉ Frederic Carsten Schmeel
Carsten.Schmeel@ukb.uni-bonn.de

¹ Department of Radiology, University Hospital Bonn, Rheinische-Friedrich-Wilhelms-Universität Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany

could not be fully standardized [4]. Hence, timing parameters and methods of parallel imaging were shown to be inconsistent despite effective protocol standardization, demonstrating the difficulty of producing fully standardized DWI protocols across different imaging platforms and sites. This lack of standardization of ADC measurement is a serious limitation to this quantitative parameter to be a reliable imaging biomarker.

In oncologic imaging studies, ADC reproducibility further depends on various critical factors including the tumor pathophysiology itself [5], curve-fitting techniques [6], and system stability [7]. Voxel-wise ADC quantification in the liver is also specifically degraded by respiratory motion artifacts, with as little or no improvement despite utilization of compensation methods such as navigator echo and respiratory gating [8]. The common lack of agreement in ADC estimates of the liver may also depend on the lower signal compared with other organs, owing to the short T2 decay, eventually leading to greater sensitivity to the noise characteristics of different MR imagers. Another more practical but not insubstantial source of error in ADC quantitation is the choice of delineation methods for region of interest (ROI) selection; even small variations in the ROI geometry and its position can substantially influence tumor ADC measurements and thus affect sensitivity to changes in tumoral ADC values [9]. Similarly, therapy-induced changes in tumor vascularization or peritumoral tissue density may affect the depiction and thus reproducibility of ROI delineation. Current consensus guidelines for DWI as a cancer biomarker recommend delineation of 3D whole tumor volumes that are more reproducible than single-slice measurements but require substantially more examination time [1]. Automated methods which may provide improved accuracy could remedy this situation but would require the highest possible image quality to accurately delineate tumor boundaries. Since abdominal DWI is regularly subject to physiological motion, this would probably necessitate regular manual interventions which query the concept of using a fully automated ROI application scheme as an add-on technique.

Summarizing the above, it appears that multi-centric and multi-device studies, highly desirable in any case, are almost always subject to measurement bias and therefore difficult to establish. Therefore, the question we are facing is how to overcome the present impasse? A possible consensus that might apply to therapy response assessment could be comparative analysis of the percentage difference between pre- and post-treatment ADC values obtained from the same scanner, rather than focusing on absolute ADC values alone. Alternatively, normalization of ADC values by calculating the ratio of the investigated organ to a reference organ could help reducing the limits of expected bias. Given that the variety of described confounders will be identified and widely corrected prior to the calculation of parametric images, this could facilitate consistent use of ADC as an imaging

biomarker for multi-center or longitudinal studies if absolute ADC values are not directly comparable between imagers. However, this still remains to be verified by larger and particularly patient studies, and major efforts need to continue to enhance protocol standardization.

Recently, a rather mathematically oriented study pursued a statistical approach to minimize ADC variability across four imaging sites with scanners from different vendors by implementing a post hoc correction model to already calculated parametric DW images [10]. The primary endpoint of this study was to define a statistical model of predictable sources of variability that contribute to measurement error (also including data sets with visible motion artifacts) and fit this to observed data in order to quantify the level of uncertainty in mean ADC repeatability. By application of the proposed model, the 95% confidence interval width used to determine a statistically significant ADC change in 20 patients with colorectal cancer liver metastases reduced from 21.1 to 2.7% after standardization. According to the authors, implementation of the proposed model will allow significant improvements in sensitivity for detection of change in ADC. They provided a lookup chart to allow investigators to estimate uncertainty due to statistical measurement error, for any given tumor volume and ADC histogram width [10]. This model may help to assess reproducibility with greater confidence, and could also be easily implemented into clinical routine.

Funding The author states that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. med. Frederic Carsten Schmeel at Bonn University Hospital.

Conflict of interest The author declares that he has no conflict of interest.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because this is not a study but an editorial comment.

Ethical approval Institutional Review Board approval was not required for this study because this is not a study but an editorial comment.

Methodology

- Observational

References

1. Padhani AR, Liu G, Koh DM et al (2009) Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia* 11:102–125
2. Donati OF, Chong D, Nanz D et al (2014) Diffusion-weighted MR imaging of upper abdominal organs: field strength and intervender

- variability of apparent diffusion coefficients. *Radiology* 270:454–463
3. Winfield JM, Collins DJ, Priest AN et al (2016) A framework for optimization of diffusion-weighted MRI protocols for large field-of-view abdominal-pelvic imaging in multicenter studies. *Med Phys* 43:95–110
 4. 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
 5. Asselin MC, O'Connor JP, Boellaard R, Thacker NA, Jackson A (2012) Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer* 48:447–455
 6. Winfield JM, deSouza NM, Priest AN et al (2015) Modelling DW-MRI data from primary and metastatic ovarian tumours. *Eur Radiol* 25:2033–2040
 7. Malyarenko D, Galbán CJ, Londy FJ et al (2013) Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. *J Magn Reson Imaging* 37:1238–1246
 8. Taouli B, Sandberg A, Stemmer A et al (2009) Diffusion-weighted imaging of the liver: comparison of navigator triggered and breathhold acquisitions. *J Magn Reson Imaging* 30:561–568
 9. Lambregts DM, Beets GL, Maas M et al (2011) Tumour ADC measurements in rectal cancer: effect of ROI methods on ADC values and interobserver variability. *Eur Radiol* 21:2567–2574
 10. Pathak R, Ragheb H, Thacker NA et al (2017) A data-driven statistical model that estimates measurement uncertainty improves interpretation of ADC reproducibility: a multi-site study of liver metastases. *Sci Rep* 26:14084