

Fig. 5 A 72-year-old man with squamous cell carcinoma of the lower thoracic esophagus. Axial DWIBS/T2WI fusion image shows abnormal high signal intensity in the middle thoracic esophageal wall. Histopathological examination revealed that the cancer had invaded the submucosa (T1b)

between lymph-node size and the frequency of nodal metastases was observed well illustrates the diagnostic inadequacy of nodal size criteria [5].

Diagnostic performance of FDG-PET/CT is also insufficient for nodal staging [6, 7]; a recent meta-analysis [7] reported a pooled sensitivity and specificity of 51 and 84% for the detection of locoregional metastases. Possible explanations for the low to moderate diagnostic performance of FDG-PET include masking of adjacent lymph nodes by tracer accumulation of the primary tumor, the inability of PET to detect lymph-node metastases below its spatial resolution, and occurrence of false-positive findings due to chronic inflammation [7]. Furthermore, although EUS combined with FNA of suspicious lymph nodes may currently be regarded as the most accurate preoperative method for determining nodal status, it is invasive and cannot be performed in the area of the primary tumor because of the risk of tumor seeding [3, 9].

DWI may be an excellent noninvasive alternative because it provides functional information and may be used for the detection and characterization of primary malignant tumors and (nodal) metastases [11]. Normal lymph nodes have a relatively restricted diffusion because of their high cellular density, while metastatic lymph nodes may have an even more restricted diffusion because of increased cellular density, and this may allow discrimination between both groups using DWI [11]. DWI was previously limited to the brain, but technological advances

(including the introduction of EPI and parallel imaging techniques, such as SENSE) and the development of the concept of DWIBS have enabled (near) volumetric extra-cranial DWI [11].

The main features of DWIBS are image acquisition during free breathing, multiple signal averaging, and background body signal suppression by means of a fat suppression pre-pulse and heavy diffusion weighting. To understand the feasibility of DWI under free breathing, knowledge of the concepts of intravoxel incoherent motion (IVIM) and intravoxel coherent motion (IVCM) is important. IVIM refers to the incoherent, random motion of water molecules within a voxel (diffusion equals IVIM). IVIM implies a change in the spatial relationship among water molecules within a voxel. IVCM refers to the coherent, synchronous motion of water molecules within a voxel. Respiratory motion can be regarded as IVCM during the small period in which the two MPGs in the Stejskal-Tanner sequence are applied. Since the acquired phase shift of coherently moving water molecules induced by the two MPGs in the Stejskal-Tanner sequence is equal in each phase-encoding step, image formation is not affected and diffusion-weighted image contrast is maintained [11].

This study first shows that DWIBS can detect almost all locally advanced (T3 and T4) tumors, whereas the majority of early (T1 and T2) esophageal cancers are not detected. Although detection of the primary tumor by DWIBS is of

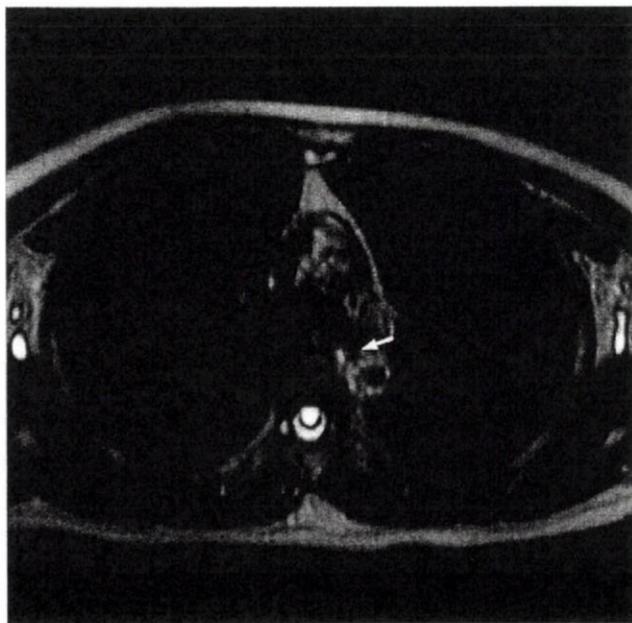


Fig. 6 A 65-year-old man with squamous cell carcinoma of the middle thoracic esophagus. On the axial DWIBS/T2WI fusion image, only one reader could point out a faint area of abnormal high signal intensity (arrow). Histopathological examination revealed that the cancer had invaded the submucosa (T1b)

less clinical importance, it gives an indication of the sensitivity of DWIBS for the depiction of (intrathoracic) lesions. The low sensitivity of DWIBS for the detection of early esophageal cancer may be explained by the small lesion size of early esophageal cancers and signal loss due to "incoherent" cardiac motion [11]. It should be noted, however, that detection rate of MT primary tumors (nearest to the heart) was not markedly lower than that of Ut and Lt primary tumors (although this finding is not conclusive because of the low number of cases). Nevertheless, detection of small intrathoracic lesions may be improved by combining DWIBS with cardiac triggering (although at the expense of prolonged data acquisition). Sensitivity of DWIBS may also be improved in the future by the development of more sophisticated multichannel coils and accelerated parallel imaging, which will allow acquisition of true isotropic high-resolution volume data sets.

Average ADC of 27 primary tumors was $1.26 \pm 0.29 \times 10^{-3} \text{ mm}^2/\text{s}$, which, to our knowledge, is the first reported ADC of esophageal cancer in the literature. However, potential clinical applications using the ADC of the primary tumor, such as in predicting response to therapy and survival, were not investigated in the current study. This study further indicates that diagnostic performance of DWIBS regarding assessment of nodal status is limited, given the average patient-based sensitivity and specificity of 77.8 and 55.6%, and the average lymph-node group-based sensitivity and specificity of 39.4 and 92.6%. This diagnostic inadequacy is also well illustrated by the overlap between ADCs of metastatic lymph nodes and ADCs of nonmetastatic lymph nodes. This finding is consistent with results of recent studies, which have shown that ADCs of metastatic lymph nodes were not significantly different from those of nonmetastatic lymph

nodes in patients with abdominal [13] and cervical or uterine cancer [14, 15].

Despite the overlap in ADCs between the two groups, mean ADC of metastatic lymph nodes was (unexpectedly) significantly higher than that of nonmetastatic lymph nodes in the present study. Perhaps the presence of microscopic areas of necrosis within the metastatic lymph nodes may explain the measured increased diffusion, although this was not verified histopathologically. Despite its inability to adequately discriminate metastatic from nonmetastatic lymph nodes, DWIBS may still have a role in identifying lymph nodes in patients with esophageal cancer, irrespective of their histological composition. DWIBS, for example, may be complimentary to nanoparticle-enhanced MR imaging, which has proven to be very useful for characterization of lymph nodes [16] but may have difficulties in tracing them. The combination of DWIBS and nanoparticle-enhanced MR imaging in patients with esophageal cancer deserves further investigation.

A limitation of this study is that only one of the included patients (4.2%) had adenocarcinoma, while 95.8% (23/24) had squamous cell carcinoma of the esophagus. This can be explained by the fact that the vast majority of esophageal cancers in Asia and Japan are squamous cell carcinomas [1, 2]. However, it would have been desirable to include more patients with adenocarcinoma, especially considering the rapid increase of adenocarcinoma in the United States and several Western European countries over the past 25 years [1, 2]. Another drawback of this study is that the cervical portion of the esophagus (Ce) [12] could not be evaluated because this region was outside the scanning range of the applied protocol.

In conclusion, this study indicates that DWI has a limited role in detecting esophageal cancer and nodal staging.

References

1. Holmes RS, Vaughan TL (2007) Epidemiology and pathogenesis of esophageal cancer. *Semin Radiat Oncol* 17:2-9
2. Pera M, Manterola C, Vidal O, Grande L (2005) Epidemiology of esophageal adenocarcinoma. *J Surg Oncol* 92:151-159
3. DeMeester SR (2006) Adenocarcinoma of the esophagus and cardia: a review of the disease and its treatment. *Ann Surg Oncol* 13:12-30
4. Kelly S, Harris KM, Berry E et al (2001) A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma. *Gut* 49:534-539
5. Schröder W, Baldus SE, Mönig SP, Beckurts TK, Dienes HP, Hölscher AH (2002) Lymph node staging of esophageal squamous cell carcinoma in patients with and without neoadjuvant radiochemotherapy: histomorphologic analysis. *World J Surg* 26:584-587
6. Bruzzi JF, Munden RF, Truong MT et al (2007) PET/CT of esophageal cancer: its role in clinical management. *Radiographics* 27:1635-1652
7. Van Westreenen HL, Westerterp M, Bossuyt PM et al (2004) Systematic review of the staging performance of 18F-fluorodeoxyglucose positron emission tomography in esophageal cancer. *J Clin Oncol* 22:3805-3812
8. Hofstetter W, Correa AM, Bekele N et al (2007) Proposed modification of nodal status in AJCC esophageal cancer staging system. *Ann Thorac Surg* 84:365-373
9. Freeman RK, Wait MA (2001) Port site metastasis after laparoscopic staging of esophageal carcinoma. *Ann Thorac Surg* 71:1032-1034
10. Takahara T, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M (2004) Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. *Radiat Med* 22:275-282
11. Kwee TC, Takahara T, Ochiai R, Nieuvelstein RAJ, Luijten PR (2008) Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology. *Eur Radiol* 18:1937-1952

-
12. Japanese Society for Esophageal Diseases (2004) Guidelines for clinical and pathologic studies on carcinoma of the esophagus, 9th ed. Preface, general principles, part I. *Esophagus* 1:61-88
 13. Akduman EI, Momtahan AJ, Balci NC, Mahajann N, Havlioglu N, Wolverson MK (2008) Comparison between malignant and benign abdominal lymph nodes on diffusion-weighted imaging. *Acad Radiol* 15:641-646
 14. Kim JK, Kim KA, Park BW, Kim N, Cho KS (2008) Feasibility of diffusion-weighted imaging in the differentiation of metastatic from nonmetastatic lymph nodes: early experience. *J Magn Reson Imaging* 28:714-719
 15. Nakai G, Matsuki M, Inada Y et al (2008) Detection and evaluation of pelvic lymph nodes in patients with gynecologic malignancies using body diffusion-weighted magnetic resonance imaging. *J Comput Assist Tomogr* 32:764-768
 16. Will O, Purkayastha S, Chan C et al (2006) Diagnostic precision of nanoparticle-enhanced MRI for lymph-node metastases: a meta-analysis. *Lancet Oncol* 7:52-60

Meeting Report

Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations

Anwar B. Padhani, Guoying Lu, Dow-Mu Koh, Thomas L. Chenover, Harriet C. Thoeny, Taro Takehara, Andrew Padhani, Brian B. Ross, Marc Van Calster, David Collins, Dima A. Hanjri, Gordon J.S. Rusth, Rami Yabl, and Peter L. Choyke

From the Cancer Imaging Research Center, National Cancer Institute, Bethesda, MD (Anwar B. Padhani, Guoying Lu, Dow-Mu Koh, Thomas L. Chenover, Harriet C. Thoeny, Taro Takehara, Andrew Padhani, Brian B. Ross, Marc Van Calster, David Collins, Dima A. Hanjri, Gordon J.S. Rusth, Rami Yabl, and Peter L. Choyke); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Anwar B. Padhani); the Department of Radiology, University of Michigan, Ann Arbor, MI (Guoying Lu); the Department of Radiology, University of California, San Diego, San Diego, CA (Dow-Mu Koh); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Thomas L. Chenover); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Harriet C. Thoeny); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Taro Takehara); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Andrew Padhani); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Brian B. Ross); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Marc Van Calster); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (David Collins); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Dima A. Hanjri); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Gordon J.S. Rusth); the Department of Radiology, University of Toronto, Toronto, Ontario, Canada (Rami Yabl); and the Department of Radiology, National Cancer Institute, Bethesda, MD (Peter L. Choyke).

Abstract

On May 3, 2008, a National Cancer Institute (NCI)-sponsored open consensus conference was held in Toronto, Ontario, Canada, during the 2008 International Society for Magnetic Resonance in Medicine Meeting. Approximately 100 experts and stakeholders summarized the current understanding of diffusion-weighted magnetic resonance imaging (DW-MRI) and reached consensus on the use of DW-MRI as a cancer imaging biomarker. DW-MRI should be tested as an imaging biomarker in the context of well-defined clinical trials, by adding DW-MRI to existing NCI-sponsored trials, particularly those with tissue sampling or survival indicators. Where possible, DW-MRI measurements should be compared with histologic indices including cellularity and tissue response. There is a need for tissue equivalent diffusivity phantoms; meanwhile, simple fluid-filled phantoms should be used. Mono-exponential assessments of apparent diffusion coefficient values should use two b values (> 100 and between 500 and 1000 mm^2/sec depending on the application). Free breathing with multiple acquisitions is superior to complex gating techniques. Baseline patient reproducibility studies should be part of study designs. Both region of interest and histogram analysis of apparent diffusion coefficient measurements should be obtained. Standards for measurement, analysis, and display are needed. Annotated data from validation studies (along with outcome measures) should be made publicly available. Magnetic resonance imaging vendors should be engaged in this process. The NCI should establish a task force of experts (physicists, radiologists, and oncologists) to plan, organize technical aspects, and conduct pilot trials. The American College of Radiology Imaging Network infrastructure may be suitable for these purposes. There is an extraordinary opportunity for DW-MRI to evolve into a clinically valuable imaging tool, potentially important for drug development.

Neoplasia (2009) 11, 102–125

Address all correspondence to: Peter L. Choyke, MD, Molecular Imaging Program, NCI, Building 10, Room 1B40, Bethesda, MD 20892. E-mail: pchoyke@nih.gov
Received 14 October 2008; Revised 29 November 2008; Accepted 1 December 2008

Copyright © 2009 Neoplasia Press, Inc. All rights reserved 1522-8002/09/\$25.00
DOI 10.1593/neo.81328

Sections

1. Introduction
 2. Biophysical Basis
 3. Clinical View
 4. Neuroradiologic Perspective
 5. Oncologic Viewpoint
 6. Pharmaceutical Industry View on the Potential Role of DW-MRI in Drug Development
 7. Role of DW-MRI: A View from The Imaging Systems Industry
 8. Recommendations for Phase 1/2 Trials of Anticancer Therapeutics
 - 8.1. Diffusion-Weighted MRI Measurements
 - 8.2. Analysis of DW-MRI Data
 - 8.3. Correlation with End Points
 - 8.4. Data Display
 - 8.5. Standardization
 - 8.6. Validation
 - 8.7. Reproducibility
 - 8.8. Multicenter Methodology
 9. Software
 10. Recommendations for Future Methodology Development
 - Appendix 1. Nomenclature and Definitions
 - Appendix 2. Roadmap for Biomarker Development as Response Biomarker
 - Appendix 3. Measurement Error Methodology
 - Appendix 4. Diffusion Phantoms and QA Procedures
 - Appendix 5. Sample 1.5-T Protocols
 - Appendix 6. Acknowledgements and Workshop Participants List
- References
Figures

Introduction

Imaging biomarkers are important tools for the detection and characterization of cancers as well as for monitoring the response to therapy [1]. With rapid technological developments, new imaging methods appear rapidly and their utility requires systematic evaluation. Diffusion-weighted magnetic resonance imaging (DW-MRI) depends on the microscopic mobility of water. This mobility, classically called Brownian motion, is due to thermal agitation and is highly influenced by the cellular environment of water. Thus, findings on DW-MRI could be an early harbinger of biologic abnormality. For instance, the most established clinical indication for DW-MRI is the assessment of cerebral ischemia where DW-MRI findings precede all other MR techniques [2].

In oncologic imaging, DW-MRI has been linked to lesion aggressiveness and tumor response, although the biophysical basis for this is incompletely understood. Parameters derived from DW-MRI are appealing as imaging biomarkers because the acquisition is noninvasive, does not require any exogenous contrast agents, does not use ionizing radiation yet is quantitative and can be obtained relatively rapidly, and is easily incorporated into routine patient evaluations. However, these desirable features are offset by many challenges that face the validation of any imaging-based biomarker.

From the outset, it is important to recognize the many pioneering contributions of Stejskal and Tanner, Le Bihan et al., and Chenevert et al. from which as come the knowledge that tissue diffusivity mea-

surements are not always random due to tissue organization and therefore, have components that can be attributed both to the vascular and the extravascular compartments [3–5]. This weighting is imparted by the experimental conditions used in measurements (the b values). This document focuses on extravascular diffusion measurements where the measured signal is related to tissue cellularity, tissue organization and extracellular space tortuosity, and on the intactness of cellular membranes that are intrinsically hydrophobic. Classically, the low values of diffusion found in most tumors have been attributed to their increased cellular density; however, this remains a point of contention because diffusivity is influenced by extracellular fibrosis, the shape and size of the intercellular spaces, and by other microscopic tissue/tumor organizational characteristics such as glandular formations (as in well differentiated adenocarcinomas).

There is an extraordinary opportunity for DW-MRI to evolve into a clinically useful method that is useful for pharmaceutical drug development and for predicting therapeutic efficacy. Potentially, DW-MRI could have clinical utility at all stages of a cancer patient's journey from detection to diagnosis, for staging and assessing therapy response, and finally, for assessing relapse. As a pharmacodynamic indicator, DW-MRI may have a significant impact on pharmaceutical drug development. It should be recognized that pharmaceutical drug development and clinical therapeutic efficacy assessments are related but are nonetheless different. In pharmaceutical development, the questions revolve around whether a drug produces a measurable effect, on the magnitude of those effects and the potential biologic implications. In clinical trials, questions revolve around whether changes in individual patients can be measured reliably and reproducibly and whether they predict important clinical outcomes related to therapy.

To truly realize its potential, it is imperative for DW-MRI to become robust so as to provide similar information at different institutions using differing equipment. To date, no accepted standards in measurement or analysis methods have been established. Indeed, it is evident that current implementations of imaging protocols and analysis by different companies vary significantly; even the same manufacturer can alter its methods with "upgrades." There is also a lack of transparency and a divergent nomenclature among the vendors regarding their particular implementation of DW-MRI, which impedes efforts to standardize the technique. Furthermore, it is unclear whether whole body or localized analyses are preferred for clinical use and for drug development. Finally, multiple analytic approaches to DW-MRI have been proposed and it is not clear which is the best for any given situation. For instance, it is unclear what the optimal set of " b values" should be, how parameters should be adjusted with the tumor type and site, and whether the data should be analyzed monoexponentially, biexponentially, or multiexponentially, all of which will influence apparent diffusion coefficient (ADC) values. Diffusion-weighted MRI depends mostly on images obtained with b value >100 sec/mm² in extracranial tissues. However, it is acknowledged that the usefulness of information depicted by lower b values has not yet been fully investigated.

Our purpose was to attempt to summarize the current understanding of the pathophysiologic basis for DW-MRI imaging, to describe widely accepted methods that depict diffusivity in the extravascular-extracellular space, and to provide recommendations on standards for measurement, image display, and analysis methods. We have done this to bring all stakeholders toward a consensus on how to conduct multi-institutional trials that will assess the efficacy of DW-MRI for tumor assessments.

Emerging Challenges for Widespread Implementation of DW-MRI as a Means of Assessing Cancer

- Rapid evolution of body imaging protocols
- Divergence among and between vendors on data measurements/analysis and lack of transparency on how measurements are made
- No accepted standards for measurements and analysis
- Multiple data acquisition protocols depending on body part and usage of data
- Qualitative to quantitative assessments
- Lack of understanding of DW-MRI at a microscopic level
- Multiexponential decay components which affect the calculated ADC values
- Incomplete validation and documentation of reproducibility
- Divergent nomenclature and symbols
- Lack of multicenter working methodologies, accepted quality assurance (QA) standards, and physiologically realistic phantoms

Biophysical Basis

Diffusion measurements reflect the effective displacement of water molecules allowed to migrate for a given time [6]. Whereas temperature modulates molecular mobility in pure water (by approximately 2.4% per degree Celsius), it is rarely considered a significant factor in DW-MRI intact tissues because other biophysical properties have far greater influences on tissue water mobility. Signal-to-noise ratio (SNR) and T_2 -relaxation rates set practical limits on the diffusion measurement interval to within 40 to 80 milliseconds on standard clinical systems.

Using pure water at body temperature (37°C) as a reference standard, the average displacement of water molecules during a 50-millisecond interval is approximately 30 μm [9]. Because this is comparable to or greater than the dimensions of cells, there is a high probability that water molecules will interact with cells and their hydrophobic membranes and macromolecules will impede the motion of water. As such, the observed or "apparent" diffusion of water within tissues is typically several-fold less than in pure water. Moreover, diffusion in biologic systems is affected by water exchange between intracellular and extracellular compartments and the tortuosity of the extracellular space (which in turn is affected by cell sizes, organization, and packing density). Thus, although the spatial resolution of DW-MRI is typically on the order of millimeters, DW-MRI is exquisitely sensitive to changes in diffusion measured on the cellular scale (e.g., micrometers). A clear example of the ability of DW-MRI to document directional diffusion from which architectural features can be derived is the anisotropy depicted in highly directional structures (e.g., myelinated white matter fiber tracts) [3,7].

Other biophysical processes can potentially increase apparent water mobility. These include active transport, flow and perfusion, and macroscopic/bulk movements such as cardiac and respiratory motions. Flow, perfusion, and motion have detrimental effects on the accuracy of measurements of tissue water diffusion and its change over time. Motion correction, cardiac and respiratory gating, and overaveraging can reduce the magnitude of these effects.

Given these complexities, undisputed consensus on the appropriate biophysical interpretation of water diffusion measurements of *in vivo* systems is difficult to achieve. Indeed, some authors propose a "low-mobility intracellular space" and a "high-mobility extracellular space," which are averaged on clinical DW-MRI. Although the latter model has intuitive appeal, it is not well supported by empirical observations

[8,9]. Alternative models based on exchange between distinct diffusion domains that coexist in the same physical compartment offer better fits of empirical data, although the development and application of these models require measurements over a broader range of b values and more diffusion settings than are typically used in clinical settings. There are a limited number of validation tools available to confirm that properties, such as water diffusion, compartmental tortuosity, and interactions with cell membranes, actually influence diffusion measurements on DW-MRI.

Blood flow signal is rapidly attenuated at low b values (e.g., $b < 100\text{-}150 \text{ sec/mm}^2$) and may be mistakenly attributed to diffusion. This phenomenon, also known as the intravoxel incoherent motion (IVIM), has been used to assess tissue perfusion [5]; however, for clarity in this communication, we exclude blood flow and perfusion from diffusion phenomena and confine our comments to water mobility in the extravascular space.

Higher minimum b values are required to suppress the perfusion in vascular-rich tissues, indicating that appropriate minimum b values may vary across applications depending on intrinsic vascularity. Even after elimination of perfusion effects, tissues are known to exhibit multiexponential signal decay. Typically, very high b values ($b = 1000\text{-}5000 \text{ sec/mm}^2$) are required to reliably quantify the biexponential decay constants, "Dfast" and "Dslow" [10,11]. Alternatively, nonmonoexponential decay behavior may be fit to a stretched exponential model that yields a distributed diffusion coefficient and an index representing degree of intravoxel diffusion heterogeneity [12]. Therefore, the notion of "low" and "high" b values is relative and dependent on the tissue being studied and the SNR available to maximize diffusion contrast.

Physical Basis

- Diffusion-weighted MRI is sensitive to thermally driven molecular water motion, which *in vivo* is impeded by cellular packing, intracellular elements, membranes, and macromolecules. Thus, DW-MRI provides insights to cellular architecture at the millimeter scale.
- Sensitivity to diffusion-based contrast is primarily controlled by the b value with the appropriate b -value range dependent on tissue diffusion properties, SNR, and the need to suppress perfusion effects effectively at low b values.
- Tissues are known to exhibit multiexponential signal decay over a broad range of b values, indicating the need for a better understanding to elucidate the fundamental biophysical properties of water movement within the cellular matrix.

Clinical View

Diffusion-weighted MRI is already being incorporated into general oncologic imaging practice because of its many clinical advantages [13,14]. A particular advantage of DW-MRI is that it does not require intravenous contrast media, thus enabling its use in patients with reduced renal function. Its clinical uses include improved tissue characterization (differentiating benign from malignant lesions), for monitoring treatment response after chemotherapy or radiation, for differentiating posttherapeutic changes from residual active tumor, and for detecting recurrent cancer. Potential additional roles include predicting treatment outcomes (before and soon after starting therapy),

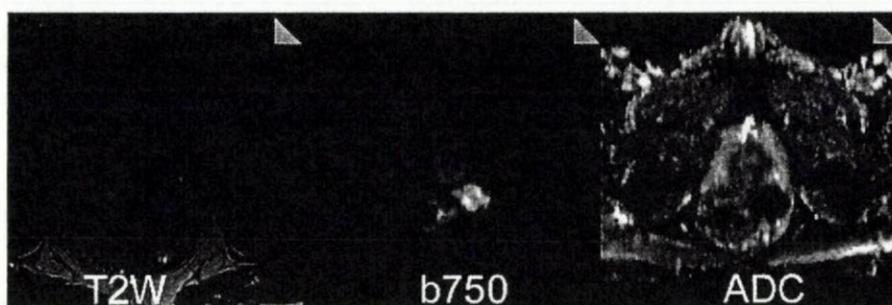


Figure 1. Restricted diffusion within rectal cancer with extension into the perirectal space. T2-weighted image demonstrate a well-circumscribed lesion in the perirectal space. Diffusion-weighted image obtained at a b value of 750 demonstrates a high signal, and corresponding ADC map demonstrates relatively restricted diffusion within the tumor.

for tumor staging, and perhaps also for detecting lymph node involvement by cancer.

Tissue Characterization

The reason(s) why malignant tumors have lower ADC values are poorly understood but is probably related to a combination of higher cellularity, tissue disorganization, and increased extracellular space tortuosity, all contributing to reduced motion of water (Figures 1 and 2). Correlations with cellularity have been found for some primary and secondary neoplasms [15–20] but not for all tumors such as adenocarcinomas and necrotic lesions that correlate only weakly [15,21]. Diffusion-weighted MRI is able to differentiate between benign and malignant focal hepatic lesions in many cases based on the higher ADC of benign lesions compared with malignant lesions [22]. However, when cystic, necrotic, or treated metastases are included, results in the liver are not as good [23]. In line with these findings, reduced ADC values of malignant breast tumors compared with those of benign lesions and normal tissue have also been noted [16,24,25]. Sumi et al. [26] showed that lymphomatous nodes had significantly lower ADC than benign nodes. However, in the same study, metastatic cervical lymph nodes in patients with head and neck cancers had significantly higher ADC values than benign nodes. This apparent discrepant result can be explained by the common occurrence of necrosis in nodes with metastatic squamous cell

carcinomas [26]. In a confirmatory study, the ADC values of lymphomas were reported to be lower than those of squamous cell carcinomas [27]. These and other studies indicate that false-positive results occur with abscess and infective processes and false-negatives occur with cystic, necrotic lesions and in well-differentiated neoplasms (particularly adenocarcinomas).

Monitoring Treatment Response

Because cellular death and vascular changes in response to treatment can both precede changes in lesion size, changes in DW-MRI may be an effective early biomarker for treatment outcome both for vascular disruptive drugs and for therapies that induce apoptosis [14,28,29]. Preclinical work has shown that DW-MRI is able to discriminate between nonperfused but viable and nonperfused, nonviable (necrotic) tissues, when tumors are treated with the vascular disruptive agent combrestatin-A4-phosphate [30]. In most malignant tumors, successful treatment is reflected by increases in ADC values. Rising ADC values with successful therapy have been noted in several anatomic sites, including breast cancers [31,32], primary and metastatic cancers to the liver [33,34], primary sarcomas of bone [15,35], and in brain malignancies [36].

Soon after initiation of therapy, transient decreases in ADC can also be observed; this seems to be related to cellular swelling, reductions in blood flow, or to extracellular space (the latter maybe

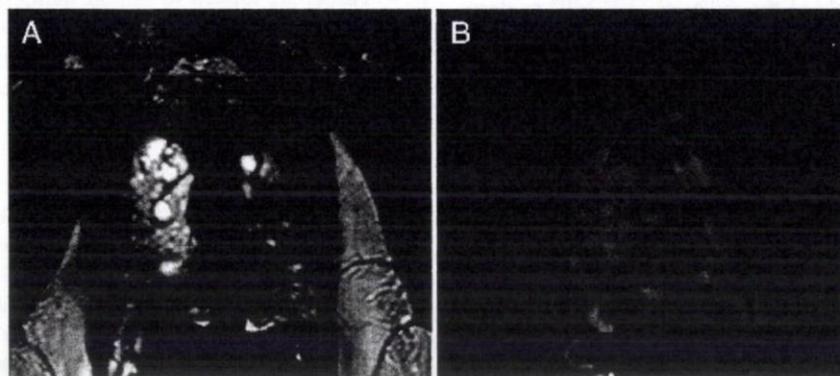


Figure 2. Restricted diffusion within a prostate cancer. T2-weighted scan demonstrates a well-circumscribed tumor in the left peripheral zone. The ADC map demonstrates restricted diffusion within the tumor.

mediated by vascular normalization if antiangiogenic drugs are given). For instance, it has been noted that anti-vascular endothelial growth factor therapies in brain tumors lead to an initial reduction in vasogenic edema that lowers ADC values [37]. The extent and duration of such ADC reductions is likely to depend on the type of treatment administered, tumor type, and the timing of imaging with respect to the treatment. Cellular swelling has been noted to occur in the early phases of apoptosis in response to anticancer treatment [14]. Apparent diffusion coefficient values can be reduced by fibrosis and dehydration after successful treatment as reported in rectal cancers [35] and in brain gliomas [37]. These additional observations indicate that ADC changes are dependent on complex interplays of biophysical processes, emphasizing the need to better understand at a histologic level, tissue changes reflected in ADC maps.

Posttherapeutic Changes Versus Recurrent/Residual Active Tumor

The differentiation of posttreatment changes and residual or recurrent tumor is a common diagnostic dilemma. In head and neck tumors, for instance, recurrent tumor and chondroradionecrosis are often impossible to distinguish clinically or by imaging [38]. Diffusion-weighted MRI has the potential to distinguish postradiation changes from recurrent cancer based on ADC value differences. Higher ADC values likely represent posttherapeutic extracellular edema, whereas lower values are suspicious for active disease. Simple visual assessments of signal intensity on high-*b* value DW images may be helpful for image interpretation, with hyperintensity associated with lower ADC values suggesting active tumor. In this respect, DW-MRI may have advantages over fluorodeoxyglucose-positron emission tomography (FDG-PET) assessments, which can be limited shortly after radiation therapy because areas of inflammation often have high uptake on PET scans and can lead to false-positive results [29].

Predictive Parameter for Treatment Outcome

Diffusion-weighted MRI has been shown to have the potential to prospectively predict the success of some treatments in a number of different tumors [30,33,36,38–43]. For example, strong negative correlations between pretreatment tumor ADC in patients with rectal cancer and size changes after chemotherapy and chemoradiation have been found [40]. This and other similar observations have led to the hypothesis that tumors with higher ADC levels are more likely to have areas of necrosis, which in turn predicts poor outcomes related to hypoxia-mediated radioresistance. This relationship between poor outcomes and high pretherapy ADC may not apply to all tumors and to all therapy types. For example, in an animal tumor model treated with a vascular disruptive agent, low tumor ADC values still had viable tumor cells on histologic diagnosis after therapy, whereas tumors with higher ADC values had a greater degree of cell kill [39].

Tumor Staging

There is early experience indicating that DW-MRI may be helpful for primary tumor staging [44] and for detecting nodal [26,27] and distant metastases [45]. In this regard, whole-body DW-MRI seems to be particularly promising [46]. Determining threshold ADC values and confounding effects that can allow/impe the differentiation of benign and malignant lymph nodes will be important goals for future clinical studies.

Clinical View

- In general, tumors have lower ADC values, whereas normal/benign/reactive tissues have correspondingly higher values.
- Apparent diffusion coefficient values for distinguishing malignancy from normal/reactive tissues and benign disease are dependent on histologic characteristics such as tumor type, differentiation, and necrosis.
- Threshold ADC values for the different tumor types and organs need to be defined taking into account data acquisition parameters.
- Diffusion-weighted MRI may be an effective early biomarker for treatment outcome for both antivascular drugs and therapies that induce tumor cell apoptosis. It is unclear how effective DW-MRI will be for monitoring the effects of other classes of anticancer therapies.
- Diffusion-weighted MRI protocols and analysis methods need to be tailored to individual tumor types, anatomic sites, and therapies.
- Successful treatment is generally reflected by an increase in ADC values; however, transient early decreases in ADC values can be seen after treatment.
- Complex interplays of biophysical processes are reflected in changes in ADC, underlining the need to define more clearly the acquisition and analytic protocols.
- The prognostic significance of pretreatment ADC values stems from the relationship between necrosis and poorer patient outcomes.

Neuroradiologic Perspective

Diffusion-weighted MRI has a number of roles in neuroradiology including the diagnosis of acute stroke as early as 30 minutes after the onset of ischemia [47], compared with the hours-to-days range for computed tomography and other MRI sequences. For this reason, DW-MRI has easily and quickly surpassed all other imaging techniques in the initial evaluation of acute stroke patients. In the acute stroke setting, DW-MRI can also be used with perfusion imaging to predict the likely clinical outcome after thrombolytic therapy and estimate the chances of secondary hemorrhagic transformation [48,Chalela et al,2007,83]. Other uses of DW-MRI include differentiating brain abscesses from other abnormalities that may mimic abscesses, with high sensitivity (96%) and specificity (96%) [49], with the exception of differentiating toxoplasmosis from lymphoma in HIV-positive patients [50]. Diffusion-weighted MRI can also be used for predicting the extent of neuronal damage after status epilepticus [51], for differentiating arachnoid cysts from intracranial epidermoid cysts, and for evaluating residual epidermoid tumor after surgical resection [52].

In the differential diagnosis of cystic brain lesions, DW-MRI can help distinguish abscesses from necrotic primary brain tumors such as high-grade gliomas, with lower ADC values usually detected in abscesses. Lower water diffusivity in abscesses is probably related to the presence of microorganisms, macromolecules, and intact inflammatory cells [49]. When intracranial masses are solid, the main determinant of diffusivity is the volume of the extracellular space. With a larger extracellular volume, which may be caused by edema/fluid accumulation, ADC values are higher, whereas tumor hypercellularity has the effect of restricting diffusion by decreasing the extracellular

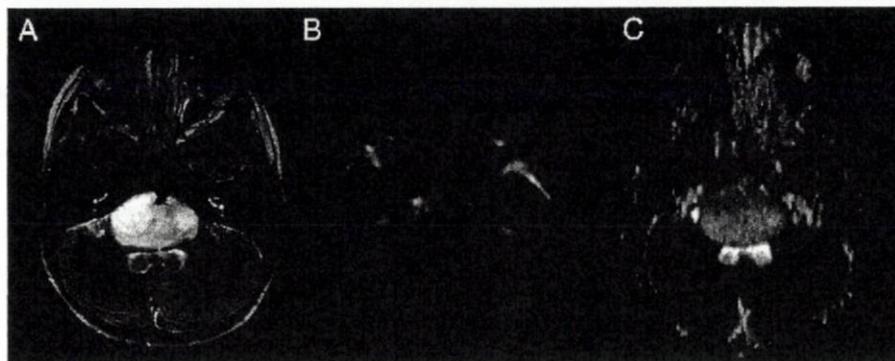


Figure 3. Minimally restricted diffusion in brainstem glioma: Axial T2-weighted image (A) at the level of middle-cerebellar peduncles shows slightly heterogeneous, predominantly hyperintense midline mass without significant surrounding edema. There is associated effacement of the fourth ventricle. Axial isotropic DW image (B) and ADC map (C) demonstrate no significant restricted diffusion. Instead, the ADC map demonstrates increased signal intensity compared with normal brain parenchyma, reflecting increased diffusion of water.

volume. Lower ADC values in lymphomas, compared with gliomas, correlate well with measures of cellularity [16] (Figures 3 and 4). Apparent diffusion coefficient values also correlate with tumor cellularity in astrocytomas [20,53], although the value of DW-MRI in tumor grading is still debated. Diffusion-weighted MRI has also been investigated as a biomarker of response to treatment in brain tumors, with increased diffusion values detected shortly after treatment initiation suggesting a favorable outcome [42,54,55].

Diffusion tensor imaging (DTI), which provides directionality to diffusion measurements, can be used to assess the relationships between tumor and nearby white matter tracts, potentially differentiating tumor infiltration of white matter tracts from displacement, which can be useful for preoperative planning [56]. Fractional anisotropy (FA), the degree to which diffusion is directed in a particular direction, is generally reduced in primary brain tumors owing to disorganized architecture resulting from neuronal death, axonal loss, and irregular tumor cellular growth [57]. Fractional anisotropy reductions also correlate with tumor cellularity and percentage tumor infiltration [58].

Neuroradiology View

- The main clinical application of DW-MRI in neuroradiology is in the diagnosis of acute ischemia.
- In the evaluation of brain tumors, DW-MRI reflects tumor cellularity and it can help differentiate brain abscesses from necrotic tumors.
- Diffusion tensor imaging can help assess the relationship between tumor and nearby white matter tracts, potentially differentiating tumor infiltration from displacement.

Oncologic Viewpoint

Diffusion-weighted MRI has the potential to assist in new drug development and in clinical practice. To be accepted in either area requires that DW-MRI be validated as an accurate biomarker. This will require systematically conducted prospective studies where patients are assessed by both DW-MRI, and standard criteria. Diffusion-weighted MRI will only be accepted if it is shown to provide accurate information earlier,

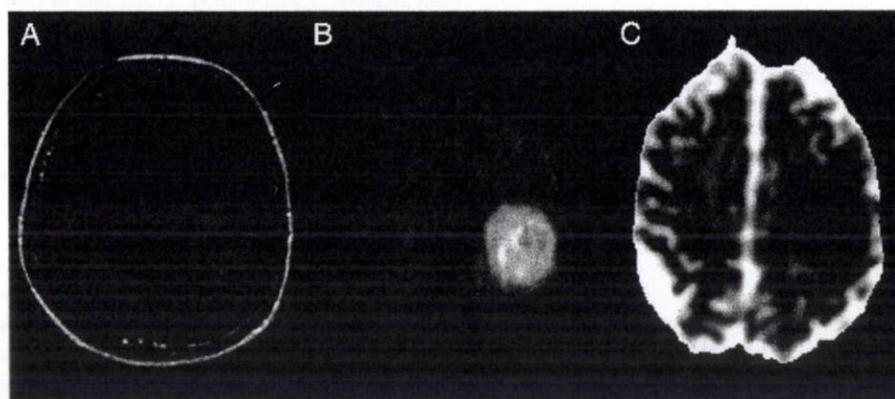


Figure 4. Restricted diffusion in CNS lymphoma: Axial enhanced T1-weighted image (A) of the brain shows a slightly heterogeneously enhancing hyperintense left frontal mass with minimal surrounding edema. Axial isotropic DW image (B) and ADC map (C) demonstrate restricted diffusion, suggestive of increased tumor cellularity.

quicker or easier than current methods, or can provide information unattainable with other modalities.

Potential in Clinical Drug Development

In clinical drug development, it would be important to see if a drug in a phase 1 study alters ADC and in which direction. It would be relatively easy to add DW-MRI to studies that were already performing serial MRI studies but it would seem sensible to look at earlier time points. This is because of the need to identify the time point of maximal response; it is only after defining this point correctly in phase 1 trials that DW-MRI could be incorporated into phase 2 and 3 studies where the opportunities to perform multiple repeat studies are limited. Furthermore, serial DW-MRI measurements that include early time points are essential to avoid false-negative results. Data from multiple studies with a wide variety of different agents incorporating DW-MRI would have to be available, before "go-no-go" decisions were made based on whether changes in ADC were seen.

If changes in ADC were seen in the phase 1 or early phase 2 studies of a new drug, it would be useful to incorporate DW-MRI measurements into phase 3 trials. Changes in ADC would then be correlated with Response Evaluation Criteria in Solid Tumors (RECIST)-defined responses, progression-free survival, and survival. If there were robust correlations between changes in ADC and one of the other efficacy end points, one could examine whether DW-MRI added value to the trial. One added value would be an earlier prediction of drug activity than is possible with other biomarkers. Another would be a better prediction of response than by standard criteria.

Care should be exercised in comparing responses according to RECIST and responses according to DW-MRI. For general use, it would be necessary to produce internationally accepted definitions for response/nonresponse on DW-MRI (these are not currently available). To produce such definitions requires analyzing numerous trials where DW-MRI results and standard end points are available. Correlation with survival or progression-free survival is far more important than correlation with RECIST response, which is just another surrogate marker of activity. If DW-MRI does predict response, it might indicate responses in some patients where RECIST does not and *vice versa*. If, for instance, a rise in ADC is associated with apoptosis, this effect might be associated with stable disease rather than partial response, yet such a drug might be valuable in extending survival. In such a situation, RECIST might indicate that a drug was inactive, but higher ADCs might be an early indicator that the drug is effective as maintenance therapy.

To validate DW-MRI as a biomarker, similar methods could be used as were used to validate the serum tumor marker CA-125. For example, rather than trying to correlate DW-MRI with RECIST response in individual patients, one could determine whether DW-MRI was able to classify the drug as active or not. Such an approach is particularly useful when different patients might be classified as responders by different techniques [59].

Another area that DW-MRI could have value in clinical trials is if it could predict which patients are most likely to benefit/not benefit from a drug/approach. Enriching a population to increase the proportion of patients benefiting is becoming increasingly important as drug development costs rise. It might, however, be difficult to determine an absolute ADC value that has predictive power (given the complexities of data acquisition and analysis methods particularly when examinations are done on different scanners).

Potential of DW-MRI in Individual Patient Management

A technique that can detect drug efficacy at an earlier time point has great potential in clinical oncology practice where there is a desire to stop ineffective therapy as quickly as possible especially if that therapy has severe adverse effects and is very expensive or where there are alternative approaches to treatment available. However, to change therapy based on a new technique requires having confidence about its accuracy. Large studies would be required to determine the sensitivity and specificity of DW-MRI for predicting activity in groups of patients. However, to change therapy on individuals also requires knowledge of measurement error. Depending on the alternative strategy to be used, a very high individual value in the positive predictive value/negative predictive value (PPV/NPV) would be required. Furthermore, it would be necessary to demonstrate that DW-MRI was able to provide information that was more accurate or was not available by other techniques and that it was widely available at a realistic cost. This aspect of DW-MRI development is illustrated in Appendix 2.

One could envisage that DW-MRI could give information about a tumor that would provide both prognostic and predictive information. The use of imitinib for gastrointestinal stromal tumors and trastuzumab for HER2-positive breast cancers are examples of the few situations where the activity of the new molecular targeted agents can be predicted by pretreatment investigations. There are no tests that predict which patients are most likely to benefit from new antiangiogenic, vascular disruptive agents, or other novel proapoptotic drugs. Only well-designed prospective studies will demonstrate whether DW-MRI can help predict which patients are most likely to benefit from such agents. Such studies might ultimately show that an absolute pretreatment ADC value below a specified range or an absence of change in serial ADC values once treatment has started is associated with a worse prognosis. However, the availability of yet another prognostic marker is unlikely to be widely used (given the higher cost of imaging tests compared with serology, for example) unless such information leads to changes in therapy.

Pharmaceutical Industry View on the Potential Role of DW-MRI in Drug Development

Many new therapeutics are presently entering development as a result of increased understanding of the molecular and genetic pathways controlling cellular function. Many of the structural and functional components of the cancer/host cell surface, cytoplasm, and nucleus are being explored for their potential value as therapeutic targets. Targeted molecular approaches typically seek to inhibit the cellular processes characteristic of the cancer phenotype. These new drugs may have complex and possibly even contradictory effects on water diffusion.

Central to the success of this process are the go-no-go decisions made in early clinical trials, an important aim of which is to reduce the high cost of pivotal (phase 3) trials. Because no single biomarker or assay is used to make such judgments, imaging will need to establish its place before being integrated into decision-making processes.

Strategic responses of the pharmaceutical industry to overcome current bottlenecks that result in delays to delivery of new drugs to market include using novel clinical trial designs and investments in novel technologies. Clinical trials are now being designed with the expectation of real-time data analysis and delivery. Imaging, together with other biomarkers, is recognized as providing accurate and reproducible data capable of enhancing decision making at critical milestones in the drug development process.

With this in mind, the important end points for imagers to consider include

- Early assessment of response
- Prognosis
- Survival estimate
- Recommendation of an optimal biologic dose
- Identification of the mechanism of drug action

Translational and early clinical development is well suited as a focus for DW-MRI in drug development. The nonionizing nature of DW-MRI and avoidance of contrast agents are conducive to serial patient examinations in exploratory early-phase drug trials. By their nature, early-phase clinical trials would allow the timely inclusion of preclinical DW-MRI information to select clinical imaging schedules and assess the likely clinical sensitivity and specificity of the technique. Key questions in early clinical trials include Has there been a change caused by the drug detectable by DW-MRI? What is the confidence that such a change has occurred? What is the magnitude of the change? What is the meaning or predictive value of the changes observed?

The limited sample size in early-phase clinical trials provides the opportunity for the rapid accumulation and interpretation of imaging data across many tumor types and therapeutics. Although this approach needs to be balanced against the limits of small sample size, studies should have a clearly stated aim(s) focused on one or more of the end points given above.

There is no reason to think that DW-MRI will be preferred for any specific class of drug. Any therapeutic could be investigated by DW-MRI, and the mechanism of action of a drug should not exclude consideration of this technique as long as the biology and sensitivity of the technique support its rational use. As a starting point, the drug mechanism(s) most likely to be detected by DW-MRI are those likely to alter the microenvironmental architecture (i.e., apoptosis and angiogenesis). The preclinical and evolving clinical experience of DW-MRI in clinical trials suggest its use as a pharmacodynamic biomarker (the manner in which the drug affects its intended target). Such data might provide insight into (1) dose scheduling in single or combination therapy and optimal drug formulation (e.g., oral *vs* intravenous administration).

Because DW-MRI has the potential to provide unique information for decision making, procedural rigor will be needed to establish it as a biomarker. Technical reproducibility needs to be determined to define significant thresholds of change in diffusion indices. Great emphasis needs to be placed on the need for reproducible examinations suited for the multicenter and global structure of early-phase clinical trials. In addition, there needs to be a rational basis for the choice of scanning times given the mechanism of the drug being assessed. Acquisition sequences, data transfer, and analyses all need to be standardized to allow future meta-analyses and comparisons of results.

Pharmaceutical Perspectives

- Diffusion-weighted MRI is one of several technology-based tools that could support decision making in the drug development process, the aim being to increase the likelihood of success in later more costly trials.

(continued)

- Information from DW-MRI has to be placed in the context of other information from the clinical trial and the relative value of DW-MRI data compared with other biomarkers needs to be established. Diffusion-weighted MRI can be used for more than simply making go-no-go decisions. If DW-MRI was found to be providing information about patient outcome then its value could be increased.
- As the pharmaceutical industry moves toward complex multi-targeted therapies, the anticipated effects on ADC measurements become more difficult to predict. It is likely that the imaging protocols and analyses will need to be highly customized to the individual drugs/combinations.
- Improved information is needed about the sensitivity, specificity, biologic meaning, and predictive value of DW-MRI before it will be routinely implemented in drug development environment. Such information may be gained from appropriate patient stratifications in well-conducted clinical trials, using standardized protocol acquisitions, central data collections, and analyses and assessments of intrapatient, intersite, and intermanufacturer scanner reproducibility.

Role of DW-MRI: A View from The Imaging Systems Industry

Diffusion-weighted MRI clearly differentiates itself from other imaging modalities both within the MRI "space" and outside as the only imaging modality able to depict water movement at a cellular level. Evaluations of the value of DW-MRI as a "contrast mechanism" leading to new clinical applications could open up new market segments encouraging more sales for MRI equipment and innovations in this area. If DW-MRI can be shown to be a robust and reproducible early biomarker of response to anticancer therapies, which is useful for drug development or for making patient decisions, then this would allow the installed MRI market to increase further. Improved understanding of the mechanisms that determine tissue diffusivity and ADC changes in response to therapy would encourage more rapid dissemination of the method.

To enable such developments to occur, it is absolutely necessary that there are agreements among all stakeholders on standards for both acquisition protocols, repeatability/reproducibility and for the postprocessing procedures, to ensure that quantitative ADC values have similar meanings across vendors and institutions. The setting down of such standards in cooperation of the scientific imaging community will allow vendors to focus their research and development resources on improving measurement and analysis methods.

Imaging System Industry View

- Diffusion weighting is unique to nuclear magnetic resonance and helps MRI to differentiate itself from other imaging modalities.
- Because the technique is quantifiable and can be repeated easily, new applications have the potential to open new market segments especially to the pharmaceutical sector but also for clinical trials and screening.
- Standardization will allow the vendors to focus their research and development resources on improving measurement and analysis methods, repeatability, and reproducibility.

Recommendations for Phase 1/2 Trials of Anticancer Therapeutics

Diffusion-Weighted MRI Measurements

Recognizing that there are several DW-MRI data acquisition techniques, the most commonly implemented basis sequence is the single and double spin-echo Stejskal-Tanner echo planar image (EPI) experiment [60]. The following comments apply to clinical imaging at 1.5 and 3.0 T but may vary and need adjustment according to field strength. To ensure high quality images for both qualitative and quantitative assessments, scanning factors should be optimized to maximize SNR and reduce artifacts (e.g., from motion, incomplete fat suppression, residual eddy currents induced by diffusion gradients, and EPI-related artifacts). In the body, DW-MRI can be performed using breath-hold, free breathing, or respiratory/cardiac-triggered techniques as dictated by specific anatomic locations [61]. Scanning parameters should be prescribed to allow accurate and reproducible ADC quantification, and the chosen parameters should ideally be achievable across MR platforms to allow meaningful comparison of results. The scanning parameters should be clearly stated in reports and manuscripts.

Ensuring high-quality DW-MR images. The following summarizes the key factors that would help to optimize image quality [61].

Optimizing SNR

1. Parallel imaging should be used in the body as this shortens echo-train lengths and reduces susceptibility and magnetic field inhomogeneity-related artifacts.
2. Multiple averaging commensurate with field strength should be used. The ability to acquire additional averages at higher b value (e.g., $>500 \text{ sec/mm}^2$) can be advantageous to compensate for reductions in SNRs at high b values.
3. The echo time (TE) should be held as short as achievable (typically 50-90 milliseconds at 1.5 T and even shorter at 3 T).
4. Tetrahedral encoding or other simultaneous application of gradient schemes (e.g., three-scan trace (Siemens), gradient overplus (Philips)) should be used to achieve the shortest possible TE, which also reduces the effect of eddy current-induced image distortions.

Reducing artifacts

The following imaging techniques can help reduce motion artifacts

1. Parallel imaging in the body reduces the time of acquisition and phase encoding and it should be used wherever possible.
2. The shortest TE achievable should be used.
3. Multiple averaging or breath-hold single shot techniques.
4. Respiratory triggering [62] and/or cardiac gating when imaging the chest and upper abdomen.
5. Antiperistaltic agents for examinations of the lower abdomen and pelvis.

Fat suppression

1. Optimized fat suppression techniques should always be used in the body to reduce ghosting artifacts. When performing DW-MRI over large FOVs on a 1.5-T system, short-tau inversion recovery

(STIR) may be more useful than other methods in achieving uniform fat suppression.

2. For examinations targeted to specific organs or anatomic regions, spectral spatial fat saturation (e.g., spectral presaturation attenuated by inversion recovery (SPAIR), Chemical Shift Selective), can be advantageous [45]. Considerations should also be given to using water excitation as a method for fat suppression.
3. Owing to B1 field inhomogeneities at 3 T, optimal fat suppression is more challenging, and frequency-selective spectral fat saturation may be more successful.
4. Where available, use advanced or higher-order shimming techniques.

Eddy currents and EPI-related artifacts

1. Eddy currents related to diffusion gradients and EPI techniques lead to geometric distortions and image shearing, which can be reduced by increasing readout bandwidths or reducing the echo spacing, conversely reducing the length of image readout. For body imaging, this is typically 1 to 2 kHz per pixel. However, increasing bandwidths increases noise and Nyquist ghosting, and so bandwidths or echo spacing settings should be optimized.
2. Patients who have recently had (within 7 days) superparamagnetic iron oxide particles as diagnostic agents may not be good candidates for DW-MRI owing to susceptibility induced by the presence of the iron oxide in some tissues such as the liver. However, ultrasmall superparamagnetic iron oxide particles, which are also taken up in normal lymph nodes, may improve specificity of DW-MRI by suppressing signal from normal but not pathologic nodes.

Diffusion-weighted MRI for qualitative assessment. To maximize tumor visualization and characterization, adequate suppression of background signals arising from normal tissue is desirable. Diffusion-weighted MRI should be performed with sufficient degrees of diffusion weighting (by appropriate choices of b values), with considerations given for the anatomic region, tissue composition, and pathologic processes as indicated in Table 1. This may require the customization of DW-MRI protocols for different tumor types and tumor locations.

Both native high b value DW-MR images and the ADC maps are useful for visual assessments of DW-MRI data; both should be evaluated with corresponding morphologic images. Cellular tissues generally demonstrate high signal intensity on high- b value images, but yield low ADC values. This pattern is occasionally seen in viscous fluids also such as postoperative cavities and abscess emphasizing the need to undertake correlative imaging with other morphologic MRI sequences including those using contrast medium enhancements.

Table 1. b Values That May Be Appropriate for Qualitative Tumor Evaluations.

b Values (sec/mm^2)	Anatomic Regions
>1000	Brain, prostate, uterus, cervix, lymph nodes
750-1000	Brain, neck, breast, chest, general abdomen (including colorectal, pancreas, kidneys, peritoneum), lymph nodes, pelvis (including prostate, uterus, cervix, ovaries, bladder). Also for whole-body imaging (diffusion-weighted whole-body imaging with background body signal suppression)
100-750	Liver (primary and metastatic disease)
<100	Used for the detection of liver lesions ("black blood technique")

Conversely, cystic or necrotic tissues show greater signal attenuation on high- b value images and have high ADC values (in this context, the definition of "necrosis" needs to be established). The latter pattern is sometimes also seen in well-differentiated neoplasms.

Interestingly, some normal and pathologic tissues exhibit high signal intensity on high- b value DW-MRI and also return high ADC values. In these cases, the high signal intensity observed on DW-MRI cannot be attributed to limitations in water diffusion but result from the intrinsically long tissue T_2 -relaxation times; an effect known as " T_2 -shine through." In some anatomic regions (e.g., prostate gland), ADC maps may be more helpful for disease detection because T_2 -shine through from the normal peripheral zone and may mask disease even on high- b value DW-MR images.

Fibrosis may appear low in signal intensity on DW-MRI and return low ADC values, but the range of the imaging appearances of fibrotic tissue has not been fully characterized.

Table 1 summarizes b values that may be used as a guide when performing DW-MRI for qualitative assessment [18,22,39,63-69]. For some tissues (e.g., prostate, lymph nodes), b values >1000 sec/mm^2 are occasionally needed to mitigate the effects of " T_2 -shine through."

Diffusion-weighted MRI for ADC quantification. Meaningful comparisons of DW-MRI from different imaging centers with data acquired from different platforms are more likely to be realized if DW-MR images are quantified.

Apparent diffusion coefficient quantification obtained using breath-hold DW-MRI is less reproducible compared with ADC obtained using free breathing techniques. However, ADC values obtained using free breathing techniques mask tissue heterogeneity owing to partial volume averaging effects. The relative balance between these two trends may vary in different parts of the body thereby necessitating different approaches.

In the absence of conclusive data demonstrating the superiority of ADC derived from either breath-hold or free breathing techniques for assessing treatment response, both techniques should be investigated.

The following should be considered when performing DW-MRI for ADC quantification:

1. Imaging should cover the entire target lesion.
2. The target lesion should not be chosen in areas prone to susceptibility, ghosting, and other EPI-related artifacts (e.g., base of the neck, left lobe of the liver, adjacent to the heart, adjacent to metallic surgical sutures/prosthetics/seeds, or across areas with visible ghosting artifacts). Target lesions can include bony lesions (sclerotic or lytic).
3. Target lesions should measure at least 2 cm in the minimum diameter.
4. There is controversy about whether partially necrotic lesions should be excluded from evaluations. This is because some therapies seem to work better in tumors that have higher ADC values (e.g., vascular targeting agents) [39]. Other treatments (such as chemotherapy and radiation) seem to work better with more cellular tumors. These observations underline the need to prospectively define tumor "necrosis" in clinical trials.
5. For ADC quantification in the liver or upper abdomen, consider performing measurements in the fasting state to minimize air in the stomach, which may degrade images of the left hepatic lobe.

6. The repetition time (TR) should be sufficiently long to avoid T1-saturation effects, which result in underestimation of ADC values.
7. The TE should be kept as short as possible as this results in better SNR and it minimizes motion and susceptibility EPI-related artifacts.
8. Three or more b values should be used. This should include $b = 0$ sec/mm^2 , a b value of ≥ 100 sec/mm^2 , and a higher b value of ≥ 500 sec/mm^2 (typically 500-750 sec/mm^2). b values of >1000 sec/mm^2 are less optimal for ADC calculation because of poorer SNR at very high b values, unless the aim is to evaluate the exponential behavior of the signal attenuation.
9. The choice of three b values as recommended above should enable calculation of perfusion-insensitive ADC values (by excluding the $b = 0$ sec/mm^2 image from the ADC calculation). The degree of perfusion bias in ADC measurement increases with volume fraction of flow and decreases b -value range. For example, consider tissue having a true ADC = 1.0×10^{-3} mm^2/sec . If the volume fraction of flow is 0.05 (i.e., 5% of signal from flowing blood), the apparent ADC measured between points $b = 0$ and 1000 sec/mm^2 is overestimated by approximately 5.1%. If the flow fraction increases to 10%, the measured ADC is inflated by 10.5%. Moreover, if the measurement is made between $b = 0$ and 500 sec/mm^2 , the perfusion-induced ADC overestimation increases to 21%. It is therefore recommended to have low a b value sufficient to extinguish the flow signal (e.g., low $b \geq 100$ sec/mm^2) while maintaining adequate diffusion contrast between low- and high- b value settings. An obvious drawback of this requirement is additional scan time because even moderately low b values should be acquired along three orthogonal axes to avoid anisotropy effects. Fortunately, this penalty is usually not too severe when using single-shot DW-EPI techniques. On MR systems incapable of performing measurements using >2 b values, imaging could be performed using just the two higher b values.

Recording of scanning parameters. To ensure maximum transparency, all pertinent scan parameters should be recorded and clearly stated in reports, manuscripts, and other scientific publications. This would facilitate investigators replicating the imaging technique on similar platforms, translating techniques onto other imaging systems, and enabling comparisons of ADC values between vendors.

Reporting of Imaging Parameters Should Ideally Include the Following:

Field strength, MR system and model, gradient performance, software version, field of view (FOV), matrix size, technique (e.g., breath-hold, free breathing, respiratory-triggered), type of imaging sequence, TR, TE, number of partitions, section thickness, number of averages (including the use of high b values averaging), fat suppression technique, choice of b values, use of tetrahedral encoding, receiver bandwidth, duration of application of diffusion gradient (Δt), and time between application of diffusion-gradients (Δt)

Accurate recording of imaging parameters may then be used to retrospectively evaluate the relative value of techniques for optimizing DW-MRI. Many of these parameters are already recorded in DICOM header information and are retrievable. Obtaining access to this header

information is critical for accurate image analysis. Magnetic resonance imaging vendors should support access to this information in a move toward greater transparency/standardization.

Diffusion-Weighted MRI Measurements

- Diffusion-weighted MRI should be optimized to maximize SNR and to reduce artifacts that arise from motion, eddy currents, ghosting, and susceptibility effects.
- For visual qualitative analysis, DW-MRI should be performed using b values, which result in sufficient background suppression to allow signal intensity differences in target tissues to be observed.
- For ADC quantification of a target tissue or lesion, two b values should be used (with one b value ≥ 100 sec/mm² and the other ≥ 500 but ≤ 1000 sec/mm²) with a monoexponential decay.
- There should be transparency in accessing and recording scan parameters on all MR platforms in a move toward greater standardization.

Analysis of DW-MRI Data

Analysis methods. For many clinical applications, single b -value DW-MRI at relatively high diffusion weighting offers exceptional sensitivity to detect disease (this is evident from experiences in the brain for the early detection of stroke (on high- b value images) and in the liver for the detection of lesions using "black blood" low- b value images). However, image signal analysis from single b -value images is inadequate for even rudimentary quantitative analysis of water mobility in tissue.

Multiple b values are necessary to calculate the ADC. At least two b values are needed for basic ADC calculations and this can be done on most clinical systems. Implicit in two b -value ADC calculations is the application of a monoexponential decay model. That is, an ADC map is generated by the natural logarithm of the ratio of low- b value over high- b value image, scaled by inverse of the b -value difference. Albeit overly simple, this two-point method is adequate in instances where multiexponential features are negligible over the acquired b -value range. Moreover, the adoption of this basic analysis has led to reasonable agreement across centers and MRI vendors for ADC quantification of the human brain (b -value range 0-1000 sec/mm²). Many observers believe this may be sufficient for clinical usage, although it is likely that physical measurement of water diffusion in tissue is more complex than described by the monoexponential decay model.

In highly vascular tissue, blood flow/perfusion may impart significant signal attenuation over the low b -value range (from $b = 0$ to $b = 100$ sec/mm²), which artificially inflates diffusion estimates. As described above, nonzero lower b values should be used to eliminate vascular contributions to the calculated ADC. The minimum b value threshold to suppress perfusion effects will depend on the vascular properties of tissues, although for most applications, a lower b value of 100 to 150 sec/mm² is probably adequate.

It is recommended to also continue to acquire the nominal " $b = 0$ " image to provide anatomic information and to maintain consistency with prior work. Usually, the $b = 0$ image can be obtained at nearly no cost in scan times using single-shot techniques, particularly be-

cause acquisition along three-orthogonal axes is not performed for the $b = 0$ weighting.

For applications where DW-MRI is acquired over larger ranges of diffusion sensitivities, and assuming perfusion effects have been effectively removed by the proper choice of the lower b value, simple monoexponential models may not adequately characterize the decay curve. Usually, evidence of true multiexponential features (not related to perfusion effects) requires substantially higher b values (e.g., 2000-6000 sec/mm²), much greater than is typically acquired in clinical studies owing to practical SNR limits. Proper analysis of these data types requires multiexponential models where signal decays are modeled as weighted sums of two or more exponentials (provided that the signals at the highest b value are above the noise level) [10,11] or alternative models such as stretched exponentials that allow a distribution of diffusion coefficients in each voxel [12]. As with other curve fitting challenges, reliability to accurately isolate multiple decay coefficients depends on the difference between the true D_{fast} and D_{slow} , SNR, b -value range, and number of b values acquired. Rejection of low SNR pixels and/or incorporation of SNR weights in the multiexponential fitting routine should be used to mitigate fitting errors. An unfortunate tradeoff in acquisition of DW-MRI over many b values and/or averaging to increase SNR to support multiexponential diffusion analysis is the commensurate increases in scan times which may not be practical in many clinical settings.

Diffusion in some tissues is known to be directionally dependent, that is, anisotropic (e.g., in the central nervous system and in muscle). If it is known *a priori* that the tissue of interest is isotropic (e.g., most tumor models) then a single gradient direction is usually sufficient to properly document diffusion properties. In general, however, it is safer to assume the lesion of interest and its surrounding tissues may have directional dependencies so it is best to measure water mobility along at least three orthogonal diffusion gradient directions yielding, say, ADC_x , ADC_y , and ADC_z . The simple average of these into a mean diffusivity value effectively removes confounding influences of the relative orientation between tissue and the imaging system. This mean diffusivity bears the same desirable rotational independence as the trace of the full diffusion tensor without having to acquire or process DTI [70].

If further information is specifically desired regarding the strength and spatial patterns of anisotropy, at least six gradient directions are required to generate the full diffusion tensor, although additional gradient directions (9-32 commonly) generally improve the quality of the tensor analysis results.

Most MRI vendors offer the option to acquire and process DTI scans in a reasonably efficient manner. Intrascan image registration should be applied if there are systematic shifts and image distortions at various gradient directions before tensor analysis. Multiple indices are available to quantify the degree of anisotropy (e.g., FA, relative anisotropy [70]). In addition, the direction of the strength of anisotropy can be color-encoded using the principal eigenvector of the diffusion tensor. Furthermore, the connectivity of anisotropic domains can be represented in tractography, which allows visualization of tissue fiber tracts in three dimensions [56]. As suggested above, diffusion anisotropy is relatively strong in the CNS. Outside the CNS with the exception of the kidney and muscle, however, anisotropy is rather modest, and therefore, most tumor analyses have been directed toward isotropic diffusivity indices (i.e., ADC calculations).

Regions of interest definitions. To study diffusion properties of tumor, proper delineation of lesion boundaries must be identified for subsequent quantification. Ideally, the region of interest (ROI) is contoured around lesions using images with the highest contrast between lesion and normal tissue. Subjective placement of smaller ROIs within lesions is not recommended particularly for response assessment studies.

Traditional high-contrast anatomic images, such as T2-weighted and contrast medium-enhanced T1-weighted, which are independent of the DW-MRI sequences are preferred, but translation of ROIs to the DW-MR image set is then required. Transfer of such ROIs to the DW-MRI data set requires image registration unless prescription of the traditional and DW-MRI scans was identical (ignoring for the moment other systematic distortions). In some instances, the DW images themselves can offer strong lesion/tissue contrast, in which case these are sufficient for ROI definition. Ideally, the b_0 T2-weighted image (or a very low b value image) should be used, although, occasionally, higher b -value images may have to be used.

There is debate as to which b -value image best delineates tumor from normal tissue/necrotic tissues. When ROIs are drawn on high- b value images for the estimation of ADC values, such ROIs are said to represent "viable tumor" because the detrimental effects of necrosis are ameliorated. However, such a method for defining ROIs is occasionally prone to error because of T₂-shine through effects. Furthermore, in the presence of necrosis/cystic structures, lesion extent may be underestimated. It is also important to remember that well-differentiated tumors may not be seen on high-value images. Whatever the method used to define ROIs, a standard, recorded strategy should be applied to ensure consistency within any given study.

In the ADC calculation methods described above, low SNR pixel values should be eliminated before the ADC map calculation, and these pixels should be flagged as "not-a-number" for exclusion. However, elimination of low SNR pixels for ADC calculations and/or using high- b value images for ROI definitions can be problematic when evaluating therapeutic effects of some drugs. For example, chemotherapy for teratoma can cause a poorly differentiated tumor to become well differentiated (a favorable outcome measure) and some drugs/therapies induce necrosis and cystic degeneration. In both cases, ROIs placed solely on areas of "viable tumor, however, defined" would underestimate/mask favorable therapeutic effects. Pixel counting of zero ADC values before and after being induced by therapy would be a way of dealing effectively with these issues.

Conservative ROI definitions would only include apparently viable tumor based on robust Gd contrast enhancement on T1-weighted images. A more generous tumor extent would include contrast-enhanced and hyperintense tissues on T2-weighted images. However, inclusion of necrotic and cystic zones can include extremes in water ADC values, which may adversely bias image analysis. It is important that standardized software be developed in which criteria of undesirable tissue be clearly defined and that individual subjective decision making by observers is kept to a minimum. Different scenarios may be adopted to exclude these nonviable tissue regions. It should be kept in mind that a particular treatment might induce more nonviable voxels, which, if eliminated from analysis, would falsely reduce the apparent impact on ADC.

The entire three-dimensional volume of interest (VOI), a composite of ROIs over multiple slices, of the lesion should be delineated particularly if the tumor is being followed over time.

Heterogeneity assessments. Volume of interest analyses methods fall into three general areas: whole-tumor summary statistics, histogram, and voxel-wise analyses.

- (1) Whole-tumor summary statistics (e.g., mean and median) are a common method for reduction of the tumor into a single quantity. The advantage of this technique is its simplicity, although it fails to fully address the important issue of tumor heterogeneity.
- (2) The histogram-based approach subclassifies different tumor diffusion environments [32]. For example, the volume of the tumor within a specified range of the diffusion histogram has been investigated as an approach to document tumor evolution in response to treatment. It is vital that consistent "binning" procedures be used to analyze data across a multi-institutional trial because errors in histogram comparisons have occurred where this was not standardized [71]. When using histograms, redistributions of high and low values may occur, and this may not be reflected in mean changes. This occurs because histograms cannot depict the spatial information as to the origin of changes, which may be crucial to detect the evolution of diffusion changes within lesions. Ideally, advanced histogram analysis techniques should be used so that a single scalar value can be used for developing response criteria on DW-MRI; an example approach is principal component analysis [71].
- (3) Ideally, to track the spatial origin of changes induced by treatments, it is necessary to have spatial tags to accurately monitor the change in diffusivity [55,72]. The retention of spatial information requires voxel-wise approaches incorporating registration of image data sets, typically between treatment interval examinations. This method seems to work well in the brain and in bones but is less likely to be applicable to whole-body measurements owing to problems associated with image registration and changes in lesion sizes with therapy [55,73]. This approach enables ADC in individual tumor voxels to be followed and so enables the depiction of the fractions exhibiting a significant change (increase or decrease) in ADC. The spatial location of ADC changes within the tumor can be made available to potentially guide spatially directed therapies.

Summary

Choice of monoexponential *versus* multiexponential modeling of signal decay with b value depends on features apparent in the data, SNR, number, and range of acquired b values.

Data typically obtained in most clinical applications for b -value ranges of 100 to 1000 sec/mm² are reasonably well modeled using monoexponential decay fits.

Tumor ROI/VOI definitions may be done on traditional high-contrast images such as T2-weighted or T1-weighted contrast-enhanced images. High-image contrast, high- b value DW images can also be used.

Descriptions of diffusion properties within lesions or tissues of interest may be reported at several levels classified as follows:

- 1) traditional summary statistic over the entire ROI/VOI;
- 2) histogram analysis, which allows segmentation of the tissue based on diffusion properties; and
- 3) voxel-by-voxel analyses where spatial information is retained over interval examinations such that fractional volume of tissue exhibiting change in diffusion properties is measurable. However, the latter requires methods of tracking individual voxels over time.

Correlations with End Points

The determination of outcome measures or end points must be dictated by the nature of the question being addressed (clinical, biologic, physical, or pharmaceutical). For instance, if the purpose of a trial is to determine whether DW-MRI can characterize the biologic aggressiveness of a tumor, then ADC values need to be correlated with recognized measurements of aggressiveness. This could include tumor grade, time to progression, progression-free survival, or overall survival. However, if the goal is to determine whether DW-MRI is an early marker of treatment success, then intermediate end points such as pathologic response could be used. Potentially, DW-MRI results could be compared with other biomarker changes such as serum markers of cancer (e.g., carcino embryonic antigen, prostate-specific antigen, etc.), RECIST, and WHO measurements of tumor size. However, firmer and more robust end points reflecting therapy efficacy in patient outcomes are preferred where possible, such as time to progression, progression-free survival, and overall survival.

If DW-MRI is being evaluated as an early biomarker of therapy response then the timing of follow-up studies should be such that DW-MRI is acquired before changes in size are expected to occur. Intermediate time points may become influenced by necrosis and liquefaction, and DW-MRI may become less useful. Long-term data points may "normalize" because liquefactive necrosis resolves and the residual mass contains fibrotic dehydrated tissues.

For response assessment studies, it is important to have predetermined whether DW-MRI changes are expected to occur, the magnitude/direction of the likely change, and the timeline as to when and for how long changes are expected to last. Animal validation studies before human studies may provide information on the appropriateness of using DW-MRI and on the optimal timing for doing imaging in human studies.

Data Display

Diffusion experiments generate large numbers of magnitude b -value images. When these are combined with morphologic images,

many hundreds of images are produced, which need to be reduced for diagnostic interpretations.

The most valuable images required for interpretation are high- b value images and ADC maps, which should always be evaluated with morphologic imaging. Because high- b value DW-MR images have high background suppression, tumor localization is usually straightforward. However, very high signal on high b value may also be due to T2-shine through effects; conversely, liquefaction or necrosis can result in an underestimation of lesion extent, so comparisons with anatomic images are important.

Although no color scales are especially suited for the display of high- b value magnitude images, convention has it that "inverted grayscale" be used (ADC maps however, are better displayed using conventional grayscale). Indeed, whole-body DW imaging with background suppression can produce images that superficially resemble FDG-PET scans (Figure 5). This is because of the high contrast on high- b value images, which, when used with three-dimensional displays, are amenable to multiplanar reconstructions and three-dimensional renderings (maximum-intensity projections, surface shaded display, volume renderings).

A common method of analyzing high- b value images is to use fusion imaging techniques. Modern three-dimensional fusion imaging visualization software works in three steps. (1) **Superimposition:** data sets do not need to be acquired in the same plane and to have identical FOVs and matrix sizes, but most ADC data sets are aligned and obtained with similar parameters. (2) **Alignment:** algorithms work with multiple degrees of freedom (translation and rotation) based on anatomic landmarks with the ability to work automatically with manual overrides if necessary. (3) **Visualization:** blending of grayscale with pseudo color images with adjustable balance between the two superimposed data sets. When blending is used for data display, the level of blending should be kept constant across a study and reported in manuscripts.

Other potential artifacts appearing on fused images include misregistration of anatomic and DW images due to bladder filling and internal organ including movements. Susceptibility artifacts caused by luminal air are exaggerated on high- b value images, although their effects are minimized on ADC maps.

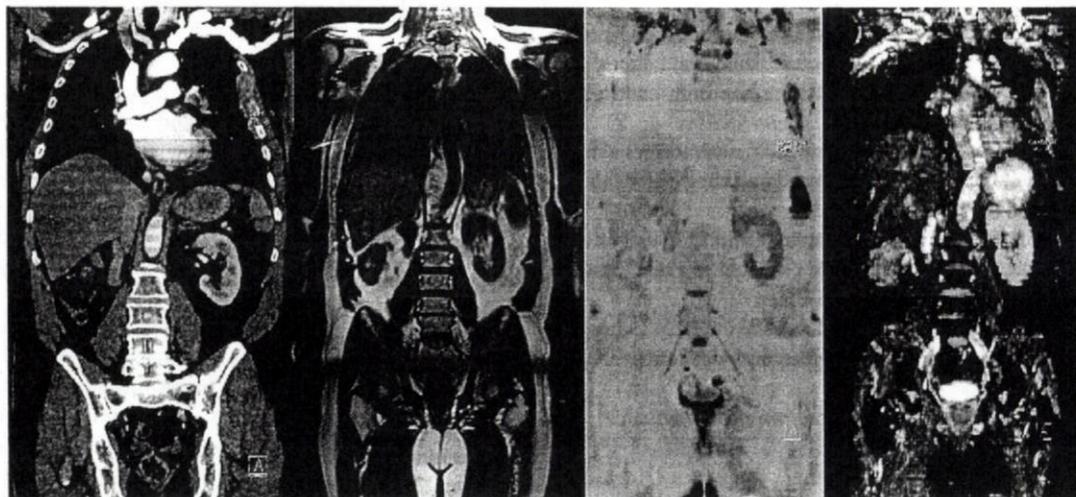


Figure 5. Metastatic renal cancer. Whole-body DW image demonstrates (left to right). Coronal computed tomographic image, coronal T1-weighted MRI at the same plane of section, DW imaging (b800) demonstrating left chest wall metastasis and the fused ADC and T1-weighted MRI.

Standardization

A major challenge to the widespread implementation of DW-MRI is the lack of a standard approach to data collection and analysis. This creates challenges for support of DW-MRI by commercial MRI vendors and makes deployment of DW-MRI techniques limited to sites with significant experimental MRI expertise. Furthermore, the lack of standard approaches impairs validation and makes the ultimate qualification of DW-MRI as a biomarker extremely difficult.

In large part, the lack of standardization is related to the technical challenges in performing DW-MRI acquisitions. In most practical applications of DW-MRI, performing "ideal" data acquisitions is impractical owing to limits in technology and patient compliance.

Approaches that accommodate technical limitations through compromises in acquisition and/or in analysis have been developed to allow the practical implementation of this technique. Examples include reducing the number of *b* values for modeling of data, reducing spatial resolution, limiting volume of imaging, averaging free breathing studies instead of gating, using empiric analyses (e.g., visual assessments signal intensity of high-*b* value images), creative acquisition time reducing techniques, and so on.

Standardized data sets should be acquired systematically using "ideal techniques" with great intrinsic redundancy to test the effect of various technical compromises on measuring the signal associated with response. Such data should be made widely available for investigators to test their analytic software. These ideal data sets should be limited to single organs/single treatments starting with the least challenging. Ideally these should be documented, anonymous, and be available on the Web.

Similarly, it would be desirable for research groups to make their analysis methods available either by publication of open code or under specific bilateral agreements. In the longer term, specific standardized software for analysis would be advantageous, but this should not restrict the continual evolution of measurement and analysis approaches.

Standard methods of diffusion assessment should be established and validated against phantoms appropriate to specific body locations, with their measurement reproducibility being established.

Recommendations for standardization

Basic standards for measurements/analysis and reporting of tissue diffusion coefficient should be established and adhered to. They should be tested against relevant phantoms, and reproducibility should be established.

New techniques need to demonstrate specific advantages over existing methods, providing comparison data that defines the benefit.

Studies should include routine measurement and QA analysis. Standardized data sets need to be made available to allow testing and comparison of analysis approaches.

Research groups should make analysis methods available, either as open source code or by specific agreements where there are confidential or commercial issues.

Standardization of software for analysis would be desirable.

Validation

To support the use of DW-MRI parameters in decision making about pharmaceuticals, it is important to link DW-MRI to underlying pathophysiological processes both before and after interventions.

Initially, this should be performed in well-defined model systems and then, where possible, confirmed by clinical measurements using biopsy specimens or surrogate tissues. The link between DW-MRI biomarker change and therapy response should also be established in xenografts and then clinically using both clinical outcome measures as well as pathologic surrogates of outcomes. Ideally, these biologic end points should relate specifically to the mechanism of action of the compound.

Suggested histologic validation of DW-MRI includes exploring links with measurements of proliferation index (Ki 67), cellularity index (cells/high-power field), tumor grade, and apoptosis. It will also be useful to explore/correlate DW-MRI with other MR measures of perfusion (dynamic contrast-enhanced (DCE) MRI, dynamic susceptibility contrast MRI, blood oxygenation level-dependent MRI), arterial spin labeling or metabolism magnetic resonance spectroscopy, and other imaging tests (e.g., FDG-PET, thymidine-PET, or annexin imaging for apoptosis).

Initially, clinical studies should validate practical approaches developed using the standardization guidelines described above, in more generalized applications such as chemotherapy response at varieties of anatomic sites. Neoadjuvant clinical trials are particularly suitable for these purposes because pathologic materials obtained can serve as rapid intermediate readouts/end points. If these are successful then novel therapeutics in early phase 1/2 studies can be evaluated.

Validation of DW-MRI in Relation to End Points

Requires correlation between size and type of biologic effect and relevant DW-MRI parameter, in animal models, supported by clinical biopsy or histology data.

Time course of effects will define the timing of imaging in clinical trials.

Attempt to derive hypothesis-driven relationships between imaging and specific biologic end points.

Biologic end points should relate to the purported mechanism of activity of the compound.

It would be desirable to be able to predict the magnitude of the MR effect based on animal models, allowing trial design to monitor dose-related change.

Reproducibility (See Also Appendix 3 for Detailed Methods)

To allow appropriate study design and to assess the significance of change, centers should demonstrate the reproducibility of their clinical measurements, in a manner that is traceable, providing information on individual and intergroup reproducibility. This information should be combined with evidence of the expected magnitude of therapeutic effect, such that studies can enable assessments of dose-related changes.

Reproducibility assessments are facilitated by incorporating baseline repeated measurements to provide information directly relevant to the body sites chosen. It is important to identify major sources of error leading to nonreproducible results. To determine whether changes in tumors induced by treatments are significant, three factors should be known. These are the natural biologic variability of parameters such as ADC, the variability inherent in the measuring instruments, and knowledge of additional errors induced by appraisers or analysis techniques. This implies that diffusion parameter measurement

changes cannot be taken at face value without due consideration of measurement errors. Estimates of measurement errors enable us to decide whether changes in ADCs are "real" for both group and individual observations.

Few published studies have documented measurement error in body DW-MRI, and the major contributors toward errors are not documented. However, from previous studies of other functional imaging techniques (e.g., DCE MRI), it is likely that DW-MRI measurement error will be dependent on a number of factors. These include imaging instrumentation and setup procedures, data acquisition techniques, and the time interval between repeated measurements. Data analysis techniques are also likely to add to measurement error including modeling techniques used (including range and noise of b value images used and implicit assumptions (monoexponential vs biexponential or multiexponential fitting)). Patient-related factors include tumor type, anatomic region being evaluated, and underlying physiologic status of patients.

It is important that clinical trials evaluating DW-MRI responses to treatment assess measurement variability as an intrinsic part of clinical trial design. The measurement error estimate component should be of sufficient statistical power (i.e., on enough patients) and needs to be performed on the study patients or in other patients who are representative of those being examined in the main study. To compare measurement errors of DW-MRI parameters at diverse anatomic sites and pathologies, it is important that similar statistical methods be used and that the meaning and limitations of statistical measures are understood.

Before statistical tests are applied, assumptions intrinsic to reproducibility analysis must be verified (e.g., normality of data and the nature of any relationship between measurement error and the magnitude of the parameters). Appropriate statistical parameters include the within-subject SD and coefficient of variance, and intraclass correlation coefficient should be quoted in communications (as detailed in Appendix 3). The repeatability statistic is a useful parameter for DCE-MRI studies because it informs on whether changes in a particular patient are significant.

Reproducibility

Centers should define reproducibility of data that is traceable, for individuals and intergroup comparisons, allowing the power of studies to be defined prospectively for a defined end point. Where possible, and in the absence of existing reproducibility data specific to the method, two baseline measurements should be incorporated to allow assessment of individual patient reproducibility.

Multiple lesions per organ should be taken into account. A standardized minimum statistical approach for reproducibility analysis should be reported.

Multicenter Methodology

In multicenter trials using identical (preferred) or similar methods (such as maintaining a constant field strength, imaging a single organ, etc.), comparison of precision and accuracy should be determined on phantoms to provide a basis for pooling of data, with account taken of corrections for machine-specific factors, and for sensitivity to motion effects not seen in phantoms.

Site qualification should be undertaken by the performance of measurements validated at a central analysis site before recruitment using

standardized data from each site. Readers should refer to Appendix 4 on QA procedures and diffusion phantoms for further details.

Analyses of DW-MRI data in multicenter trials should be performed at a single center using a standardized validated software. The reliability of analyses should be assured using data from each participating center before starting the trial.

In each study, patient and lesion selection, as well as the number of studies per patient including reproducibility assessments, should be defined prospectively. Reproducibility studies should be done at each imaging site because it provides estimates of measurement error in multicenter settings but also serves as a quantitative QA measurement of site performance. Standardized QA procedures should be enforced on all institutions participating to keep the data as uniform as possible.

Every effort should be made to ensure that the study can proceed on a given MR unit even if the unit is upgraded to a higher software level. It is important for the viability of DW-MRI, as a biomarker that implemented DW-MRI methods, be impervious to upgrades and software changes; otherwise, its future as a biomarker is in question.

Robust data acquisition protocols that are able to deal effectively with physiological motions should be instituted and adhered to.

Central data collections should incorporate appropriate QA and quality control procedures. Fast feedback to imaging sites is recommended to minimize data loss due to incomplete or incorrect imaging. The number and causes of failed examinations/analyses should be prospectively recorded. Ideally, failed examination/analysis rates should be <5% to 10%.

Data analysis should use a software that is fit for the purpose, is validated, and is preferably Food and Drug Administration 21 CFR part 11-compliant. 21 CFR Part 11 sets forth the requirements that need to be met to have the Food and Drug Administration consider electronic signatures and records equally trustworthy and just as reliable as handwritten signatures. Validation of software algorithms via multicenter trials is an essential need for obtaining regulatory approval to use DW-MRI as an accepted surrogate biomarker.

To promote the comparisons of ADC values obtained from different centers and for differing therapies and to overcome the dependence of ADC on the range of b values chosen for any particular study, perfusion-insensitive ADC values (by excluding the $b = 0 \text{ sec/mm}^2$ image from the ADC calculation) should always be quoted.

Additionally, study data should be publicly available to enable alternative analytic approaches that might be superior to the ones used in the study.

Multicenter Studies

Animal validation should be undertaken before human studies to provide information on the appropriateness of using DW-MRI and may be able to indicate the optimal timing for doing imaging in human studies.

Double-baseline studies should be done to provide data about measurement error of imaging specific to the study and thus knowledge of what constitutes a significant change in an individual and in a group of patients (powering studies).

Quantified parameters such as ADC should be measured to derive physiologically meaning that can be related to drug mechanisms of action. Quantified parameters have the advantage of allowing interpatient and inpatient comparisons to be made. Good quality control and QA are keys to success for multicenter studies.

Software

Software applications that provide the means to standardize analysis and display of DW-MRI data will facilitate the advancement of this approach for diagnostic and therapeutic assessments in cancer imaging.

Software applications should allow efficient and reproducible analyses of DW-MRI data using tools that allow zoom and pan functionality for two-dimensional and three-dimensional displays. Efficient ROI/VOI delineations using anatomic and b -value images should be available.

It is preferred that software incorporates intraexamination and interexamination image registration. Image intensity-based spatial registration using mutual information is vital because it will allow for voxel-wise changes to be followed over time in individuals.

The software should have a flexible workflow and should be able to generate a variety of quantitative calculations for comparing tumor diffusion values at multiple time points and to visualize therapy response in individual, configurable layouts (Figure 6).

Overall, the standardized software will provide users with the ability to generate automatically a variety of quantitative calculations based on volumetric analysis (histograms) or voxel-based analysis over multiple time points and to visualize therapy response (an example of voxel-based display is functional diffusion maps — fDM — see Appendix 1).

Open source code should be used whenever possible to permit broad dissemination of the method, thus encouraging standardization. The application should allow to import/export DICOM

data sets and to output measurements via schema-based XML to data repositories.

Software

Diffusion-weighted MRI data require unique analysis approaches for quantification of results.

Standardization of software applications for analysis of DW-MRI data should provide for volumetric/histogram and for voxel-wise quantification approaches.

Standardized software will provide for robust, standardized approaches for analysis of single and multisite clinical trials data that can be used for regulatory submission.

Recommendations for Future Methodology Development

Measurements

Development and testing of new acquisition protocols sensitized to multiple directions with the objective of improving image quality and reducing distortions and artifacts.

Improved robustness in whole-body DW-MRI techniques with increased coverage and improved fat suppression.

Improved robustness of acquisition techniques able to deal with physiological motions as dictated by specific anatomic locations are

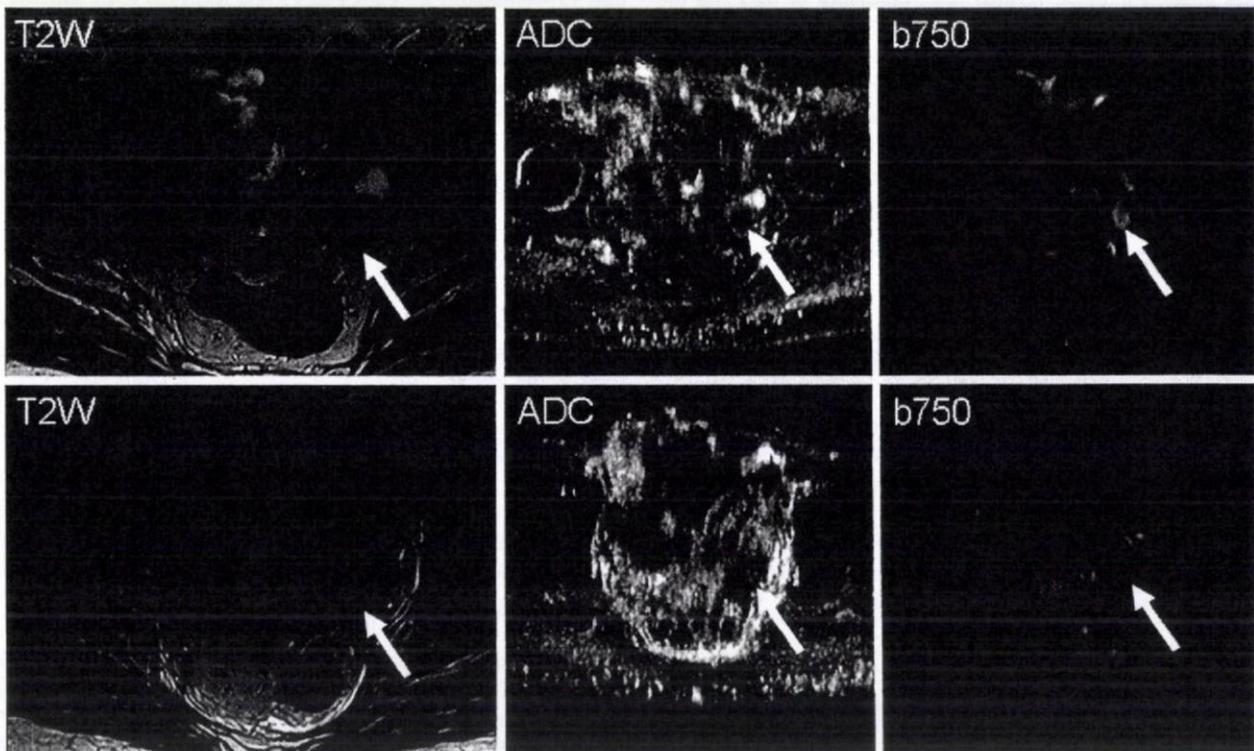


Figure 6. Diffusion-weighted MRI used in treatment response monitoring. Top row demonstrates a perirectal mass (arrow) on T2-weighted MRI, ADC, and the b_{750} gradient image in this patient with rectal cancer metastatic to a local lymph node. On chemotherapy, the lesion has decreased in size but also has higher ADC values and corresponding lower signal on the b_{750} image. The changes in ADC may be useful for monitoring response to therapy.

needed to minimize smearing of data image, to reduce partial volume averaging effects, and to preserve heterogeneity of ADC maps. The relative benefits of motion-compensated and -independent measurement techniques need systematic evaluation.

There is debate on the best phantom materials, options are described in Appendix 4. Whatever material is ultimately used, there is a pressing need for equipment manufacturers to specify clearly QA procedures that they already do for SNR measurements.

Improved phantoms mimicking the cellular environment of living tissue are required. Simple phantoms filled with liquids with different diffusion coefficients measure diffusion coefficients and ADC. Phantoms filled with beads of well-controlled size may help us in understanding the properties of diffusion of extracellular water (provided that susceptibility effects are negligible) but does not tell us anything about other water diffusion components (vascular contributions and intracellular). However, non-Gaussian distributions of the displacement would be seen so these could be used initially.

It is recognized that improvements in SNR afforded by 3-T systems could be a boost for DW-MRI. It is accepted that diffusivity is likely to be independent of field strength; however, the effect of high field strengths on data acquisitions and analyses still needs to be systematically evaluated.

Analysis Methods

Improvements in software for analyzing DW-MRI data are needed. Such software should be adapted for "clinical" and "research" uses. Ideally, the ability to generate synthetic b -value images back-calculated from ADC/ b values (less noise) would be valuable and could be used to extrapolate very high b value images. The method of calculating ADC should be clear to the reader (via fitting displays, noise of individual b -value images, and calculation error). There should be flexible scaling of ADC images on displays and the ability to generate ADC values for any given range of b values to promote the calculation and reporting of perfusion-insensitive ADC values in clinical studies. Low SNR pixel values should be eliminated before ADC map calculation, and these pixels should be flagged as "not-a-number" for exclusion from subsequent analysis. The ability to segment (threshold) high- b value/ADC images and then to obtain ADC values from "threshold ROIs" using histograms are needed.

Fusion imaging techniques are invaluable for DW-MRI, and improved registration in three-dimensional using mutual information and warping techniques is needed.

Fusion methods need to be integrated into analysis and reporting/communication tools. Fusion techniques need to be improved and extended with the ability to display, coregister, segment, fuse, and analyze multiple functional imaging methods (DCE; blood oxygenation level-dependent, magnetic resonance spectroscopic imaging, etc.) in an integrated single work space. Such tools should allow for missing functional data sets (not obtained, poor SNR, corruption, and artifacts).

It is becoming clear that DTI techniques provide information that is valuable for investigating some organs such as the kidneys, muscle, and the prostate gland. These efforts should be encouraged, and to this end, robust data acquisition sequences and analysis software should become more available.

Conclusions

Diffusion-weighted MRI is an attractive noninvasive, quantitative technique that yields parameters that relate to tissue structure, cellularity, and necrosis. The plethora of existing methods in the current

literature makes it impossible to conclude definitively that DW-MRI is qualified as a pharmacodynamic marker of therapy response. Therefore, the major outcome of this consensus document is to stimulate funding agencies, vendors, and clinical trials to collaborate on the design of multicenter studies that will test the value of DW-MRI in more controlled circumstances than have hitherto been possible. This will require focused groups of researchers to agree on precise clinical scenarios in a variety of cancer types and on the timing of imaging so as to address key biomarker development questions. Agreements will be needed on data acquisition and analysis methods that yield water diffusion biomarkers, on calibration tools, and statistical methods for analyzing trial data. Although this task is complex, the potential benefits are enormous given the ease with which DW-MRI data can be incorporated into clinical trials. We hope that this White Paper will catalyze action in this field so that these challenges can be addressed in the near future.

Appendix 1. Nomenclature and Definitions

Anisotropy: A state in which a physical characteristic varies in value along axes in different direction; a physical measurement made in one direction differs from the measurement made in another direction.

Apparent diffusion coefficient (ADC) or diffusion map: A parametric image displaying coefficients derived from DW images. The term apparent refers to the dependence of these coefficients on factors that alter the apparent mobility of free water.

b value: An attenuation factor related to the strength of applied gradients on DW images; the higher the value b value, the stronger the diffusion weighting (measured in seconds per square millimeter). Increasing b value can be obtained by 1) increasing the strength (height) of the diffusion gradient, 2) increasing its duration, and 3) increasing the gap between the paired gradient lobes.

Diffusion-weighted MRI (DW-MRI): An MRI technique where signal and contrast is determined by the microscopic mobility of water. This mobility is due to thermal agitation and is influenced by the cellular environment of the water. The term DW-MRI is preferred to DWI, which commonly refers to "driving while intoxicated."

Diffusion tensor imaging (DTI)/trace images: A sophisticated form of diffusion imaging that allows the directionality and the magnitude of water diffusion to be determined.

Diffusion coefficient: A constant of proportionality relating particle flux to the concentration gradient. The SI unit of is meters squared per second. This is too large with respect to typical values observed by MRI in tissues, and the unit millimeter squared per second is recommended.

Functional diffusion map (fDM): a color overlay image of therapeutic-induced ADC change within the tumor. The fDM provides the ability to objectively segment tumor into color regions based on the magnitude and direction of ADC change.

Fractional anisotropy (FA): A DTI parameter that grades the probability of water diffusion in the environment having a spherical probability/being normalized. An FA value of 0 indicates a perfectly spherical distribution of probability, whereas an FA value of 1 indicates a completely definite direction of diffusion.

Intravoxel incoherent motion (IVIM): Flow-related dephasing of magnetized water molecules within voxels resulting in signal losses on low- b value images ($<100 \text{ sec/mm}^2$).

Isotropy: The quality of a property that does not depend on the direction along which it is measured (has the same intensity regardless of the direction of measurement).