의학박사 학위논문

확산강조영상의 정량적 영상 바이오마커 정립을 위한 팬텀 기반 품질 평가 및 프로토콜 표준화 연구

Standardized Process of Diffusion-Weighted Imaging for Imaging Biomarkers: Quality Assessment and Protocol Validation Using Phantoms

울산대학교 대학원

의 학 과

최 세 진

확산강조영상의 정량적 영상 바이오마커 정립을 위한 팬텀 기반 품질 평가 및 프로토콜 표준화 연구

지도교수 김 경 원

이 논문을 의학박사 학위 논문으로 제출함

2024 년 08 월

울산대학교대학원

의 학 과

최 세 진

최세진의 의학박사학위 논문을 인준함

심사위원 서 종 현 인

심사위원 김 경 원 인

심사위원 김 정 곤 인

심사위원 우 동 철 인

심사위원 이 정 진 인

울 산 대 학 교 대 학 원

2024 년 08 월

국문요약

배경: 현성 확산 계수(ADC)를 신뢰할 수 있는 바이오마커로 사용하기 위해서는 MRI 장비 성능과 임상 촬영 프로토콜을 검증하는 것이 필수이다. 이 연구에서는 상용 팬텀을 사용하여 확산 강조 영상(DWI)에 대해 다양한 MRI 장비와 임상 뇌 프로토콜을 검증하고, 검증된 프로토콜을 환자에게 적용하여 영상 질을 확인하고자 한다.

방법: 정량화 영상 바이오마커 연합(QIBA)의 확산 팬텀과 클라우드 기반 분석 도구를 사용하여 네 가지 다른 MRI 스캐너와 임상 뇌 프로토콜의 성능을 검증하였다. ADC 측정의 정확성과 반복성은 QIBA 프로파일의 성능 지표를 이용하여 평가하였다. 검증된 임상 뇌 프로토콜은 17 명의 환자에게 적용되었으며, 환자영상에서 이미지 품질과 ADC 반복성을 평가하였다.

결과: 네 가지 MRI 스캐너 모두 ADC 측정에서 높은 정확도(ADC 편향, -2.3% – -0.4%), 기준 ADC 값과의 우수한 선형 상관관계(기울기, 0.9 – 1.0; R^2, 0.999 –1.000), 높은 단기 반복성(변동 계수[wCV], 0% – 0.3%)를 보였다. 임상 프로토콜 또한 높은 정확도(ADC 편향, -3.1% – -0.7%)와 견고한 반복성(wCV, 0% – 0.1%)으로 QIBA 요구 사항을 충족하였다. 검증된 임상 프로토콜을 사용하여 얻은 뇌 DWI 는 우수한 이미지 품질(평균 점수 ≥ 2.9)과 좋은 반복성(wCV, 1.8–2.2)을 보였다.

결론: DWI 의 종합적인 표준화 과정은 QIBA 요구 사항에 부합하였으며 ADC 측정이 다양한 MRI 장비와 임상 프로토콜에서 매우 정확하고 반복 가능함을 입증하였다.

차 례

표 목차 및 그림 목차

Introduction

Diffusion-weighted imaging (DWI) is a functional magnetic resonance imaging (MRI) sequence that measures the diffusivity of water molecules, influenced by factors such as tissue cellularity, microstructure, fluid viscosity, cell membrane permeability, and blood flow (1-4). The apparent diffusion coefficient (ADC) value is a functional parameter calculated as the mean diffusivity along three orthogonal directions from DWI. In current clinical practice, DWI is mainly used for the detection and characterization of lesions. However, recent advancements in therapeutics have heightened the need to utilize the ADC as a quantitative imaging biomarker for objectively assessing treatment responses, given that the ADC value reflects the cellularity of a lesion. (2, 5-7). For example, chemotherapy-related changes in tumors initially manifest as alterations in tissue cellularity, which are then followed by changes in lesion size (1, 6, 8). Assessing treatment response in tumors requires longitudinal monitoring with quantitative biomarkers at multiple time points (7, 9). Therefore, imaging biomarkers for treatment response assessment should be validated for both accuracy (i.e., how close the ADC values are to the true values) and precision (i.e., how consistent the ADC values are between repeated measurements).

DWI is relatively susceptible to specific features of MRI scanners, such as signal-to-noise ratio (SNR) and the calibration of high-strength diffusion gradient systems, which affect the cross-term factors, as well as sequence parameters (10-13). The ADC reflects the diffusivity of a lesion and is calculated by applying the signal intensity measured in DWI to an exponential equation. Consequently, when the signal intensity in DWI varies due to numerous technical factors, ensuring the reliability and repeatability of ADC measurements becomes challenging. Therefore, it is important to maintain constant MRI scanners and sequence parameters when using DWI and ADC as biomarkers (14, 15). In response to these needs, the Quantitative Imaging Biomarker Alliance (QIBA) issued the QIBA DWI profile, which provides standardized imaging protocols for DWI acquisition (16). The DWI acquisition protocols for phantoms suggested by QIBA (hereafter referred to as the QIBA phantom protocol) are designed to minimize differences in ADC values regardless of the type of MRI scanner and software version. Validation of MRI equipment performance and clinical protocols is performed using a phantom. The QIBA developed the QIBA diffusion phantom to validate the accuracy and repeatability of DWI acquisition and ADC measurement (9, 17, 18). Along with the development of the phantom, a cloud-based tool was also created for standardized phantom image analysis. After fulfilling the MRI scanner requirements, clinical brain protocols scanned with validated MRI equipment can be validated for patient scanning through an image quality assurance (QA) process. The final step of QA is to review patient DWI for image quality and ADC measurement repeatability.

The QIBA phantom protocol primarily focuses on validating MRI scanners rather than its

applicability in clinical practice. In contrast, clinical acquisition protocols for brain DWI used in routine clinical practice at each institution (hereafter referred to as clinical brain protocols) differ from the QIBA phantom protocol. These clinical brain protocols prioritize practicality in real-world settings, and further variability is introduced across institutions due to diverse MRI vendors and software. Therefore, the accuracy and precision of ADC measurements should be validated using these clinical brain protocols to effectively use ADC as an imaging biomarker.

To date, there is limited evidence of the use of the QIBA diffusion phantom and the standardized phantom image analysis tool in real-world practice and clinical trial settings (16, 18, 19). A few previous studies have focused on the repeatability and reproducibility of ADC values and the validation of MRI equipment using the QIBA phantom. However, to utilize ADC values as an imaging biomarker, it is essential to standardize the validation process of DWI and confirm the feasibility of clinical imaging protocols.

Therefore, we conducted this study using the QIBA diffusion phantom to validate four MRI scanners from different vendors with the following goals: (1) to validate MRI equipment performance using the QIBA phantom protocol with a phantom, (2) to validate clinical brain protocols in terms of accuracy and repeatability with a phantom, and (3) to apply the validated clinical brain protocols for patient scanning and image quality assurance (QA).

Materials and Methods

Overall Study Scheme

This study was performed by three steps (Figure 1). First, we evaluated the performance of various MRI scanners according to the QIBA claims of the QIBA profile using the QIBA phantom. Second, we validated the clinical brain protocols for DWI on various MRI scanners according to the QIBA claims of the QIBA profile using the QIBA phantom. Finally, we applied the validated clinical brain protocols to patient scanning in a clinical trial and performed image quality assurance (QA) for the acquired DWI scans.

Phantom Preparation

In this study, we used a commercially available phantom (CaliberMRI, Inc., Boulder, CO, USA) developed by the National Institute of Standards and Technology (NIST) with support and input from the National Cancer Institute and QIBA (hereafter referred to as the QIBA diffusion phantom) (16, 19).

The phantom is a spherical device measuring 194 mm in diameter, housing 13 vials (30 mL each) filled with varying concentrations of polyvinylpyrrolidone (PVP) ranging from 0% to 50% w/w (Figure 2). The central vial contained pure water (0%), and the vials with PVP solutions were arranged in two arrays: an inner ring and an outer ring. The PVP solution was used to generate physiologically relevant ADC values, with higher PVP concentrations resulting in lower ADC values. Given the temperature sensitivity of diffusion properties, the ADC value of each PVP concentration at 0℃, as verified by the National Institute of Standards and Technology, was used as the reference standard. To ensure homogeneity and stability of the phantom temperature during the test, the space between the cylinders was filled with crushed ice and tap water prior to scanning, according to the instructions provided by the Quantitative Imaging Biomarker Alliance. The phantom was maintained at approximately 0℃ during scanning.

DWI Acquisition Using the Phantom

Table 1 summarizes the detailed information of image acquisition parameters for the QIBA phantom protocol. When scanning the phantom using the QIBA phantom protocol, all acquisition parameters adhered to the protocol except for the number of excitations (NEX), which was set to 1 instead of 2 for the 3.0-T scanner. The QIBA phantom protocol specified the echo time (TE) as the shortest possible, and in this study, all scans were obtained with a TE of less than 60 ms.

Table 2 provides information on the MRI systems and protocols for brain DWI as suggested by QIBA and various MRI vendors. A total of three 3-T platforms (SIGNA Architect from GE Healthcare; Ingenia from Philips Healthcare; Magnetom Vida from Siemens Medical Solutions) and one 1.5-T platform (Magnetom Avanto from Siemens Medical Solutions) were used for DWI.

DWI phantom images were acquired using two acquisition protocols: (1) the QIBA phantom protocol for MRI equipment performance evaluation (Table 1) and (2) the clinical brain protocols for evaluation of DWI acquisition and ADC measurement for routine practice (Table 2). According to the QIBA profile, short-term (intra-exam) and long-term (multiday) repeatability were evaluated using the QIBA phantom protocol. We conducted four examinations for each scanner using the QIBA phantom on the same day and repeated the exams in the same manner one month later for each protocol. The specifications of the clinical brain protocols are consistent with the QIBA profile claims and recommendations for brain DWI. The clinical brain protocol met the ideal or target specification of the

QIBA brain profile, with the following exceptions that were within acceptable limits: gap thickness (2 mm), acquired matrix (128 x 128), and number of averages (1).

Quantitative Analysis of Phantom Data

Phantom Data Processing

For standardized analysis of quantitative DWI phantom data, we used commercially available cloud-based quality assessment software (CaliberMRI, Inc., qCal-MR Quality Control (QC) Software, www.qmri.com). The data and figures were adopted from the quality assessment report with permission from CaliberMRI. Quantitative analysis of DWI was performed by measuring the ADC values from volumes-of-interest (VOIs) derived from circular regions of interest (ROIs) measuring 19.6 mm in diameter on five slices. ADC maps were created from multiple DWI b-value pairs, and for the clinical DWI protocol, b-values of 1000 and 0 s/mm² were used. ADC maps were created using a mono-exponential model. The reference standard for the ADC value of each PVP concentration was the NIST-verified value provided by the QIBA profile. The analytical method used in the commercial software was identical to the standard analysis software provided by the Quantitative Imaging Data Warehouse (QIBA QIDW, rsna.org/qidw). QIBAphan, an open-source DWI phantom QC analysis software provided by QIBA QIDW, can be accessed online at https://bit.ly/2QXLo3e. QIBAphan converts QIBA DWI data from classic DICOM format into uniform data structures for generating QC statistics. Users select ROI centers in each slice of the DWIs, and the software automatically generates statistics on the ADC values across the VOIs. QIBAphan provides a QC report that includes processed output ROI statistics and performance metrics in CSV files.

Phantom Data Analysis

Phantom data analysis was performed for both the QIBA phantom protocol for MRI equipment performance evaluation and the DWI clinical brain protocols for routine clinical practice. All phantom data analyses followed the QIBA profile using the QIBA claims. For phantom data acquired using the QIBA phantom protocol, protocol compliance was checked to ensure that the acquired MRI met the recommended acquisition parameters of the QIBA phantom protocol. A radiologist (S.J.C) performed a visual inspection to assess the image quality of all phantom DWI data on-site. This inspection included verifying the presence of all required DWI series, the number of b-values as suggested by QIBA (Table 2), and checking for the presence of artifacts that could interfere with the evaluation of performance metrics by the quality assessment software.

The definition and equation of the quantitative DWI performance metrics are presented in Table 3 (9, 16). The quantitative DWI performance metrics included the bias in ADC measurement (ADC bias), the measurement repeatability estimated by the repeatability coefficient (RC) and the within-subject coefficient of variation (wCV), linearity, b-value dependence, random measurement error, and signal-to-noise ratio (SNR).

ADC bias is an estimate of measurement error, calculated as the difference between the average measurement and its true value. For efficient data demonstration, this bias is divided by the true value and presented as a percentage (%bias). Repeatability represents the precision of measurements taken in repeated exams under the same or similar examination conditions over a short period of time, using the same measurement procedure (9, 16). In the phantom study, short-term repeatability within the examination was assessed by calculating the repeatability coefficient (RC) and the within-subject coefficient of variation (wCV). Linearity refers to the ability to provide measured quantity values that are directly proportional to the value of the reference standard. Precision reflects the closeness of agreement between the measured values obtained by replicate measurements on the same or similar experimental units under specified conditions. In this study, random error was assessed as an indicator of precision. Evaluation of the signal-to-noise ratio (SNR) provides a relative system performance metric and confirms that the MRI equipment is adequate to measure ADC bias without introducing incremental bias due to a low SNR (16). SNR was calculated as spatial mean of signal image divided by mean of noise image.

Application of Validated Clinical Protocols

This prospective study was approved by the institutional review board of our hospital (2017-0467), and informed consent was obtained from all eligible patients. Patients who visited the institution for the evaluation of cerebral infarction and agreed to participate in the clinical study were included. After the clinical study trial period, we retrospectively collected brain DWI scans from these included patients. The scans were conducted using the clinical brain protocol for the initial evaluation and follow-up of cerebral/cerebellar infarction between March 2013 and September 2020. The mean age of the human subjects was 66.1 years (range, 38–87 years), and male was 58.8% (10/17). A total of 55 DWI scans were performed on these 17 patients, comprising 20 scans from 3.0-T machines and 35 scans from a 1.5-T machine. Details about the brain DWI scan protocol are presented in Table 2. If there was any image degradation caused by patient motion, susceptibility, or eddy current distortion, immediate corrections were made on-site, and the exams were retaken (21).

We evaluated the image quality and then calculated the RC and wCV of each image. DWI data quality was evaluated by an experience radiologist (K.W.K.) for the items presented by the QIBA using a 3-point scale (1, Unacceptable; 2, Acceptable; 3, Ideal) (14). Details of items for image quality assurance and scale are as follows:

Low SNR: Visualization of anatomical features in tissues of interest at all b-values was evaluated: unacceptable, poor SNR at all b-values with anatomical features are lost; acceptable, minor deterioration of image without disturbing visualization of anatomical structure; ideal, identification of all anatomical structures with accurate structure.

Ghost/parallel imaging artifacts: The presence of the discrete ghosts from extraneous signal sources along the phase-encode direction obscuring the tissue of interest was evaluated: unacceptable, presence of artifact creating erroneous ADC value; acceptable, minor artifact without disturbing assessment of performance parameter; ideal, visualization of anatomical features in tissues of interest and no artifact.

Severe spatial distortion: Severe spatial distortions affecting ADC values and the apparent size/shape/volume of tissues of interest were investigated: unacceptable, severe distortion altering apparent size/shape/volume of tissue of interest; acceptable, minor distortion without significant modify of shape of tissue; ideal, no distortion of tissue of interest.

Eddy currents: Blur or spatial misalignment between low and high b-value diffusion-weighted imaging (DWI), particularly at the edges of anatomical features, was evaluated: unacceptable, blur of anatomy causes and erroneous measurement of ADC value; acceptable, minor spatial misregistration only affecting the edge of the lesion; ideal, no evidence of blur or spatial misalignment in all images.

Fat suppression: Superposition of unsuppressed fat signal on the tissue of interest was assessed: unacceptable, unsuppressed fat signal spatially shifted obscuring the tissue of interest and renders ADC meaningless in tissue superimposed by a residual fat signal; acceptable, minor detrimental chemical shift artifacts not affecting tissue of interest; ideal, complete suppressed fat signal onto tissue of interest.

Motion artefacts: The presence of cerebrospinal fluid (CSF) pulsation in the ventricles or cardiac pulsation near large vessels and the brainstem was assessed: unacceptable, motion artifact contributes to blurring image and erroneous signal leading to unpredictable ADC values; acceptable, minor artifact without disturbing assessment of tissue of interest; ideal, no motion artifact in the image.

Nyquist ghost: Duplication of an anatomical structure or distortion of an image occurring in the phase encoding direction was evaluated. Unacceptable, artifact presence of artifact disturbing recognition of anatomical structure and obscure tissue of interest leading to unpredictable ADC values; Acceptable,

minor artifact without disturbing assessment of performance parameter; Ideal, Visualization of anatomical features in tissues of interest and no artifact.

After evaluating the quality of the DWI data, ADC values are extracted using AsanJ-Stroke Software (Asan Image Metrics, Seoul, Korea; assessed at https://datasharing.aim-aicro.com/strokevolumetry) (22), which was developed based on ImageJ (NIH, Bethesda, Maryland, USA). In each subject, a radiologist drew two ROIs in the non-diseased healthy brain hemisphere—a circle with a 20mm diameter at the center of the frontoparietal white matter and another with a 10mm diameter in the cerebrospinal fluid (CSF) of the lateral ventricle—for ADC measurements. The ROIs were manually drawn on the b=1000 s/mm² image considering lesion location (Figure 3). The ROIs were saved in separate files and then subsequently applied to the equivalent site of the ADC map. AsanJ provides the ADC pixel histogram as well as the ADC mean and standard deviation using data form ADC map. Using extracted data from the ADC, performance metrics according to the QIBA profile were evaluated. To evaluate the long-term repeatability of each scanner, the RC and wCV were calculated as the QIBA profile. And we compared the measured ADC value with reference ADC value of white matter and CSF (23, 24).

Results

Validation of MRI Equipment Performance

Protocol Compliance and Image Quality

Phantom DWI data from the four MRI systems were successfully acquired, satisfying the QIBA profile without artifacts. The quality assessment software evaluated protocol compliance according to the QIBA profile before calculating performance parameters. All phantom DWI data acquired from the four MRI scanners using the QIBA phantom protocol passed the protocol compliance. No parameters exceeded the limits of the QIBA claims. For qualitative inspection, we identified four DWI series and confirmed that each was composed of five b-values. DWI data from the four vendors had appropriate image quality without significant artifacts. Representative images of DWI and ADC maps are shown in Figure 4.

Key Quantitative DWI Performance Metrics

The ADC VOI statistics reports and derived graphs of DWI data generated by the quality assessment company contain ADC statistics and representative graphs of ADC values based on axial position, varying concentrations of the solution, and repeatability parameters of ADC. The performance metrics

results of all scanners are summarized in Table 4. All four MRI scanners met the conformance criteria of the QIBA DWI claims.

The ADC bias of the four MRI scanners ranged from -2.3% to -0.4%, which were within the QIBA DWI claims (< absolute value of 3.6). The short-term repeatability of all four MRI scanners with wCV ranging from 0% to 0.3% also met the QIBA claims (wCV ≤ 0.5%). There was an excellent linear correlation between the measured ADC values and the true values (i.e., NIST values), with slope ranging from 0.98–1.0 and R2 ranging from 0.999–1.000 in all MRI scanners, which met the QIBA claim (R2 > 0.9 and slope 0.9–1.0). The max b-value dependence, ranging from 0.1% to 1.5%, was within the QIBA claim limit (≤ 2%). The random measurement errors ranging from 0.5% to 1.7% also met the QIBA claim (≤ 2%). The SNR of all four MRI scanners were high (56.0–230.1), which met the QIBA claim (≥ 45).

Validation of Clinical Brain Protocols

Protocol Compliance and Image Quality

For evaluation of the clinical brain protocols, we have verified the protocol compliance manually, because the QC software was optimized for the QIBA protocol. For qualitative inspection, we identified four DWI series and confirmed that each series was composed of two b-values, b=0 s/mm² and b=1000 s/mm². DWI data from the four vendors had appropriate image quality without significant artifacts.

Key Quantitative DWI Performance Metrics

As presented in Table 5, the results of the performance metrics of all clinical brain protocols are as follows: ADC bias values (ADC bias, -3.1% to -0.7%), short-term repeatability (wCV, 0%–0.1%), linearity (R2, 0.995–0.998; slope, 0.95–1.00), random measurement error (0.4%–0.8%), and SNR (145.6–380.3). These performance metrics met the QIBA claims. However, the max b-value dependence could not be calculated for the clinical brain protocols because there were only two b-values.

Quality Assurance for Acquired Patient DWI

The results of image quality assessment, measured ADC values, and repeatability parameters are presented in Table 6. DWI data from MRI scanners fulfilled the ideal quality for all items except for one scan from a 1.5-T machine having mild spatial distortion and a Nyquist ghost artifact (Figure 5, scored acceptable for each item), and one scan from a 3.0-T machine had a Nyquist ghost artifact

(acceptable). None of scans had unacceptable item. For the brain parenchyma, the mean ADC value from three MRI scanners ranged 808.6 μm2/s to 843.7 μm2/s. The repeatability of ADC value measurement for the brain parenchyma showed a wCV ranging from 2.0% and 2.1%, which met the QIBA claim. The CSF, which has relatively low diffusion restriction, also showed good repeatability (2.0% and 2.1%) within the QIBA claims. The long-term reproducibility of ADC value measurement for the brain parenchyma showed good performance with wCV ranging from 1.8% to 1.9%, which met the QIBA claim. The mean ADC value of CSF from four scanners ranged 3011.0 μm2/s to 3156.5 μm2/s with a wCV ranging from 2.0% to 2.1%. The experimental ADC value of water at 37°C is 3037.7 μm²/s, and the difference between the experimental ADC of water and the measured CSF at 37°C ranged from 3.6 to 118.8 μm²/s (24).

Discussion

In this study, the MRI equipment performance of four MRI scanners and the clinical brain DWI protocol were evaluated for DWI acquisition and ADC measurement using the QIBA diffusion phantom and a standardized cloud-based phantom analysis tool. All performance metrics for the four MRI scanners were within the range of the QIBA profile claim without requiring additional calibration procedures. These results validated the accuracy and repeatability of DWI acquisition and ADC measurement with these four MRI scanners. The clinical brain protocols also demonstrated excellent accuracy and repeatability in ADC measurement during the phantom imaging analysis. This indicates that our clinical brain protocols can be reliably applied to patient scanning in clinical practice and trials. Indeed, the DWI from 17 patients acquired for a clinical trial met the ideal quality for most QC check items and showed excellent long-term repeatability according to the QIBA claims.

The recently developed standardized QIBA diffusion phantom was commercialized along with the cloud-based standardized phantom analysis tool, making it accessible for purchase by any institution or researcher. This study might be the first published report on the experience of standardizing DWI using both the QIBA diffusion phantom and the cloud-based analysis tool. Previous studies validated DWI as an imaging biomarker at individual institutions using separately developed phantoms and the use of different validation tools posed a challenge in applying a consistent validation method across various institutions. The standardized QIBA phantom facilitates a common validation method across multiple institutions, contributing to the establishment of DWI as a reliable imaging biomarker.

In our study, we validate MRI equipment using QIBA phantom protocol and clinical brain protocols. The notable differences between the QIBA phantom protocol and clinical brain protocols

were b-values (0, 500, 1000, 1500, and 2000 s/mm2 for the QIBA phantom protocol vs. 0 and 1000 s/mm2 for the clinical brain protocol), repetition time (TR) (8000 ms vs. 3000 ms), and ADC map creation method (log-linear model vs. mono-exponential model). In general, the clinical acquisition protocols of brain DWI in many institutions are optimized for a clinical setting with a shorter acquisition time with acceptable image quality. Clinical protocols adopt a shorter TR and TE with fewer b-values and NEX than the QIBA phantom protocol. Thus, the QIBA phantom protocol showed slightly higher linearity across the vendors than the clinical protocols. DWI acquisition with higher NEX and more b-values can increase the SNR of DWI and accuracy of ADC measurement. Nevertheless, our study showed that the clinical brain protocols also showed comparably good linearity in ADC values and high reproducibility, which met the QIBA claims.

The validated clinical brain protocols were applied to patients in a clinical trial and showed excellent image quality and high long-term repeatability. To ensure reliable patient DWI scanning, periodic QA procedures are very important (25). Image QA refers to the comprehensive quality management process including DWI acquisition and ADC measurement. In that sense, image QA includes validation of the MRI equipment performance and clinical acquisition protocols using a phantom as well as an image quality evaluation on actual patient DWI scans (16). According to the QIBA profile, periodic QA should be performed, especially in clinical trials using DWI/ADC measurement as quantitative biomarkers. From the preparation of the phantom and MRI equipment to obtaining the images and analysis report, it took approximately 30 minutes. Considering this, Quality Control (QC) using the QIBA phantom and profile ensures machine performance and is advantageous in terms of efficiency. In our study, we presented the whole process of image QA from phantom imaging to patient imaging.

Our study has several limitations. First, the number of MRI scanners and clinical brain protocols was relatively small. Several previous studies have demonstrated the accuracy of ADC measurements across various scanners in larger numbers. However, to the best of our knowledge, this study is the first to show the complete process of standardizing DWI using a commercially available phantom. For the utilization of DWI as a quantitative biomarker, it is crucial to establish a standardized validation process for MRI scanners using the same phantom. Therefore, further studies with a larger number of MRI scanners and protocols across multiple institutions are necessary. Second, we applied ADC measurement to brain MRI in a clinical trial with a small number of patients. Different intrinsic biophysical tissue properties of various organs may affect the ADC values and their repeatability. Thus, further studies are necessary involving a larger number of patients and different organs to validate DWI as a quantitative imaging biomarker.

In conclusion, we demonstrated a whole standardization process for DWI from validation of MRI equipment performance and clinical brain protocols to application for patient scanning. This study demonstrated the robustness of DWI with high accuracy and repeatability across diverse MRI equipment and clinically optimized protocols, which is in accordance with the QIBA claims.

Table 1. Image acquisition parameters for QIBA phantom protocol.

Parameters	QIBA phantom protocol
Field strength (T)	1.5 or 3.0 T
Receiver Coil	Head coil
Sequence	DWI EPI
Slice orientation	Axial
FOV	220 X 220 mm
Acquired Voxel size	1.72 x 1.72 x 4 mm
Acquired Matrix (frequency x phase)	128 x 128
Recon voxel size	0.86 x 0.86 x 4 mm
Recon Matrix	128 x 128 to 256 x 256
Parallel imaging acceleration	factor = 2
Phase encode direction	Anterior- posterior
Frequency encode direction	Right - Left
Oversampling	Off
Number of slices	25
Packages	1
Slice thickness	4 mm
Slice gap	1 mm
B0 Shim	Best quality volume shim
B1 Shim	Off or default
Scan mode	Multislice
Technique	Spin echo
Fast imaging mode	Echo planar imaging
Shot mode	Single shot
Echoes	1
Partial echo	Off
TE	shortest
Flip angle	90 deg
TR	8000 ms
Half scan factor	≥0.75
Water fat shift	Minimum
Fat suppression	STIR
Diffusion encoding directions	Three orthogonal
b-value	0, 500, 1000, 1500, 2000
Average high b-value	Off
Gradient mode	Maximum
NEX	2
Preparation phases	Full prep
Geometry phases	Default
Bandwidth in frequency-direction	Maximum

Abbreviations: DWI, Diffusion-weighted imaging; EPI, Echo planar imaging; FOV, Field of view; NEX, number of excitations; QIBA, Quantitative imaging biomarker alliance; STIR, Short TI inversion recovery; TE, Echo time; TR, Repetition time.

Table 2. Specifications for brain DWI scan protocols by QIBA and institution.

Parameters	QIBA clinical brain protocol requirement	Institution clinical brain protocol for			
		GE SIGNA Architect	Philips Ingenia CX	Siemens MAGNET OM Vida	Siemens Avanto
Field strength (T)	1.5 or 3.0 T	3.0 T	3.0 T	3.0 T	1.5 T
Acquisition sequence	SS-EPI	SS-EPI	SS-EPI	SS-EPI	SS-EPI
Receiver Coil	Ideal[ǀ]: 32 channel head array coil Target[ǁ]: 8-32 channel head array coil Acceptable[ǂ]: 8 channel head array coil	16- or 32-channel head coil	16- or 32-channel head coil	16- or 32-channel head coil	16- or 32-channel head coil
Fat suppression	On	STIR	STIR	STIR	STIR
Number of b-values	Ideal: >3 (including one b=0-50; one 450-550 s/mm^2; and one at highest b-value) Target/Acceptable: 2 (including b=0-50 s/mm^2 and at highest b-value)	2	2	2	2
Minimum highest b-value	Ideal/Target: b=1000 s/mm^2 Acceptable: b=850-999 s/mm^2	1000 s/mm^2	1000 s/mm^2	1000 s/mm^2	1000 s/mm^2
Diffusion encoding directions	Ideal/Target: >3-orthogonal, combined gradient channels Acceptable: >3-orthogonal, single gradient channels	3-orthogonal, combined gradient channels	3-orthogonal, combined gradient channels	3-orthogonal, combined gradient channels	3-orthogonal, combined gradient channels
Slice thickness	Ideal: <4 mm	5 mm	5 mm	5 mm	5 mm

	Target: 4-5 mm Acceptable: 5mm				
Gap thickness	Ideal/Target: 0-1 mm Acceptable: 1-2 mm	2 mm	2 mm	2 mm	2 mm
Field-of-view	220-240 mm	250 x 250 mm	250 x 250 mm	250 x 250 mm	250 x 250 mm
Acquired Matrix (frequency x phase)	Ideal/Target: (160-256) x (160-256), or 1.5-1 mm in-plane resolution Acceptable: 128 x 128, or 1.7 mm in-plane resolution	128 x 128	128 x 128	128 x 128	128 x 128
Plane orientation	Transversal-axial	Transversal-axial	Transversal-axial	Transversal-axial	Transversal-axial
Phase-encode/ frequency-encode direction	Anterior-Posterior / Right-Left	Anterior-Posterior / Right-Left	Anterior-Posterior / Right-Left	Anterior-Posterior / Right-Left	Anterior-Posterior / Right-Left
Number of averages	Ideal/Target: ≥ 2 Acceptable:1	1	1	1	2
Half-scan factor	Acceptable/Target: >0.65	0.811	0.811	0.811	0.811
In-plane parallel imaging acceleration factor	Ideal: 2-3 Acceptable/Target: 2	factor = 2.5	factor = 2.5	factor = 2.5	factor = 2.5
TR	Ideal: > 5000 ms Acceptable/Target: 3000-5000 ms	3000 ms	3000 ms	3000 ms	3000 ms
TE	Ideal: <60ms Target: minimum TE Acceptable: <120 ms	< 60ms	< 60ms	< 60ms	< 60ms
Receiver Bandwidth	Ideal/Target: maximum possible in frequency encoding direction (minimum echo spacing) Acceptable: >1000 Hz/voxel	Maximum	Maximum	Maximum	Maximum

[j]Ideal: Meeting this specification may require extra effort or non-standard hardware or software, but is expected to provide better results th

an meeting the target.

ⅡTarget: Meeting this specification is achievable with reasonable effort and adequate equipment and is expected to provide better results th an meeting the acceptable specification.

ǂAcceptable: Actors that shall meet this specification to conform to this profile.

Abbreviations: SS-EPI, Single-Shot Echo Planar Imaging; STIR, Short TI inversion recovery; TE, Echo time; TR, Repetition time

Table 3. Definition of performance metrics using phantom imaging and QIBA claims according to QIBA DWI profile

Performance metrics		Definition	QIBA claim
Bias in ADC measurement	ADC bias = $\mu - DC_{true}$; or % bias = $100\% \frac{\mu - DC_{true}}{DC_{true}}$	μ : mean ADC (mm^2/s) within the ROI DC_{true}:ADC value for 0% PVP = 1.1×10^{-3} mm^2/s	Central measurement tube (0% PVP) ≤ 4%
Repeatability	RC = $2.77 \cdot \sigma_\omega$ wCV = $100\% \frac{\sigma\omega}{\mu}$	σ_ω: standard deviation	short-term RC < $1.5 \times 10^{-5} mm^2/s$ = 0.015 < $\mu m^2/s$ (wCV < 0.5%) long-term RC < $6.5 \times 10^{-5} mm^2/s$ = 0.065 < $\mu m^2/s$ (wCV < 2.2%)
Linearity	$\mu = \beta o + \beta 1 DC_{true}$		R-squared (R^2) of the linear model fit > 0.90 95% CI for the slope within the interval 0.95 to 1.05.
b-value dependence	ADC b-value dependence = $100\% \frac{ADC_{bmin,b2} - ADC_{bmin,b1}}{ADC_{bmin,b1}}$	b_{min} = b0	Maximum difference between any of ADC derived from variable b-values to their average ≤ 2% for central tube
Random measurement error	Random measurement error =100% $\frac{\sigma}{\mu}$		≤ 2% for central tube
SNR	SNR = $\frac{Spatial\ mean\ ROI\ on\ Signal\ Image}{Spatial\ mean\ of\ ROI\ on\ Noise\ image}$		b=0 SNR ≥ 45

Abbreviations: ADC, Apparent diffusion coefficient; RC, Repeatability coefficient; Signal-to-noise ratio; wCV, within-subject coefficient of variation

Table 4. Performance metrics of MRI scanners using phantom imaging acquired by the QIBA phantom protocol

Performance Metrics		QIBA claims	GE SIGNA Architect	Philips Ingenia CX	Siemens MAGNETOM Vida	Siemens Avanto
Category	**Metric**					
Accuracy	ADC bias (%)	Abs () ≤ 3.6	-0.4%	-2.3%	-2.1%	-1.5%
Repeatability	RC	< 15 $\mu m^2/s$	3.5 $\mu m^2/s$	8.5 $\mu m^2/s$	0.3 $\mu m^2/s$	2.9 $\mu m^2/s$
	wCV	≤ 0.5%	0.1%	0.3%	0.1%	0.1%
Linearity	Slope	0.95–1.05	0.98	0.98	0.98	1.00
	R^2	> 0.90	0.999	0.999	0.999	1.0
B-value dependence	Max b-value Dependence	≤ 2%	0.3%	0.6%	0.1%	1.5%
Precision	Random measurement error	≤ 2%	0.5%	0.7%	0.6%	1.7%
SNR	SNR for 0% water on b-value 0	≥ 45	105.2	56.0	230.1	117.4

Abbreviations: ADC, Apparent diffusion coefficient; RC, Repeatability coefficient; R^2, Coefficient of determination; SNR, Signal-to-noise ratio; wCV, within-subject coefficient of variation

Table 5. Performance metrics of clinical brain protocols using phantom imaging acquired by clinical protocols of each MRI scanner

Performance Metrics		QIBA claims	GE SIGNA Architect	Philips Ingenia CX	Siemens MAGNETOM Vida	Siemens Avanto
Category	**Metric**					
Accuracy	ADC bias (%)	Abs () ≤ 3.6	-3.1%	-0.9%	-1.5%	-0.7%
Repeatability	RC	< 15 $\mu m^2/s$	2.6 $\mu m^2/s$	1.2 $\mu m^2/s$	2.9 $\mu m^2/s$	3.1 $\mu m^2/s$
	wCV	≤ 0.5%	0.1%	0%	0.1%	0.1%
Linearity	Slope	0.95–1.05	0.95	1.00	0.97	0.99
	R^2	> 0.90	0.995	0.997	0.997	0.998
B-value dependence[†]	Max b-value Dependence	≤ 2%	NA	NA	NA	NA
Precision	Random measurement error	≤ 2%	0.5%	0.4%	0.4%	0.8%
SNR	SNR for 0% water on b-value 0	≥ 45	145.6	323.4	380.3	199.7

Abbreviations: ADC, Apparent diffusion coefficient; RC, Repeatability coefficient; R^2, Coefficient of determination; SNR, Signal-to-noise ratio; wCV, within-subject coefficient of variation

[†]Max b-value dependence was omitted in the institutional protocol validation because the protocol consist of two b-values.

Table 6. Image quality and repeatability measurement of patients' imaging

		GE SIGNA Architect	Philips Ingenia CX	Siemens Avanto
Image quality evaluation†				
Number of patients (scans)		N=3 (3 scans)	N=14 (17 scans)	N=17 (35 scans)
Low SNR		3±0	3±0	3±0
Ghost/parallel imaging artifacts		3±0	3±0	3±0
Severe spatial distortion		3±0	3±0	3.0±0.2
Eddy currents		3±0	3±0	3±0
Fat suppression		3±0	3±0	3±0
Motion artefacts		3±0	3±0	3±0
Nyquist ghost		3±0	2.9±0.3	3.0±0.2
Repeatability evaluation				
Number of patients		N= 0	N=3 (6 scans)	N=14 (28 scans)
White matter	ADC value ($\mu m^2/s$) †	N.A.	807.7±16.7	801.4±16.2
	RC ($\mu m^2/s$) ‡	N.A.	46.3	44.9
	wCV (%)‡	N.A.	2.1	2.0
CSF	ADC value ($\mu m^2/s$) †	N.A.	3079.4±63.6	3013.0±59.9
	RC ($\mu m^2/s$) ‡	N.A.	176.3	166.1
	wCV (%)‡	N.A.	2.1	2.00
Reproducibility evaluation				
Number of patients		N=3 (3 scans)	N=14 (17 scans)	N=17 (35 scans)
White matter	ADC value ($\mu m^2/s$) †	843.7±15.3	813.7±15.4	803.7±15.2
	RC ($\mu m^2/s$) ‡	42.5	42.7	42.1
	wCV (%)‡	1.8	1.9	1.9
CSF	ADC value ($\mu m^2/s$) †	3156.5±65.9	3114.7±63.5	3011.0±60.4
	RC ($\mu m^2/s$) ‡	182.6	175.9	167.2

wCV (%)‡	2.1	2.0	2.0

† Data are reported as the mean ± SD.

‡ QIBA claims for long-term repeatability, RC < 6.5x10-5 mm^2/s = 0.065 < $\mu m^2/s$ (wCV < 2.

Figure 1. Whole process of image standardization of DWI/ADC measurement using phantom validation and clinical application

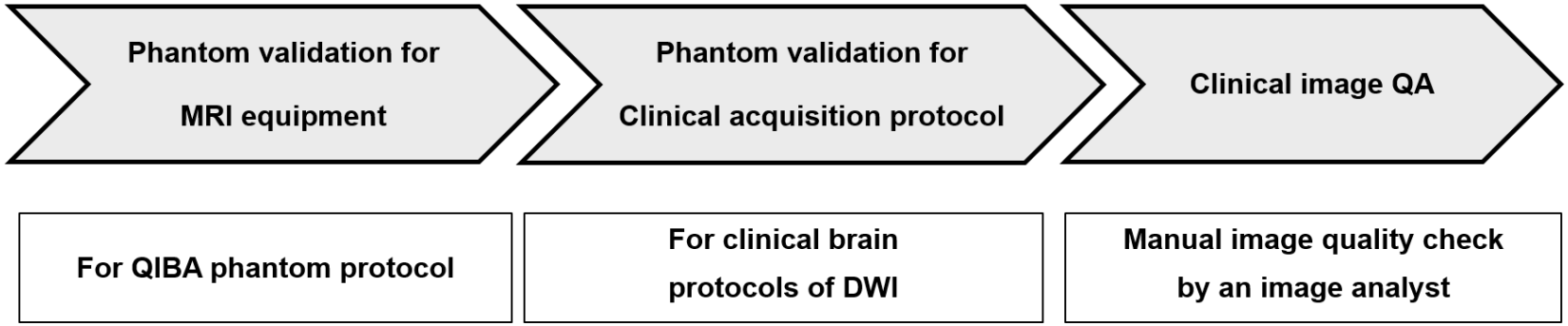


Figure 2. NIST/QIBA diffusion phantom

Figure 3. Example of ROI segmentation for brain white matter and CSF. A circled ROI was drawn at frontoparietal white matter (A) and lateral ventricle (B) on a b-1000 image and then transferred to an ADC map. The mean and standard deviation were calculated with the software.

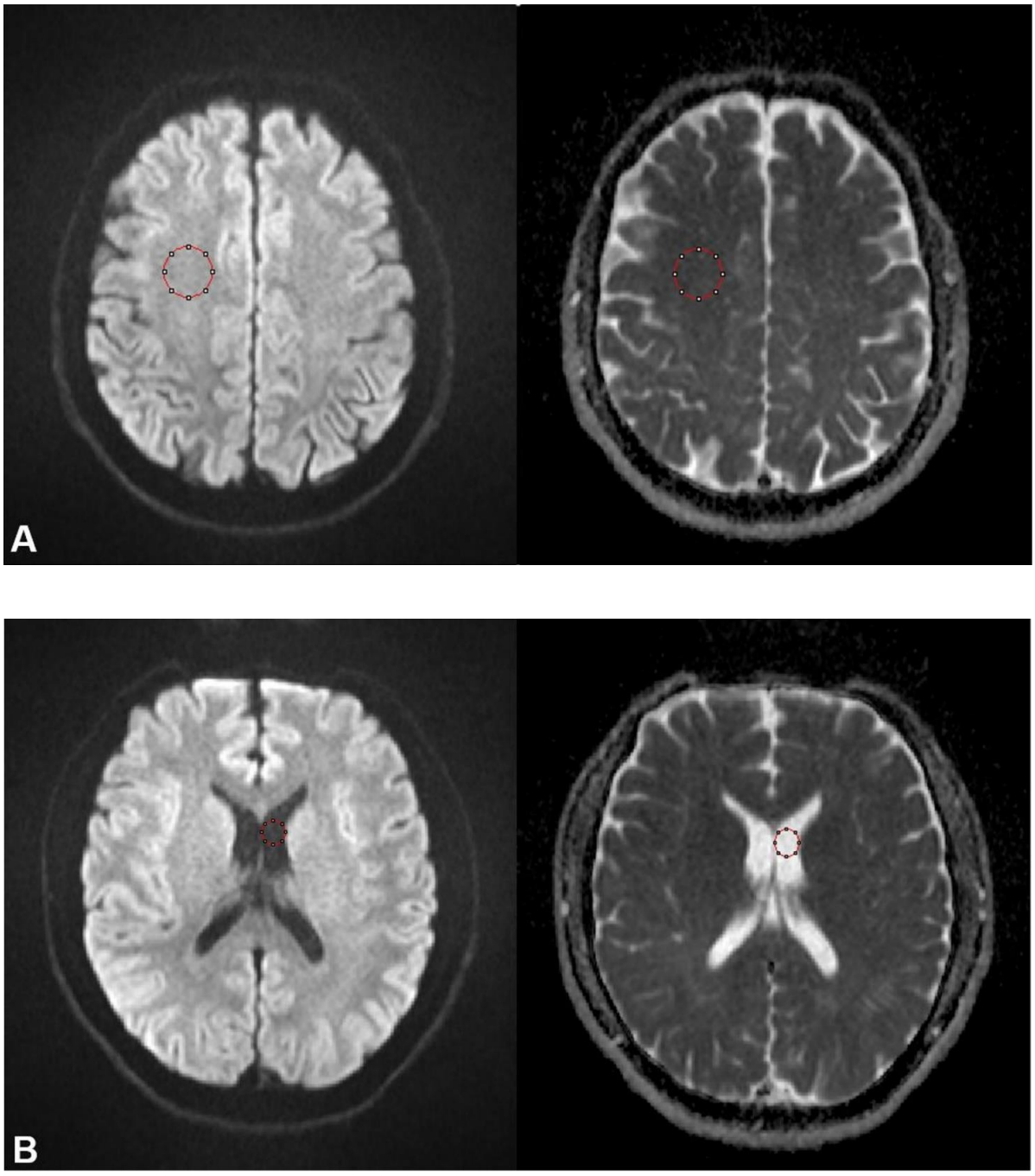


Figure 4. Representative images of DWI and ADC maps of phantom. T2-weighted images (A) and DWI with b-value of 0, 500, 1000 (B), 1500, 2000 s/mm^2 were obtained according to QIBA profile. Exponential ADC maps (C) were generated using all b-values and a mono-exponential model.

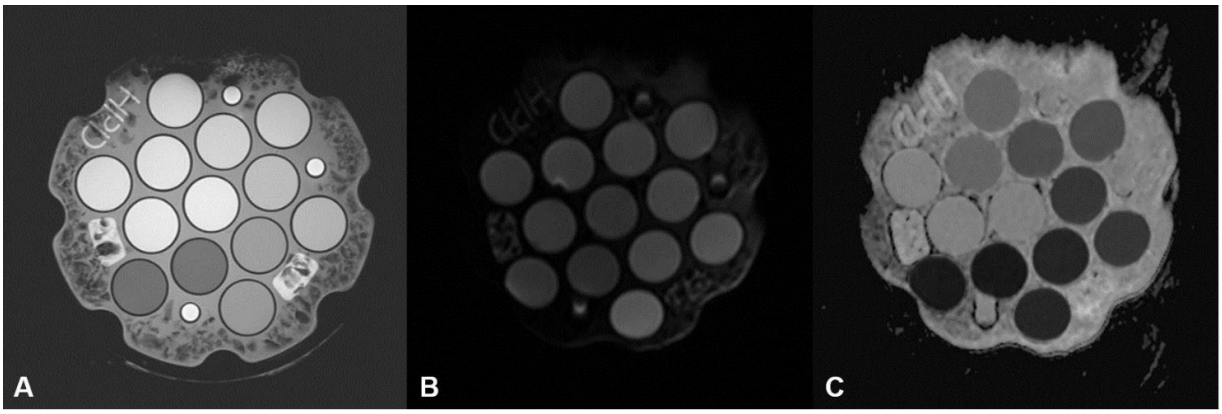
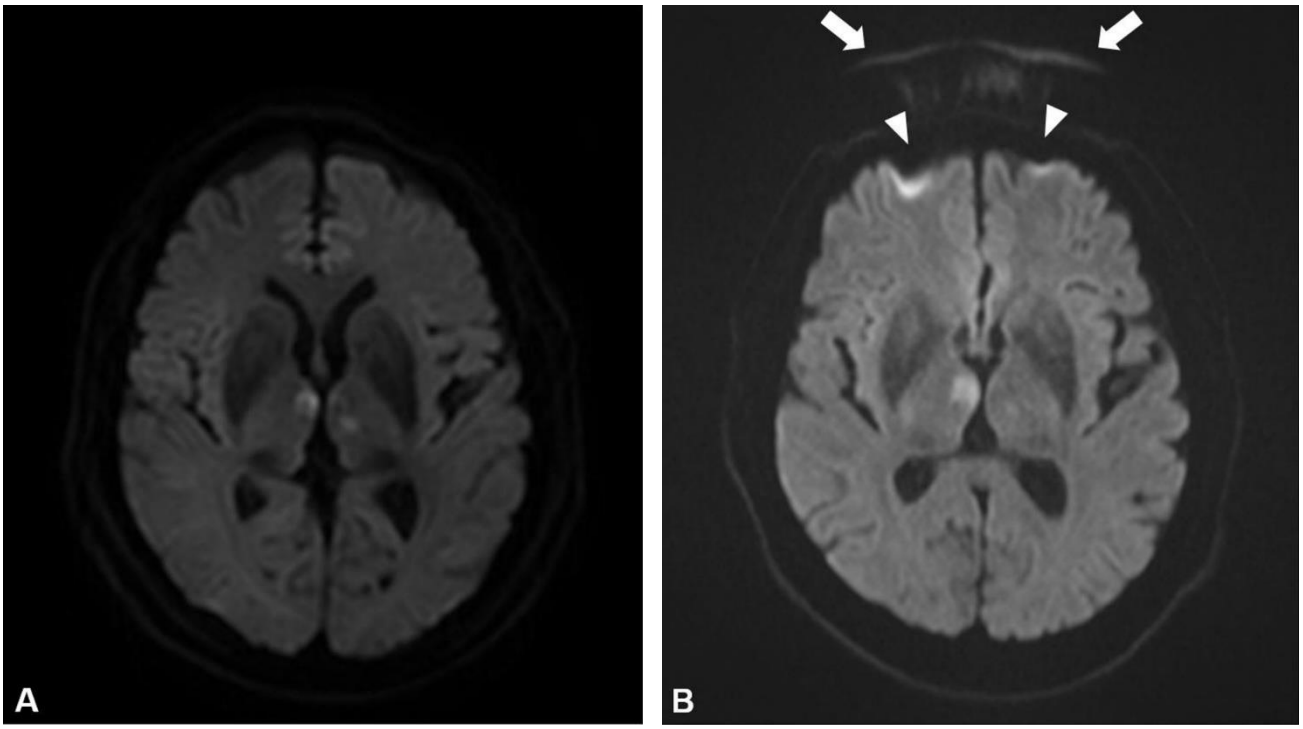


Figure 5. Examples of image quality evaluation of clinical DWI from stroke patients. This patient had a bilateral thalamic acute infarction. (A) A b-1000 image with ideal image quality. The image demonstrated the anatomical structure and tissue of interest without any artifacts. (B) An example of an image of the same patient with a Nyquist ghost artifact and spatial distortion. Nyquist ghost artifact with duplication of the frontal bone in the phase encoding direction (arrows) and mild distortion of both frontal lobes due to a susceptibility artifact (arrowheads). Spatial distortion and the Nyquist ghost artifact were scored as "acceptable" for this image.



Reference

1. Li SP, Padhani AR. Tumor response assessments with diffusion and perfusion MRI. Journal of magnetic resonance imaging : JMRI. 2012;35(4):745-63.

2. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia (New York, NY). 2009;11(2):102-25.

3. Thomassin-Naggara I, Dechoux S, Bonneau C, Morel A, Rouzier R, Carette MF, et al. How to differentiate benign from malignant myometrial tumours using MR imaging. European radiology. 2013;23(8):2306-14.

4. van der Hoogt KJJ, Schipper RJ, Winter-Warnars GA, Ter Beek LC, Loo CE, Mann RM, et al. Factors affecting the value of diffusion-weighted imaging for identifying breast cancer patients with pathological complete response on neoadjuvant systemic therapy: a systematic review. Insights into imaging. 2021;12(1):187.

5. Barnes A, Alonzi R, Blackledge M, Charles-Edwards G, Collins DJ, Cook G, et al. UK quantitative WB-DWI technical workgroup: consensus meeting recommendations on optimisation, quality control, processing and analysis of quantitative whole-body diffusion-weighted imaging for cancer. Br J Radiol. 2018;91(1081):20170577.

6. Galbán CJ, Hoff BA, Chenevert TL, Ross BD. Diffusion MRI in early cancer therapeutic response assessment. NMR Biomed. 2017;30(3).

7. García-Figueiras R, Padhani AR, Baleato-González S. Therapy Monitoring with Functional and Molecular MR Imaging. Magn Reson Imaging Clin N Am. 2016;24(1):261-88.

8. Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res. 2009;15(3):986-94.

9. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015;277(3):813-25.

10. Fedeli L, Benelli M, Busoni S, Belli G, Ciccarone A, Coniglio A, et al. On the dependence of quantitative diffusion-weighted imaging on scanner system characteristics and acquisition parameters: A large multicenter and multiparametric phantom study with unsupervised clustering analysis. Phys Med. 2021;85:98-106.

11. Giannelli M, Sghedoni R, Iacconi C, Iori M, Traino AC, Guerrisi M, et al. MR scanner systems should be adequately characterized in diffusion-MRI of the breast. PLoS One. 2014;9(1):e86280.

12. Malyarenko D, Galbán CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, et al. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging. 2013;37(5):1238-46.

13. Mattiello J, Basser PJ, LeBihan DJJomr, Series A. Analytical expressions for the b matrix in NMR diffusion imaging and spectroscopy. 1994;108(2):131-41.

14. Gallichan D, Scholz J, Bartsch A, Behrens TE, Robson MD, Miller KLJHbm. Addressing a systematic vibration artifact in diffusion-weighted MRI. 2010;31(2):193-202.

15. Liu Z, Wang Y, Gerig G, Gouttard S, Tao R, Fletcher T, et al. Quality Control of Diffusion Weighted Images. Proc SPIE Int Soc Opt Eng. 2010;7628:76280J.

16. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. J Magn Reson Imaging. 2019;49(7):e101-e21.

17. Palacios EM, Martin AJ, Boss MA, Ezekiel F, Chang YS, Yuh EL, et al. Toward Precision and Reproducibility of Diffusion Tensor Imaging: A Multicenter Diffusion Phantom and Traveling Volunteer Study. AJNR Am J Neuroradiol. 2017;38(3):537-45.

18. Paquier Z, Chao SL, Bregni G, Sanchez AV, Guiot T, Dhont J, et al. Pre-trial quality assurance of diffusion-weighted MRI for radiomic analysis and the role of harmonisation. Phys Med. 2022;103:138-46.

19. Yung JP, Ding Y, Hwang K-P, Cardenas CE, Ai H, Fuller CD, et al. Quantitative Evaluation of

apparent diffusion coefficient in a large multi-unit institution using the QIBA diffusion phantom. 2020.

20. qCal-MR quality control (QC) Software [Available from: https://qcalsoftware.com/.

21. Jones DK, Cercignani M. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. 2010;23(7):803-20.

22. Asan Image Metrics AsanJ-Stroke software [Available from: https://aim-aicro.com/software/strokevolumetry.

23. Helenius J, Soinne L, Perkiö J, Salonen O, Kangasmäki A, Kaste M, et al. Diffusion-weighted MR imaging in normal human brains in various age groups. AJNR Am J Neuroradiol. 2002;23(2):194-9.

24. Holz M, Heil SR, Sacco AJPCCP. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. 2000;2(20):4740-2.

25. Quantitative Imaging Biomarker Alliance. QIBA profile: Diffusion-weighted magnetic resonance imaging (DWI). 2020.

영문요약

Background: For the apparent diffusion coefficient (ADC) to be used reliably as a biomarker, it is essential to validate MRI equipment performance and clinical acquisition protocols before applying them to patients. This study aims to validate various MRI equipment and clinical brain protocols for diffusion-weighted imaging (DWI) using a commercial phantom and to confirm the validated protocols in patient images.

Materials and Methods: The study validated the performance of four different MRI scanners and clinical brain protocols using a Quantitative Imaging Biomarker Alliance (QIBA) diffusion phantom and a cloud-based analysis tool. Performance metrics, including accuracy and repeatability of ADC measurements, were evaluated using the QIBA profile. The validated clinical brain protocols were then applied to 17 patients, and both image quality and ADC repeatability were assessed.

Results: All four MRI scanners demonstrated high accuracy in ADC measurement (ADC bias, -2.3% to -0.4%), excellent linear correlation with the reference ADC value (slope, 0.9 to 1.0; R^2, 0.999–1.000), and high short-term repeatability (within-subject coefficient of variation (wCV), 0% to 0.3%). The clinical protocols also met QIBA claims with high accuracy (ADC bias, -3.1% to -0.7%) and robust repeatability (wCV, 0% to 0.1%). Brain DWI obtained using the validated clinical protocols exhibited excellent image quality (mean score ≥ 2.9) and good repeatability (wCV, 1.8–2.2).

Conclusion: The comprehensive standardization process of DWI demonstrated that ADC measurements are highly accurate and repeatable across different MRI equipment and clinical protocols, in line with QIBA claims.