의학박사 학위논문

# 표준 유방 확산강조영상 팬텀을 이용한 자기공명영상 기기의 품질 평가와 다기관 비교

# Quality Evaluation and Multicenter Comparison of MR Scanners Using Standard Breast Diffusion-Weighted Imaging Phantom

울 산 대 학 교 대 학 원

의 학 과

유 현 경

# Quality Evaluation and Multicenter Comparison of MR Scanners Using Standard Breast Diffusion-Weighted Imaging Phantom

지 도 교 수　　신 희 정

이 논문을 의학박사 학위 논문으로 제출함

2020 년 02 월

울 산 대 학 교　대 학 원

의　　학　　과

유 현 경

유현경의 의학박사학위 논문을 인준함

| | | |
|---|---|---|
| 심사위원 | 김 학 희 | (인) |
| 심사위원 | 차 주 희 | (인) |
| 심사위원 | 신 희 정 | (인) |
| 심사위원 | 이 종 원 | (인) |
| 심사위원 | 문 우 경 | (인) |

울 산 대 학 교 대 학 원

2020 년 02 월

## English Abstract

**Purpose:**

The purpose of this study was to compare apparent diffusion coefficient (ADC) value, spatial resolution, and signal-to-noise ratio (SNR) of diffusion-weighted imaging (DWI) obtained from nine 3-T magnetic resonance (MR) scanners from six institutions, using standard breast DWI phantoms for quality evaluation of quantitative DWI.

**Materials and methods:**

Nine MR scanners from six institutions were evaluated using a standard breast DWI phantom (QalibreMD$^{TM}$). The DWI scan protocol was standardized as much as possible, given the constraints of using multiple scanner platforms. DWI was performed before and after phantom repositioning using three *b*-values (0, 800, 1200 s/mm$^2$), single-shot or readout-segmented echoplanar imaging (EPI). In addition, further DWI was performed using two *b*-values (0 and 1000 s/mm$^2$) immediately before repositioning in order to assess the effect of different *b*-value combinations. We assessed artifacts, fat suppression, SNR, and ADC values. We also evaluated circularity and areas of variable-sized pattern holes on the spatial resolution plate.

Repeatability and reproducibility were evaluated using a Bland–Altman plot, within-subject coefficient of variation (wCV), intraclass correlation coefficient (ICC), agreement index (AI), and repeatability coefficient (RC).

**Results:**

Each ADC value obtained from test tubes with variable polyvinylpyrrolidone concentrations was matched with the corresponding reference ADC value. Mean ADC values showed no significant difference ($p$ = 0.456) across three vendors and across different institutions. ADC values were not significantly different between two groups of $b$-value combinations (set 1 as test $b$ = 0, 800, 1200 vs. set 2 $b$ = 0, 1000, $p$ = 0.826: and set 2 vs. retest, $p$ = 0.525). Temperatures at the time of acquisition were 20℃ (n = 5) and 21℃ (n = 4), and mean ADCs were not significantly different according to temperature ($p$ = 0.262). SNRs were significantly different according to institutions and vendors ($p$ < 0.001 for both). As $b$-values increased, SNR decreased. ADC repeatability between test and retest was excellent, with an ICC of 0.98 (95% confidence interval [CI] 0.97, 0.99) and wCV of 8.1% (95%CI 6.2, 11.6 %). However, ICC was 0.23 (95%CI 0.13, 0.42) and wCV was 113.2% (95%CI 87, 161.6%) for SNR

repeatability, indicating poor repeatability. SNR differed significantly within the same vendor (vendor B, $p < 0.05$). In terms of spatial resolution, there was a significant difference according to the location from the nipple (anteroposterior direction) and central plate (right-to-left direction). On the spatial resolution plate, circularity of the posterior pattern holes was better than that of the anterior pattern holes ($p < 0.001$) and circularity of central pattern holes was better than that of the peripheral pattern holes ($p < 0.05$). Circularity differed statistically significantly within the same vendor, for both the axial and the sagittal set ($p < 0.05$).

**Conclusion:**

In the multicenter setting, mean ADC can be measured with excellent repeatability and reproducibility using a standardized DWI protocol and a standard phantom. SNR may be affected by *b*-values and vendors, while spatial resolution, including circularity and area, is not affected by vendor, but varies by the location within the breast coil. In our multicenter test using standard DWI phantom and a standard acquisition protocol, ADC values and spatial resolution was not affected by vendors, while SNR was affected by vendors and *b*-values.

**Key Words**

Breast

Standard breast DWI Phantom

Quality Evaluation

Multicenter Comparison

**CONTENTS**

## LIST of TABLES AND FIGURES

## Introduction

Breast cancer is the most commonly diagnosed cancer and was the leading cause of cancer death among women, and the second most common leading cause of cancer death in 2018[1]. Breast dynamic contrast-enhanced magnetic resonance imaging (DCE MRI) has become a standard method for evaluating breast cancer preoperatively and in the high-risk screening setting as well as for treatment response monitoring. Diffusion-weighted imaging (DWI) is a functional imaging technique which reflects water diffusion properties within the tissue. Growing evidence supports DWI as a supplemental and/or alternative technique to DCE MRI[2-4]. The apparent diffusion coefficient (ADC) of DWI reflects impediment of water mobility by cellularity and interstitial tortuosity[5] and can be used as a potential quantitative biomarker; it can be obtained rapidly and without administration of contrast media from DWI[3, 4]. A recent study evaluated ADC as an imaging biomarker for monitoring therapeutic effect and reported that tumor ADC can be measured with excellent repeatability and reproducibility in a multicenter setting, using a standardized protocol and AQ procedure[2]. Furthermore, research on DWI screening has been reported[6, 7], and a multicenter DWI screening trial

commenced this year[8]. The effective implementation of the DWI could vary with the MR scanner and protocol. Image quality, including the repeatability and reproducibility of ADC and SNR measurements, should be adequately assessed, especially in multicenter studies where site variability in the quantification of DWI could reduce the sensitivity and specificity of the study[3].

Various DWI phantoms have been developed with various materials, such as polyvinylpyrrolidone (PVP)[9], polyethylene glycol[10], and nickel agarose and sucrose gel[11] in order to improve the reproducibility of ADC measurements and standardization of DWI. Jerome et al.[12] developed a combination phantom of PVP and ice-water as a temperature-controlled medium and reported that this phantom allows reliable quality assurance measurements that can be used to determine agreement between MRI scanners. Two previous studies evaluated the repeatability and reproducibility of ADC measurement[13, 14]. Malyarenko et al.[13] reported that the standard deviation (SD) of bore-center ADCs measured across 35 scanners was lower than 2%, the day-to-day repeatability of the measurements was within 4.5%. However, significant (> 10%) vendor-specific and system-specific spatial

nonuniformity ADC bias was detected for the off-center measurement, which was consistent with gradient nonlinearity. Newitt et al.[14] reported that the concordance of the majority of implementations was excellent for both phantom and in vivo ADC measurements. However, despite overall good concordance, implementation biases in ADC measures might sometimes be significant and might be sufficiently large to be of concern in multicenter studies. Recently, a sub-study from the American College of Radiology Imaging Network ACRIN 6698 Trial[2] reported that ADC repeatability and reproducibility was excellent in this multicenter setting using a standardized protocol and quality assurance (QA) procedure. However, they did not evaluate the spatial resolution of DWI. In contrast to a diagnostic setting, a screening setting needs a high spatial resolution in order to detect small cancers. However, only a few studies evaluating spatial resolution on MR phantom images have been reported[13, 15, 16].

Therefore, the purpose of this study was to compare ADC measurements, spatial resolution, and the SNR of DWI obtained from nine 3-T MR scanners from six institutions, using standard breast DWI phantoms for quality evaluation of quantitative DWI.

## Materials and Methods

### *Breast DWI Phantom*

We used a standard breast DWI phantom (QalibreMD™, Boulder, CO, USA) developed in collaboration with the National Institute for Standards and Technology (NIST), the University of California at San Francisco (UCSF)[17]. It consists of two distinct parts that are mimics designed to match the ADC, T1, and T2 of breast tissues, and that are surrounded by a flexible silicone shell; its design allows it to be easily applied to existing breast coils. The two parts are connected by a backboard made of polycarbonate, and we evaluated diffusion (ADC) sites only (Fig. 1). We placed this standard DWI phantom into the dedicated breast coil and obtained axial and sagittal scans using a standard scan protocol (Table 1). Interstitial solutions of the phantom are fiber mimics and the contents of each test tube vial are composed of PVP (Fig. 1A and 1B). The 10, 14, and 18% weight-by-weight (w/w) PVP tubes were used for determination of ADC measurement resolution, to decipher the sensitivity of ADC measurements for distinguishing between benign and malignant lesions. The differences between the expected ADC values of 10% and 18% w/w PVP were similar to the differences

between benign and malignant tumors: $1.61 \pm 0.33 \times 10^{-3}$ $mm^2/s$ compared to $1.25 \pm 0.29 \times 10^{-3}$ $mm^2/s$[15, 18, 19]. In terms of spatial resolution, there are 14 resolution pattern holes and 2 resolution insets. On the axial scan, each resolution pattern is a row of holes with sizes: 10 mm, 8 mm, 6 mm, 4 mm, 2 mm, and 1 mm with a center spacing of 11 mm, 9 mm, 7 mm, 5 mm, and 3 mm, respectively (Fig. 1C). Each resolution inset has patterns and sub-patterns of holes with 1.25 mm, 1 mm, 0.75 mm, 0.5 mm diameter. The sub-patterns are rotated 10 degrees from each other. On the sagittal scan, there were 42 10-mm-sized holes with 15-mm spacing across both halves (Fig. 1D).

***Scanners and DWI Acquisition Protocol***

From February 2019 to November 2019, DWI was performed using a standard DWI phantom and a standard scan protocol for nine 3-T MR scanners from the following institutions; Asan Medical Center (AMC), Samsung Medical Center (SMC), Seoul National University Hospital (SNUH), Seoul National University Bundang Hospital (SNUBH), Severance Hospital (Severance), and National Cancer Center (NCC). The nine 3-T MR scanners included

one Architect (GE Healthcare, Milwaukee, WI), two Skyra (Siemens Medical Solutions, Erlangen, Germany), and six Ingenia or Ingenia CX (Philips Medical System, Amsterdam, Netherlands).

The DWI scan protocol was standardized as much as possible, given the constraints of the three scanner platforms (Siemens, Philips, and GE). We attempted to employ the protocol proposed by QIBA (The Radiology Society of North America's Quantitative Imaging Biomarkers Alliance)[20] and the EUSOBI (European Society of Breast Radiology guidelines)[21] and focused on spatial resolution, because a screening study requires high spatial resolution in order to detect small cancers. Our DWI scan protocol was as follows (Table 1). For DWI scans, we used a single-shot (ss) or readout-segmented (rs) echoplanar sequence (EPI), using eight, 16, and 18 channels with dedicated bilateral breast coils, spectral attenuated inversion recovery (SPAIR), or short T1 inversion recovery (STIR) fat suppression; two sets of *b*-value combinations ($b$ = 0, 1000 s/mm$^2$, and $b$ = 0, 800, 1200 s/mm$^2$); 3-mm slice thickness with no gap; and in-plane spatial resolution less than 1.3 mm. For test–retest assessment, we performed DWI using a combination of three *b*-values (set 1, $b$ = 0, 800, 1200 s/mm$^2$), and subsequently

with a combination of two *b*-values (set 2, $b = 0$ and 1000 s/mm$^2$). Then, we removed the DWI phantom from the bore and replaced it with a phantom within a breast coil, thereafter repeatedly performed DWI with the three *b*-value combination (set 3, $b = 0$, 800, 1200 s/mm$^2$). Two different sets of DWI using two *b*-values ($b = 0$ and 1000 s/mm$^2$) and three *b*-values were performed sequentially in order to assess the effect of different *b*-value combinations.

***ADC and SNR analysis***

Digital imaging and communications in medicine (DICOM) files of DWI scans were transferred to Image J and assessed using a standardized QA protocol for the following categories: artifacts, fat suppression, and SNR. A row of circular regions-of-interest (ROIs) was placed at the center of the test tube, with variable percentage PVP, in order to exclude edge artifacts, and we obtained mean ADC values by averaging the value from each ROI within the same PVP test tube (Fig. 2A). ADC values were calculated by the following equation: $S_D = S_0 * e^{-b * ADC}$, where $S_D$ is the diffusion-weighted signal intensity, $S_0$ is the signal intensity without diffusion-weighting, *b* is the diffusion sensitization factor, which varies by

the strength and timing of the applied diffusion gradients (in s/mm$^2$), and the ADC is the rate of diffusion, defined as the average area occupied by a water molecule per unit time (in mm$^2$/s). An ADC map can be calculated using image acquisitions at two or more different *b*-values, quantitatively reflecting a composite of tissue factors affecting net water mobility in each voxel including microcirculation, cellular density, organization, and membrane integrity [22].

SNR was defined as a ratio between the mean signal intensity (SI) within the ROI ($S_{ROI}$) and the SD of the background noise ($\sigma_{BG}$) ($SNR = S_{ROI} / \sigma_{BG}$)[23-25]. SNR is an important factor for image quality, and low SNR at a high *b*-value image could bias he estimation of ADC. In our study, we calculated SNR by measuring the average of a row of ROIs in each test tube and then divided this by an average of a row of ROIs on the background at each *b*-value DWI (Fig. 2B). Test–retest repeatability and inter-institutional reproducibility were evaluated. Previous studies[3, 13, 26] have reported that ADC values were closely related to temperature. For this reason, we recorded the room temperature during the DWI acquisition. Sufficient time was allowed for the phantom to achieve thermal equilibrium to the scan room (> 1 hour), and then DWI was performed after the temperature of the phantom was equal to that of room

temperature.

***Spatial Resolution***

Spatial resolution is an important factor for detecting small cancer, especially in a screening setting. Poor spatial resolution can cause image distortion as well as ADC misregistration[2, 20]. In our study, we used a standard DWI phantom, in which there were two spatial resolution plates: the axial and sagittal sets.

We first calculated the percentage of visualized pattern holes for subjective evaluation of spatial resolution in the axial set, by size and *b*-values. An example image is shown in Fig. 1C and 1D; one radiologist with 9 years of experience in breast MRI evaluated the degree of distortion and the circularity of pattern holes on the spatial resolution plates, using ASAN J, which is an in-house software modified from Image J (public domain Java image processing program inspired by NIH)[27]. The radiologist clicked each pattern hole and an ROI was semi-automatically generated. Then, the software calculated the "area" and "circularity" of each pattern hole[27]. A value of 1.0 indicated a perfect circle, and the shape was

increasingly elongated as the value approached 0.0. It was calculated by the equation: $4\pi \times$ area / perimeter$^2$. Circularity determinations may not be valid for the smallest hole. In addition, we analyzed the following parameters: aspect ratio (AR), and roundness (Round). AR indicates the major axis/ minor axis, while Round was calculated by $4 \times$ area/ ($\pi \times$ major axis$^2$), or the inverse of the aspect ratio.

On the axial scan, there were 14 groups with variable pattern holes; in one group, there were six different sizes of pattern holes. Among them, we evaluated three groups of holes on the $b = 0$ image, as shown in Fig. 1C. On the sagittal scan, there were 42 pattern holes and among these, we selected 24 holes (four sets of six holes). These 24 holes included two horizontal rows, for analyzing the effect of right-to-left direction, and two vertical rows, for analyzing the effect of anteroposterior direction (Fig. 1D).

***Statistical Analysis***

The linear mixed-effect model, as a multilevel test, was used to evaluate the possible differences in ADC and SNR values between test tubes with varying percentages of PVP, and

between different vendors. We used scanners and PVP% test tubes as fixed effects, and institutions and set numbers (test and retest) as random effects. Statistical analyses were performed using R 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). Repeatability and reproducibility of ADC and SNR were assessed using the intraclass correlation coefficient (ICC), within-subject coefficient of variation (wCV), agreement index (AI), and repeatability coefficient (RC) [2, 28]. Both RC and wCV are based on-within subject SD (wSD), RC = 2.77 × wSD[20], and wCV = 100% × wSD / mean. RC has the same unit as a marker, while wCV is unit-less. ICC was derived from a two-way random-effects model for absolute agreement, it ranged from –1 to 1 and was unit-less. For wCV, values less than 10% indicates very good or excellent agreement, from 10 to less than 20% means good agreement, from 20 to less than 30% means acceptable agreement, and from 30% upward indicates that agreement is not acceptable[29].

## Results

### *ADC Values and SNR*

The ADC values of various PVP test tubes approximated the range of ADC values from malignant tumors to normal breast tissue. The values obtained for malignant tumor to benign tissue in previous in vivo clinical studies using 1.5-T systems[18, 19] were as follows ($\times 10^{-3}$mm$^2$/s): benign mass 1.72 ± 0.43 to 1.74 ± 0.46, benign lesion 1.61 ± 0.33, malignant lesion 1.41 ± 0.22, and malignant mass 1.25 ± 0.29 to 1.32 ± 0.23. A comparison of the results of a previous phantom study[15] and our acquired ADC values is shown in Table 2; our results are also shown in Fig. 3. Mean ADC values were not significantly different across all scanners from three vendors (Fig. 4A) ($p$ = 0.456). In a comparison of the mean ADC of three vendors, the $p$-value was 0.880 for vendors A versus B, 0.377 for vendors A versus C, and 0.266 for vendors B versus C, respectively. Mean ADC values according to the $b$-value combination was also not significantly different ($p$ = 0.826 for set 1 versus set 2, and $p$ = 0.525 for set 2 versus set 3). Temperatures at the time of acquisition were 20℃ (n = 5) to 21℃ (n = 4), and mean ADCs did not differ significantly according to the temperature ($p$ = 0.262). SDs of ADC showed

statistically significant differences across all three vendors ($p < 0.001$), but did not differ significantly according to the temperature ($p = 0.111$) and *b*-value combinations ($p = 0.543$ for set 1 versus set 2, $p = 0.572$ for set 2 versus set 3).

The SNR showed statistically significant differences between the three vendors (Fig. 4B) ($p < 0.001$). The mean value and SD of SNR were 84.43 and 67.98 for vendor A, 285.34 and 333.81 for vendor B, 211.67 and 139.28 for vendor C, respectively, and differed significantly between vendor A versus vendor B, vendor B versus vendor C, and vendor A versus vendor C ($p < 0.001$ for all). As the *b*-value increased, SNR decreased ($p < 0.001$). The mean and SD of SNR were 398.75 and 384.38 for $b = 0$, 145.97 and 110.73 for $b = 800$, 128.00 and 85.63 for $b = 1000$, and 92.73 and 71.31 for $b = 1200$, respectively (Fig. 4C). SNR was not significantly different according to the temperature ($p = 0.802$). SNR differed statistically significantly according to different *b*-value combinations: both $p < 0.05$ in the comparison of set 1 versus set 2, and of set 2 versus set 3. Among all acquired DW images from three vendors, the Nyquist artifact was noted in one vendor (Fig. 2A and B, asterisk).

***ADC and SNR Repeatability and Reproducibility***

The test and retest repeatability of the mean ADC obtained from various PVP% test tubes are shown in Fig. 5A, while SNR repeatability is shown in Fig. 5B. The overall mean ADC ranged from 0.069 to 2.298 × $10^{-3}$ $mm^2/s$ and SD ranged from 0.001 to 0.39 × $10^{-3}$ $mm^2/s$. For overall mean ADC measurements, wCV = 8.1% (95%CI 6.2, 11.6 %), ICC = 0.98 (95%CI 0.97, 0.99), RC = 0.23 × $10^{-3}$ $mm^2/s$ (95%CI 0.22, 0.25), and AI = 0.84 (95%CI 0.74, 0.87), indicating excellent repeatability. Overall mean SNR ranged from 4.08 to1993.0, and the SD ranged from 10.08 to 993.26. For the overall mean SNR, wCV = 113.2 % (95%CI 87, 161.6 %), ICC = 0.23 (95%CI 0.13, 0.42), RC = 687.7 (95%CI 658.9, 719.1), and AI = 0.60 (95%CI 0.51, 0.67), which indicates poor repeatability.

***Spatial Resolution***

In the axial set, we calculated the percentage of visualized pattern holes by size and *b*-values for subjective evaluation of spatial resolution. On $b$ = 0 images, in 55.6% of all scanners, 1-mm pattern holes were visualized in groups 1 and 2, and 33.3% in group 3. On *b*

= 800 and $b$ = 1000 images, only one or two scanners could detect 1-mm-size pattern holes, and no scanner visualized 1-mm pattern holes with $b$ = 1200. For 2-mm pattern holes, 100% were visualized with all $b$-values in groups 1 and 2, and 77.8% were visualized with all $b$-values in group 3. In 85.7% of scanners, pattern holes larger than 4 mm were all visualized with all $b$-values and in all groups, except for one scanner. That scanner could detect more than 8-mm pattern holes with $b$ = 1200 in group 3 but could detected more than 4 mm pattern holes in all groups on $b$ = 0 images. These results were shown in Table 3.

Overall "circularity" on axial scans showed no significant differences between the three vendors ($p$ = 0.985 for across three vendors, $p$ = 0.895 for vendor A versus B, $p$ = 0.988 for vendor A versus C, and $p$ = 0.911 for vendor B versus C). As shown in Fig. 1C, on the axial scan, three groups of pattern holes were divided according to the location. The circularity showed significant differences between groups 1, 2, and 3. Circularity of group 1 was better than that of groups 2 and 3 ($p < 0.001$ for both). In each group, there were six different sizes of pattern holes; the circularity of circles 2 to 6 was not significantly different, compared with circle 1 (the largest circle) ($p$ = 0.467, 0.955, 0.794, 0.644, and 0.209, respectively).

For “area”, from the largest to the smallest circle, the values ($mm^2$) were as follows; 78.5, 50.24, 28.26, 12.56, 3.14, and 0.8, respectively. These results were calculated based on the physical size of each hole in the phantom. Areas obtained on the axial scan were as follows (mean ± SD pixels); 105.06 ± 69.6, 94.64 ± 90.44, 86.85 ± 105.76, 78.66 ± 115.87, 75.33 ± 115.00, and 2.0 ± 1.26, respectively. As 1 pixel is 1.3 mm, and area of 1 × 1 pixel is 1.69 $mm^2$; thus, the areas obtained were as follows ($mm^2$): 177.55 ± 117.62, 159.94 ± 152.84, 146.78 ± 178.73, 132.94 ± 195.82, 127.31 ± 194.35, and 3.38 ± 2.13, respectively. Overall areas showed no significant differences across the three vendors ($p = 0.713$ for overall area, $p = 0.491$ for vendor A versus B, $p = 0.490$ for vendor A versus C, and $p = 0.999$ for vendor B versus C). These results are summarized in Table 4 and Fig. 6A.

On the sagittal scan, circularity was significantly decreased from posterior to anterior ($p < 0.001$) as well as from the center to the periphery ($p < 0.05$) within the same scanner. The differences of circularity in the anteroposterior direction were more significant ($p < 0.001$) than those in the right-to-left direction ($p = 0.006$). The circularity of each hole showed no significant difference across vendors and scanners (between A–A', $p = 0.405$ and between B-

B', $p = 0.247$). In terms of area, however, there was a significant difference between A–A' ($p < 0.001$) and between B–B' ($p = 0.003$) across the three vendors. The results of vendor A were similar to those of vendor C, but the results of vendor B differed significantly from those of vendor A and of vendor C ($p < 0.05$ for both). In terms of AR and roundness, the results of vendor C differed significantly from those of vendor A and vendor B ($p < 0.05$ for both). These results are also summarized in Table 4 and Fig. 6B. There were six different scanners associated with vendor B and two scanners with vendor C. Among them, the circularity, area, AR, and round all showed significant differences for different scanners from the same vendor, in both the axial and the sagittal set ($p < 0.05$).

## Discussion

The purpose of this study was to compare ADC measurement, spatial resolution, and SNR of DWI using standard breast DWI phantoms for quality evaluation of quantitative DWI. For this study, both the phantom study and quality control of DWI could be an important issue. First, in a phantom, Keenan et al.[15] suggested that the ideal breast phantom has the following characteristics: 1) it can simulate heterogeneous distributions of fat and fibroglandular tissue in breast parenchyma, 2) it can allow quantitative evaluation of T1, T2, and diffuse imaging protocols, 3) it has high compatibility with commonly used breast coil types, 4) it allows evaluation of the left and right side of the coil, and 5) it can measure image distortion. In previous studies, the DWI phantoms or acquisition protocol varied[9-11, 26], and they mostly used a head coil[13, 30, 31]. In several previous studies[32-34], the reproducibility of the ADC measurements was particularly good at both field strengths of 1.5 T and 3.0 T, for various vendors, and there was good agreement with various phantom studies, but these did not use a breast-specific phantom. The phantom we used in this study allows ADC measurements and physically fit into the MRI breast coil, maximizing functionality. A previous study[15] analyzed

similar phantoms; they reported that the breast phantom can serve as a quality control tool to facilitate standardization of quantitative measurements for breast MRI.

Second, in the quality control of DWI, Chenevert et al.[26] reported that the ADC value of ice water is approximately $1.1 \times 10^{-3}$ $mm^2/s$ if the temperature is maintained at 0□; temperature is thus an important factor. Keenan et al.[15] reported the ADC values for different PVP concentrations, and their results approximated those of our study, but still differed from our measurements although the values were similar to those of previous in vivo studies[18, 19]. This may be due to different vendors or different MR scanners (1.5- and 3-T scanners in the previous study versus all 3-T in our study), but may also be due to different definitions of room temperature. Their study was performed in a room temperature ranging from 17 to 24℃, while our room temperature ranged from 20 to 21℃. Wagner et al.[35] calibrated measurements to the sample temperature. They reported that, at lower concentrations, such as 10% PVP, good agreement was observed, but at larger concentrations, such as 40% PVP, the deviation became as large as 10% in the temperature calibrations, due to the high viscosity of larger concentrations. According to their study, ADC measurements without temperature calibration

or temperature fixation do not seem to be advisable, because the observed temperature dependency appears to be too great to be acceptable. Numerous previous studies[9-12, 32, 33] reported that the diffusion properties of all tested fluids depended on their temperature. Thus, room temperature should be considered in terms of the difference in our results from those of previous studies.

Following on the finding of a previous study[26] that ADC value of ice water is approximately $1.1 \times 10^{-3}$ $mm^2/s$ if the temperature is maintained at 0 , Sorace et al.[36] reported that ADC values of an ice-water phantom was $1.22 \times 10^{-3}$ $mm^2/s$. They reported that the average difference in repeatability of ADC was less than 0.01% and reproducibility was 2.7% on average. Another previous study[13] reported that the day-to-day repeatability of the measurements was within 4.5%, intra-examination repeatability of the ice-water phantom was within 1%, and the overall SD was 2.3%. These studies can help set the variability limits for the DWI technique for quality evaluation and control. Belli et al.[3] reported quality assurance for quantitative DWI in a multicenter comparison of various vendors. In this multicenter study[3], using a standard doped water phantom, there was no significant difference in ADC

values between 1.5-T and 3-T scanners, and no differences among vendors. In a recent study performed in a clinical setting[2)], the reported test–retest repeatability and reproducibility and ADC measures in breast DWI in patients with breast cancer, showed excellent repeatability and reproducibility in a multi-institution setting, using a standardized protocol and QA procedure. Newitt et al. [2)] reported excellent ADC repeatability (wCV = 4.8%, ICC = 0.97 [95%CI 0.95, 0.98]). In previous clinical studies of test and retest repeatability in normal breast, wCV was 3.72%[36)] (n = 10), and 4.5% wCV for ADC derived from diffusion tensor imaging[37)]. Some clinical studies have also investigated repeatability and reproducibility of breast ADC measures in a single site in the normal breast[38, 39)] or in malignant breast lesions[40, 41)]. They reported that the wCV or reproducibility was 3.2–8.3%. In our study, in terms of the overall mean ADC value, there was no significant difference among all vendors ($p$ = 0.456). The overall mean ADC value was not influenced by PVP%, the vendors, room or bore temperature, or the test and retest effect. Furthermore, our overall repeatability values of all mean ADC measurements were excellent (wCV = 8.1%, ICC 0.98, RC = 0.23 × $10^{-3}$ $mm^2/s$, and AI = 0.84). Repeatability was not significantly influenced by the vendor, bore temperature, or

different *b*-value combinations. Our results acquired at multiple centers revealed similar differences in terms of repeatability and reproducibility in a previous phantom study[3,13,26,36]. Despite the slightly different protocol of different equipment, the results showed good agreement.

The overall SNR, on the other hand, showed a significant difference among vendors ($p < 0.001$). In particular the scanner of vendor A showed a statistically significantly lower SNR than that of the other vendors. For high *b*-values, the SNR significantly decreased. This result was meaningful because variable SNR can bias ADC measurements and spatial resolution. We also found that SNR differed significantly across *b*-value combinations. The value for set 2 was better than those of sets 1 and 3. This was probably due to the low SNR of the highest *b*-value ($b = 1200$).

Only a few previous studies had evaluated spatial resolution previously[13, 15, 16]. Malyarenko[13] evaluated the ADC-dependence of spatial resolution, and found a larger offset induced by a larger ADC error. To evaluate system-to-system reproducibility as a function of ROI location, the average of ADC measurements of the central tube ROI at two extreme

superior–inferior (S/I) offsets and the average of right–left (R/L) offsets were used, separately. They reported that measurement variability was significantly higher with a ± 70 mm offset from the center in the S/I direction for the head coil and ± 110 mm R/L offset for the torso coil, respectively. They concluded that ADC bias was consistent with the systematic measurement of slope nonlinearity. Keenan et al.[15] defined each hole based on pixel intensity, calculated the center of each defined hole, and measured center-to-center distances between all holes in the right/ left (R/L) and anterior/posterior (A/P) directions. They used rigid geometry components, more specifically, a regularly spaced grid in the center of the diffusion phantom side, and they compared these with CAD drawings of the phantom. They reported that the measured R/L center-to-center distance was 19 mm and the A/P center-to-center distance was 20 mm, while the actual spacing was 20 mm. Tuong and Gardiner[16] evaluated spatial resolution with a phantom that had a resolution plate of measurable shapes of various sizes on the STIR image. Their phantom consisted of multiple measurable lines, steps, and circles of varying sizes, ranging from 1 to 20 mm. They reported that, on STIR images, more than 2-mm-sized structures were visualized in 85% of instances. We also evaluated spatial resolution

with an axial set of visualized pattern holes of various sizes on images obtained with various *b*-values, with a phantom containing various pattern holes ranging in size from 1 to 10 mm. In axial sets, sizes measuring ≥ 2 mm were visualized in 77.8% of instances in all groups, and in 100% in groups 1 and 2 on $b = 0$ images. Pattern holes with diameters of ≥ 1 mm were seen in 33.3% of instances in group 3 and 55.6% of instances in groups 1 and 2 on $b = 0$ images. Thus, more than half of 1-mm-sized pattern holes were visible, depending on the location, on $b = 0$ images. Based on these results, size cutoffs would be 2 mm for pattern holes, and DWI could be used for screening with $b = 0$, as previously reported in several studies[6, 7]. However, pattern holes of the phantom were structures with clear margins and a homogeneous background; there may be limitations to using this cutoff in a clinical setting. Before discussing cutoff values for screening, further phantom or in vivo studies are needed to observe the effect of a heterogeneous background.

In addition, we evaluated spatial resolution with an analysis program, with two resolution insets of the DWI phantom, and analysis of each pattern hole individually, which has not been reported in previous studies. Assessment of the morphological features and size

of lesions is of great importance in the characterization of a suspicious findings on MRI; our study is therefore relevant from this point of view. Our results showed that spatial resolution in the posterior was better than in the anterior in both the axial set and sagittal set. This means that chest wall lesions could be distorted in a clinical setting. Moreover, spatial resolution in the central area was better than that in the periphery (R/L direction).

Interestingly, in analysis of the circularity of size 1 or 6 circles in one group in the axial set of spatial resolution pattern holes, the test and retest values were almost consistent, but the closer the pattern holes were to size 3, the greater was the difference between the test and retest. This result was probably due to technical issues and thus circularity assessments may not be valid for very small particles. The smallest circle on the ASAN J is represented by dots or squares. If the analysis program or pixel value is more detailed, it may have been consistent by the size difference.

For image quality evaluation, we also assessed artifacts. Among all acquired images, the Nyquist artifact was noted (Fig. 2A and B, asterisk) in images from one vendor. This may be due to parallel imaging undersampling, showed the duplicating signals from the subject on

DWI. It may also cause reduced SNR[4]. This artifact commonly occurred but was not noted in imaged from scanners from the other two vendors (vendors B, C). For that specific scanner, we performed ADC measurements several times with different sequences; however, this artifact was observed in all images, with all *b*-values. This artifact could be reduced by using parallel imaging acceleration, reducing echo train length, lowering phase encoding resolution, or avoiding multi-shot (segmented) EPI.

Our study had several limitations. First, as mentioned in a previous study[15] evaluating phantoms presents a technical problem. Due to the limitation of phantoms, they might not completely reflect the human breast tissues because of the intrinsic heterogeneity of breast parenchyma, which can affect the quality of the breast MRI. Several studies[9-11] have utilized phantoms made of various materials, but these could not fully represent the human body. However, the design of the phantom we used was presented to balance the anatomical features, and therefore to maximize functionality. Keenan et al[15] reported that the ADC values of PVP were nearly in agreement with in vivo ADC values using a diffusion phantom that was similar to that used in the present study. The results of our study also showed good agreement

with those previous studies. Most recently, research on phantom development using 3D printers for multi-purpose and multi-modality imaging has been reported[42]. These newly developed phantoms might be able to overcome this technical problem.

Second, the number of included scanners were relatively small (six institutions, nine scanners). Differences in values such as the SNR may be due to the heterogeneous distribution of a small number of scanners and vendors (one GE, six Philips, and two Siemens). Therefore, further studies with a larger number of institutions and scanners will be necessary.

In a previous study, Partridge et al.[4] reported that DWI holds potential to improve the detection and biological characterization of breast cancer. Therefore, standardization of the DWI protocol is the most important issue. Multicenter studies need to be performed to improve repeatability and reproducibility of ADC measurement. In this study, the repeatability and reproducibility of mean ADC value were evaluated using standard breast DWI phantom in a multi-center study, using standardized protocols. We found excellent agreement in the test and retest measures of the overall mean ADC value. The repeatability of ADC measurement was not influenced by different vendors or scanners or room temperature, even though the

SNR varied between vendors. Although the SNR varied according to the *b*-value or multi-*b*-value combinations, and between scanners, these results did not affect the measured ADC values. We also found excellent agreement of circularity of spatial resolution; all scanners showed differences in both the AP and RL direction. Our multicenter test using a standard DWI phantom and a standard acquisition protocol showed that ADC values and spatial resolution would not be affected by vendors, while SNR was affected by vendors and *b*-values.

**Table 1.** Scanners and DWI Acquisition Protocol

| Vendor | Siemens | Philips | GE |
|---|---|---|---|
| Equipment | 3T Skyra | 3T Ingenia (Ingenia CX) | 3T architect |
| Scan dimension | Bilateral axial/ sagittal | Bilateral axial/ sagittal | Bilateral axial/ sagittal |
| DWI technique | rs[a] –EPI[a] (RESOLVE[b]) | ss[a] –EPI (SENSE[c]) | ss –EPI (ASSET[d]) |
| Fat suppression | SPIR[e] | STIR[f] | SPAIR[g] |
| TR/TE (ms) | 9930/ 69 | 9161/ 70 (TI = 230) | 7500/ 76.6 (TI = 115) |
| Field of view (mm) | 340 × 207 | 340 × 212 | 340 × 204 |
| Flip angle (degree) | 180 | 90 | 90 |
| Number of average | 1 | 6 | 3 |
| Thickness (mm) | 3 | 3 | 3 |
| Intersection gap (%) | 0 | 0 | 0 |
| Matrix | 236 × 156 | Recon.[h] 230 × 200 Acq.[i] 256 × 160 | 256 × 152 |
| Voxel size (mm) | 1.3 × 1.3 × 1.3 | 1.1 × 1.1 × 3 (1.3 × 1.3 × 1.3) | 1.3 × 1.3 × 1.3 |
| Slices | 50 | 50 | 50 |
| *b*-value set 1,3 | 0, 800, 1200 | 0, 800, 1200 | 0, 800, 1200 |
| ($s/mm^2$) set 2 | 0, 1000 | 0, 1000 | 0, 1000 |

[a] Single-shot (ss) or readout-segmented (rs) echoplanar sequence (EPI)

[b] Readout segmentation of long variable echo trains

[c] Sensitivity encoding

[d] Array coil spatial sensitivity encoding

[e] Spectral presaturation with inversion recovery

[f] Short TI inversion recovery

[g] Spectral attenuated inversion recovery

[h] Reconstruction

[i] Acquisition

**Table 2.** A comparison of the results of a previous phantom study[15] and our acquired mean ADC values. The results of previous clinical 1.5-T systems in vivo values for malignant tumor to benign tissue, as measured in the literature[18, 19] ($\times 10^{-3}$mm$^2$/s); benign mass 1.72 ± 0.43 to 1.74 ± 0.46; benign lesion 1.61 ± 0.33; malignant lesion 1.41 ± 0.22; and malignant mass 1.25 ± 0.29 to 1.32 ± 0.23.

| | Measured 25-75 percentile ranges PVP% from previous study[15] ($\times 10^{-3}$ mm$^2$/s), 17-24℃ | Range of mean ADC values From our study ($\times 10^{-3}$ mm$^2$/s), 20-21℃ |
|---|---|---|
| 0% | 2.140 to 2.612 | 1.917 to 2.298 |
| 10% | 1.689 to 1.802 | 1.794 to 1.866 |
| 14% | 1.516 to 1.593 | 1.376 to 1.421 |
| 18% | 1.358 to 1.400 | 1.250 to 1.322 |
| 25% | 0.937 to 1.028 | 0.782 to 0.837 |
| 40% | 0.606 to 0.695 | 0.270 to 0.319 |

**Table 3**. Percentage (%) of visualized pattern holes by *b*-values and groups in axial spatial resolution set

| Groups and | | Sizes of pattern holes | | | | | |
|---|---|---|---|---|---|---|---|
| *b*-values (s/mm$^2$) | | 1mm | 2mm | 4mm | 6mm | 8mm | 10mm |
| Group 1 | 0 | 55.6 | 100 | 100 | 100 | 100 | 100 |
| | 800 | 0 | 42.9 | 100 | 100 | 100 | 100 |
| | 1000 | 0 | 66.7 | 100 | 100 | 100 | 100 |
| | 1200 | 0 | 42.9 | 100 | 100 | 100 | 100 |
| Group 2 | 0 | 55.6 | 100 | 100 | 100 | 100 | 100 |
| | 800 | 14.3 | 85.7 | 100 | 100 | 100 | 100 |
| | 1000 | 11.1 | 100 | 100 | 100 | 100 | 100 |
| | 1200 | 0 | 71.4 | 100 | 100 | 100 | 100 |
| Group 3 | 0 | 33.3 | 77.8 | 100 | 100 | 100 | 100 |
| | 800 | 0 | 42.9 | 100 | 100 | 100 | 100 |
| | 1000 | 11.1 | 66.7 | 100 | 100 | 100 | 100 |
| | 1200 | 0 | 0 | 85.7 | 85.7 | 100 | 100 |

**Table 4.** Summary of spatial resolution parameters according to three vendors

| P value (axial) | Circularity | Area | Aspect ratio | Roundness |
|---|---|---|---|---|
| Across all vendor | 0.985 | 0.713 | 0.930 | 0.853 |
| A vs. B | 0.895 | 0.491 | 0.784 | 0.743 |
| A vs. C | 0.988 | 0.490 | 0.961 | 0.609 |
| B vs. C | 0.911 | 0.999 | 0.754 | 0.856 |

| P value | Circularity | | Area | | Aspect ratio | | Roundness | |
|---|---|---|---|---|---|---|---|---|
| (sagittal) | A-A' | B-B' | A-A' | B-B' | A-A' | B-B' | A-A' | B-B' |
| Across all vendor | 0.405 | 0.247 | <0.001* | 0.003* | 0.039* | 0.036* | 0.044* | 0.009* |
| A vs. B | 0.293 | 0.161 | <0.001* | 0.003* | 0.419 | 0.918 | 0.613 | 0.983 |
| A vs. C | 0.799 | 0.626 | 0.571 | 0.297 | 0.029* | 0.078 | 0.049* | 0.047* |
| B vs. C | 0.291 | 0.230 | <0.001* | 0.009* | 0.028* | 0.018* | 0.026* | 0.004* |

* $p < 0.05$

**Figure 1.** The Standard Breast DWI Phantom (QalibreMD$^{TM}$).

A. coronal, and B. axial DWI image of variable concentration of PVP test tubes C. axial, and D. sagittal images of spatial resolution pattern holes. Among 14 groups of variable pattern holes in axial scan, three groups were evaluated. Among 42 pattern holes in sagittal scan, four sets of six circles were evaluated for effect of right-to-left direction and anteroposterior direction.

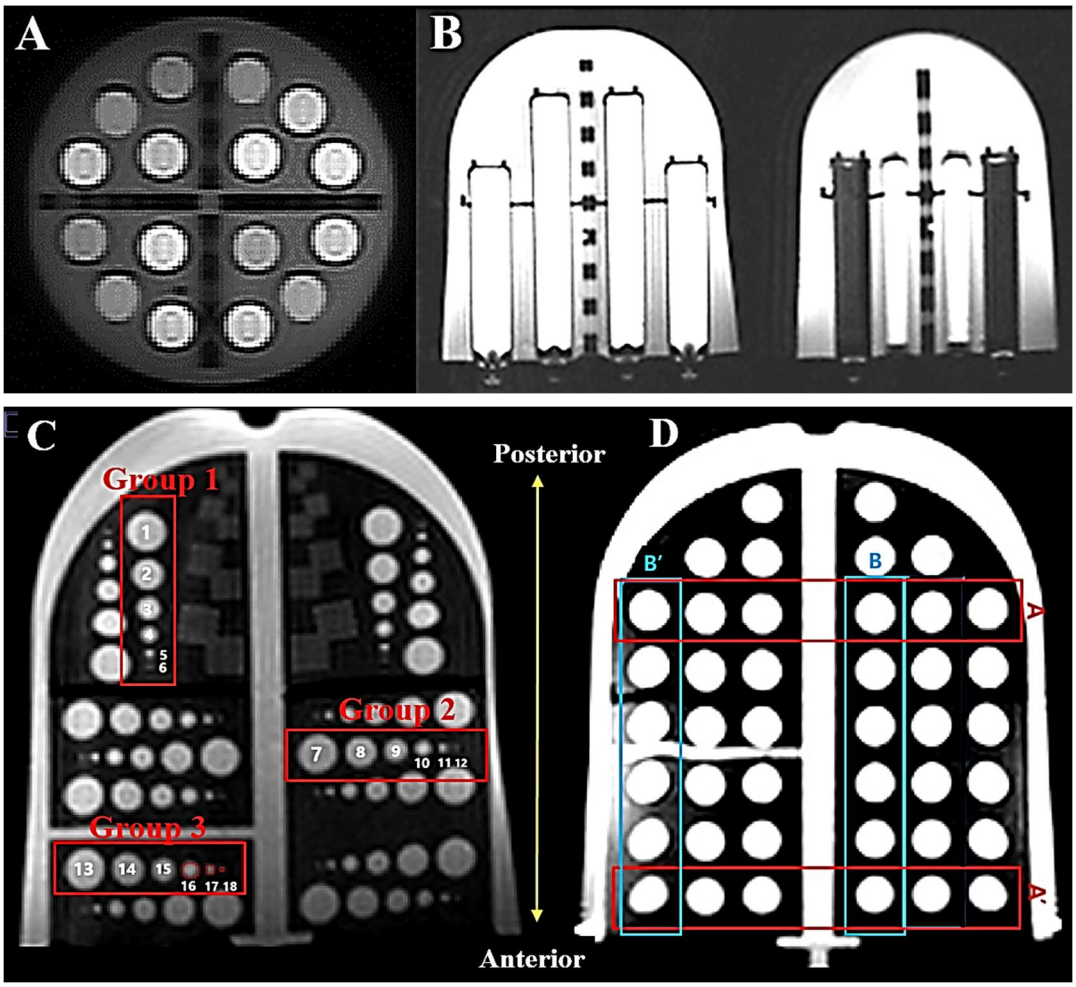


**Figure 2.** Method of ADC (A) and SNR (B) measurements on each representative slices using DWI phantom. Make five to ten ROI measured and derived mean ± SD (A) of ADC value. For SNR, an average of ROIs within each test tube devided by an average of ROIs in the background (B)**.** The vendors were A, B, and C from left to right**.** Nyquist artifact was noted on vendor A image (asterisk on A and B).

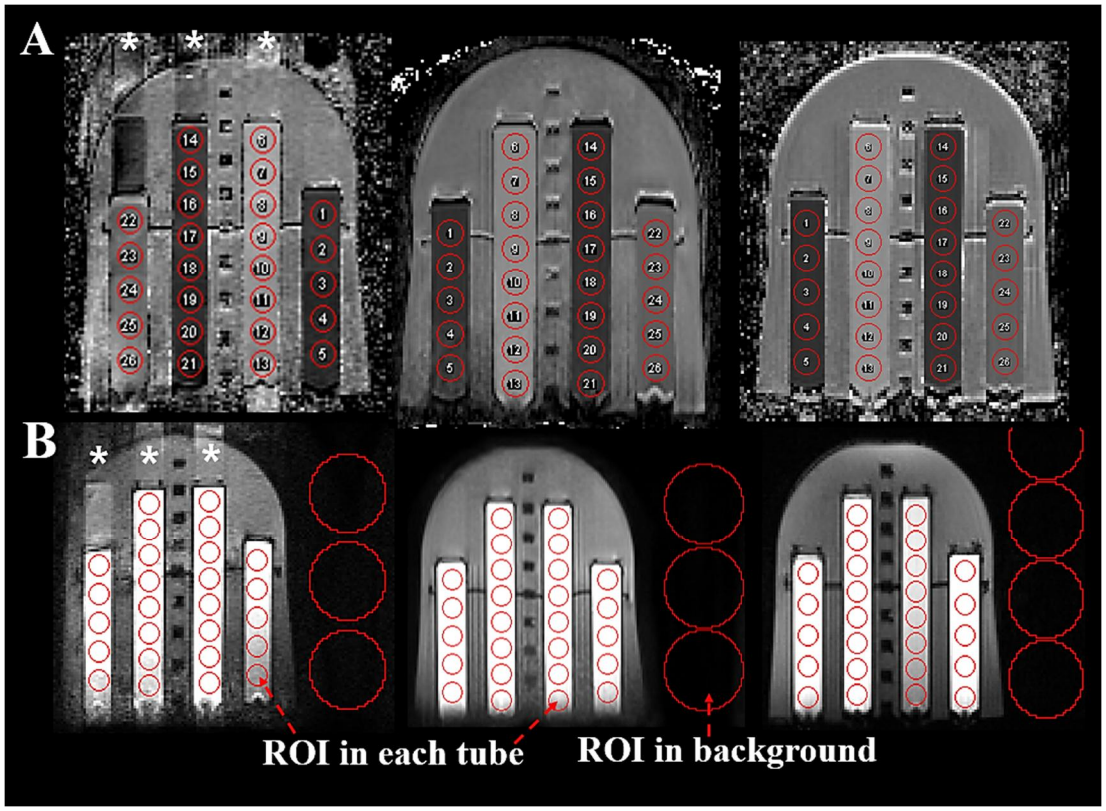


**Figure 3.** Boxplot displaying ADC values according to various PVP test tubes. The box extends from the 25th to the 75th percentile. The line is the median ADC value. The lines extend to the largest and smallest observed values within 1.5 box lengths.

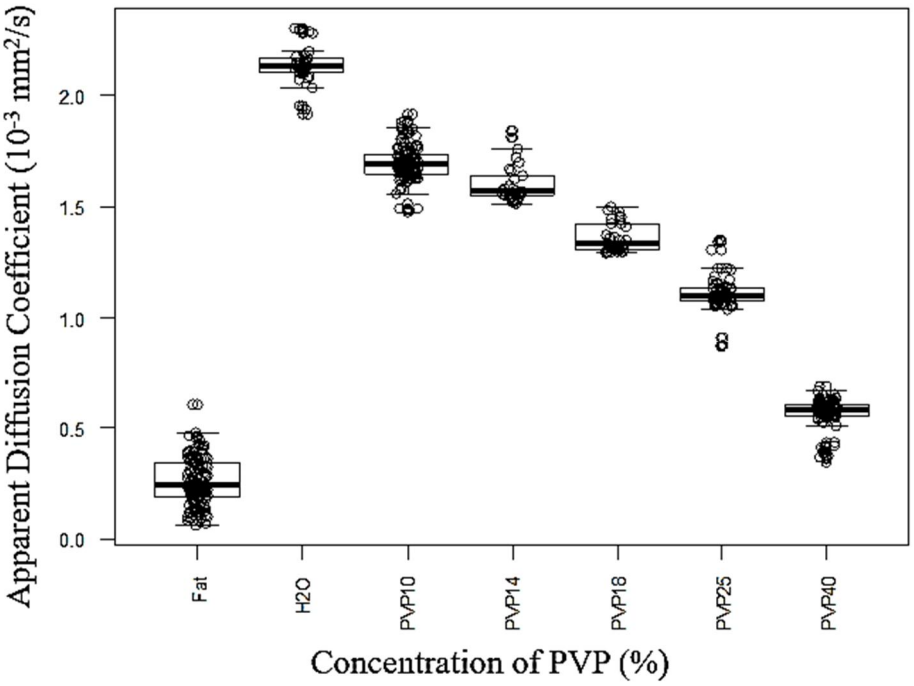


**Figure 4.** Boxplots displaying the distribution of mean ADC values (A) and SNR (B and C) according to vendors and *b*-values. While there was no statistically significant difference of mean ADC values ($p = 0.456$) across three vendors, SNR showed a significant difference between three vendors ($p < 0.001$). As *b*-values increased, SNR decreased ($p < 0.001$).

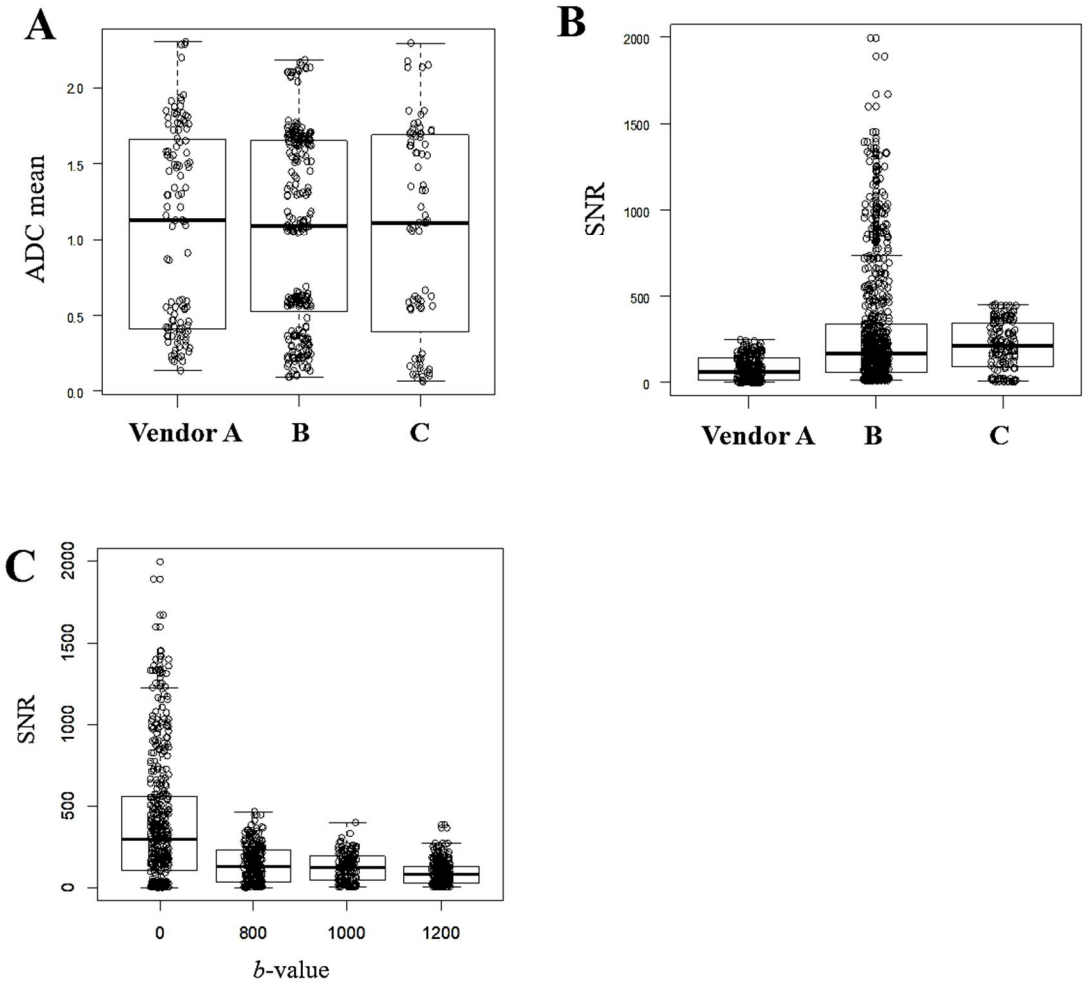


**Figure 5.** Bland–Altman plots for the difference of ADC measurements (A) and SNR (B) between test and retest for entire PVP test tubes. Mean difference and 95%CI (1.96 × SD) are shown as horizontal straight lines and dashed upper and lower lines, respectively. Repeatability of all mean ADC values of PVP% test tubes was excellent, while the distribution of net SNR values between test and retest was wider and SNR repeatability was poor because wCV exceeded 30%.

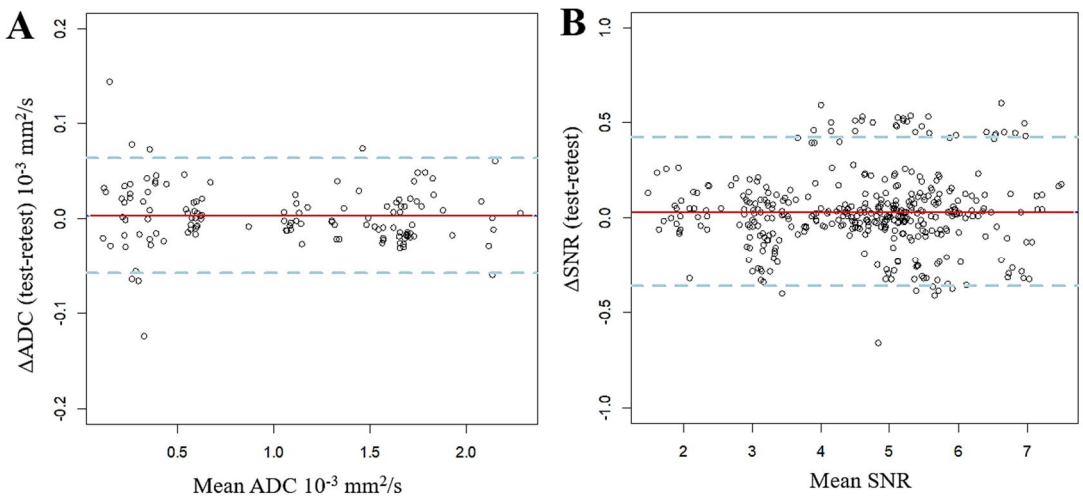
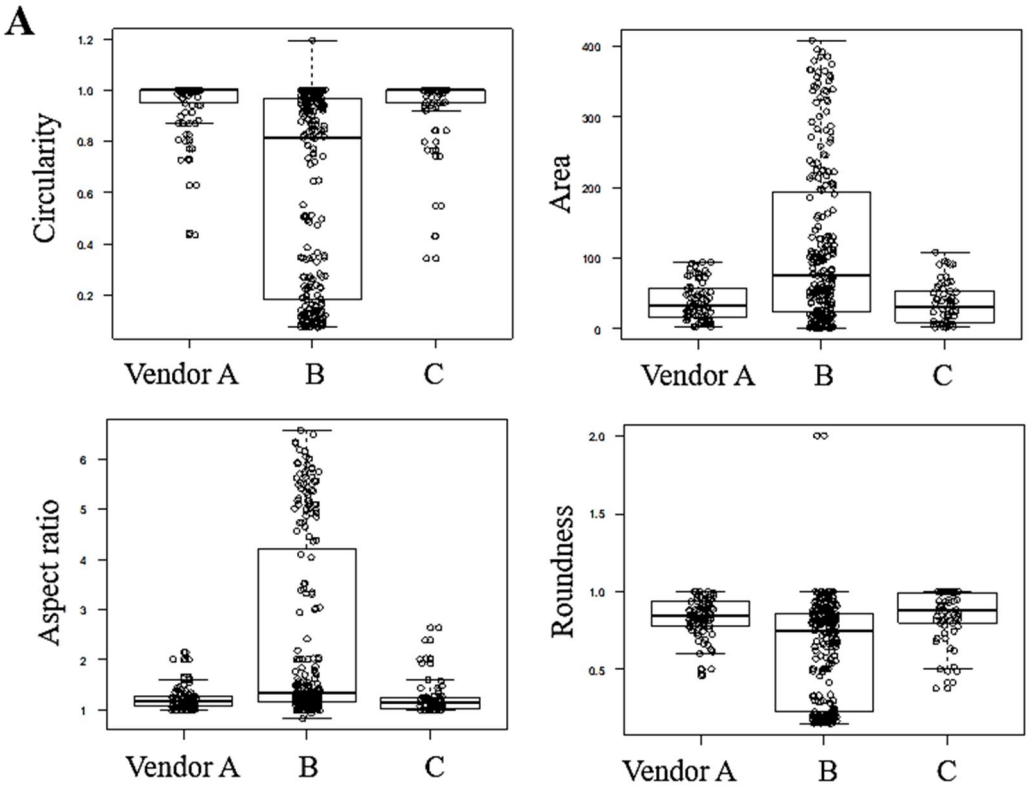
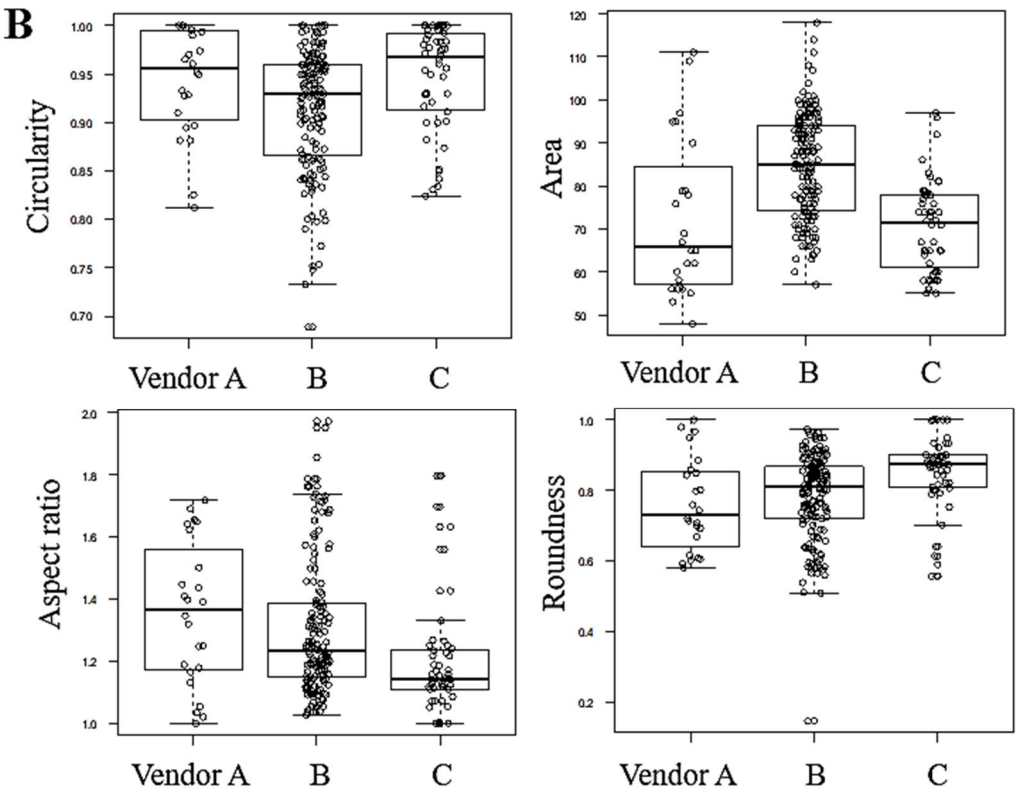


**Figure 6.** Summary of spatial resolution parameters according to three vendors for axial (A) and sagittal (B) sets. In axial set, there were no statistical significance between all vendors for all parameters of spatial resolution. However, in sagittal set, there was statistically difference across three vendors for spatial resolution parameters except circularity.




B
Circularity
1.00
0.95
0.90
0.85
0.80
0.75
0.70
Vendor A
B
C
Area
120
110
100
90
80
70
60
50
Vendor A
B
C
Aspect ratio
2.0
1.8
1.6
1.4
1.2
1.0
Vendor A
B
C
Roundness
1.0
0.8
0.6
0.4
0.2
Vendor A
B
C

## References


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

2. Newitt DC, Zhang Z, Gibbs JE, Partridge SC, Chenevert TL, Rosen MA, et al. Test–retest repeatability and reproducibility of ADC measures by breast DWI: Results from the ACRIN 6698 trial. Journal of Magnetic Resonance Imaging. 2018;49(6):1617-28.

3. Belli G, Busoni S, Ciccarone A, Coniglio A, Esposito M, Giannelli M, et al. Quality assurance multicenter comparison of different MR scanners for quantitative diffusion-weighted imaging. J Magn Reson Imaging. 2016;43(1):213-9.

4. Partridge SC, Nissan N, Rahbar H, Kitsch AE, Sigmund EE. Diffusion-weighted breast MRI: Clinical applications and emerging techniques. J Magn Reson Imaging. 2017;45(2):337-55.

5. Chenevert TL SL, Taylor JM, Robertson PL, Greenberg HS, Rehemtulla A, Ross BD. Diffusion Magnetic Resonance Imagingan Early Surrogate Marker of Therapeutic Efficacy in

Brain Tumors. J Natl Cancer Inst 2000;92(24):2029-36.

6. Amornsiripanitch N, Bickelhaupt S, Shin HJ, Dang M, Rahbar H, Pinker K, Partridge SC. Diffusion-weighted MRI for Unenhanced Breast Cancer Screening. Radiology.2019 Oct 8:182789. doi: 10.1148/radiol.2019182789. [Epub ahead of print].

7. Kang JW, Shin HJ, Shin KC, Chae EY, Choi WJ, Cha JH, et al. Unenhanced magnetic resonance screening using fused diffusion-weighted imaging and maximum-intensity projection in patients with a personal history of breast cancer: role of fused DWI for postoperative screening. Breast Cancer Res Treat. 2017;165(1):119-28.

8. ClinicalTrials.gov [Internet]. Moon WK (MD): Seoul National University Hospital (Korea). 2019 Feb 7- . Identifier NCT03835897, Breast Cancer Screening With Diffusion-weighted MRI in Women at High Risk for Breast Cancer; 2019 Feb 7 [cited 2019 Feb 11]; [about 4 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT03835897?term=03835897&draw=1&rank=1.

9. Pullens P, Bladt P, Sijbers J, Maas AI, Parizel PM. Technical Note: A safe, cheap, and easy-to-use isotropic diffusion MRI phantom for clinical and multicenter studies. Med Phys.

2017;44(3):1063-70.

10. Gatidis S, Schmidt H, Martirosian P, Schwenzer NF. Development of an MRI phantom for diffusion-weighted imaging with independent adjustment of apparent diffusion coefficient values and T2 relaxation times. Magn Reson Med. 2014;72(2):459-63.

11. Lavdas I, Behan KC, Papadaki A, McRobbie DW, Aboagye EO. A phantom for diffusion-weighted MRI (DW-MRI). J Magn Reson Imaging. 2013;38(1):173-9.

12. Jerome NP PM, Orton MR, Parkes HG, Winfield JM, Boss MA, Leach MO, deSouza NM, Collins DJ. Development of a temperature-controlled phantom for magnetic resonance quality assurance of diffusion, dynamic, and relaxometry measurements Med Phys 2016;43(6):2998-3007.

13. Malyarenko D, Galban 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.

14. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI

Quantitative Imaging Network. J Med Imaging (Bellingham). 2018;5(1):011003.

15. Keenan KE, Wilmes LJ, Aliu SO, Newitt DC, Jones EF, Boss MA, et al. Design of a breast phantom for quantitative MRI. J Magn Reson Imaging. 2016;44(3):610-9.

16. Tuong B, Gardiner I. Development of a novel breast MRI phantom for quality control. AJR Am J Roentgenol. 2013;201(3):W511-5.

17. Quantitative MRI Breast Phantom. QalibreMD, 14 Oct. 2019, http://www.qalibre-md.com/breastphantom/.

18. Partridge SC, Demartini WB, Kurland BF, Eby PR, White SW, Lehman CD. Differential diagnosis of mammographically and clinically occult breast lesions on diffusion-weighted MRI. J Magn Reson Imaging. 2010;31(3):562-70.

19. Partridge SC, Mullins CD, Kurland BF, Allain MD, DeMartini WB, Eby PR, et al. Apparent diffusion coefficient values for discriminating benign and malignant breast MRI lesions: effects of lesion type and size. AJR Am J Roentgenol. 2010;194(6):1664-73.

20. Quantitative Imaging Biomarkers Alliance (QIBA) Profile: Diffusion-Weighted Magnetic Resonance Imaging (DWI)

http://qibawiki.rsna.org/images/7/7e/QIBADWIProfile_as_of_2019-Feb-05.pdf.

21. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. Radiology. 2019;292(3):520-36.

22. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology. 1986;161(2):401-7.

23. Yoshida T, Urikura A, Shirata K, Nakaya Y, Endo M, Terashima S, et al. Short tau inversion recovery in breast diffusion-weighted imaging: signal-to-noise ratio and apparent diffusion coefficients using a breast phantom in comparison with spectral attenuated inversion recovery. Radiol Med. 2018;123(4):296-304.

24. Kim YJ, Kim SH, Kang BJ, Park CS, Kim HS, Son YH, et al. Readout-segmented echo-planar imaging in diffusion-weighted mr imaging in breast cancer: comparison with single-shot echo-planar imaging in image quality. Korean J Radiol. 2014;15(4):403-10.

25. Yoshida T, Urikura A, Shirata K, Nakaya Y, Terashima S, Hosokawa Y. Image quality assessment of single-shot turbo spin echo diffusion-weighted imaging with parallel imaging

technique: a phantom study. Br J Radiol. 2016;89(1065):20160512.

26. Chenevert TL, Galban CJ, Ivancevic MK, Rohrer SE, Londy FJ, Kwee TC, et al. Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. J Magn Reson Imaging. 2011;34(4):983-7.

27. National Institutes of Health, US. Department of Health and Human Services, 2019, https://imagej.nih.gov/ij/

28. Zhang Z, Wang Y, Duan F. An AUC-like index for agreement assessment. J Biopharm Stat. 2014;24(4):893-907.

29. Kong KA. Statistical Methods: Reliability Assessment and Method Comparison. The Ewha Medical Journal. 2017;40(1):9-16.

30. Sasaki M YK, Watanabe Y, Matsui M, Ida M, Fujiwara S, Shibata E. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology. 2008;249(2):624-30.

31. Zhu T, Hu R, Qiu X, Taylor M, Tso Y, Yiannoutsos C, et al. Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study.

Neuroimage. 2011;56(3):1398-411.

32. Delakis I ME, Leach MO, De Wilde JP. Developing a quality control protocol for diffusion imaging on a clinical MRI system. 2004;49(8):1409-22.

33. Lavdas I, Miquel ME, McRobbie DW, Aboagye EO. Comparison between diffusion-weighted MRI (DW-MRI) at 1.5 and 3 tesla: a phantom study. J Magn Reson Imaging. 2014;40(3):682-90.

34. Miquel ME, Scott AD, Macdougall ND, Boubertakh R, Bharwani N, Rockall AG. In vitro and in vivo repeatability of abdominal diffusion-weighted MRI. Br J Radiol. 2012;85(1019):1507-12.

35. Wagner F, Laun FB, Kuder TA, Mlynarska A, Maier F, Faust J, et al. Temperature and concentration calibration of aqueous polyvinylpyrrolidone (PVP) solutions for isotropic diffusion MRI phantoms. PLoS One. 2017;12(6):e0179276.

36. Sorace AG, Wu C, Barnes SL, Jarrett AM, Avery S, Patt D, et al. Repeatability, reproducibility, and accuracy of quantitative mri of the breast in the community radiology setting. J Magn Reson Imaging. 2018. Mar 23. doi: 10.1002/jmri.26011. [Epub ahead of print]

37. Partridge SC, Murthy RS, Ziadloo A, White SW, Allison KH, Lehman CD. Diffusion tensor magnetic resonance imaging of the normal breast. Magn Reson Imaging. 2010;28(3):320-8.

38. Aliu SO, Jones EF, Azziz A, Kornak J, Wilmes LJ, Newitt DC, et al. Repeatability of quantitative MRI measurements in normal breast tissue. Transl Oncol. 2014;7(1):130-7.

39. O'Flynn EA, Morgan VA, Giles SL, deSouza NM. Diffusion weighted imaging of the normal breast: reproducibility of apparent diffusion coefficient measurements and variation with menstrual cycle and menopausal status. Eur Radiol. 2012;22(7):1512-8.

40. Giannotti E, Waugh S, Priba L, Davis Z, Crowe E, Vinnicombe S. Assessment and quantification of sources of variability in breast apparent diffusion coefficient (ADC) measurements at diffusion weighted imaging. Eur J Radiol. 2015;84(9):1729-36.

41. Jang M, Kim SM, Yun BL, Ahn HS, Kim SY, Kang E, et al. Reproducibility of Apparent Diffusion Coefficient Measurements in Malignant Breast Masses. J Korean Med Sci. 2015;30(11):1689-97.

42. He Y, Liu Y, Dyer BA, Boone JM, Liu S, Chen T, et al. 3D-printed breast phantom

for multi-purpose and multi-modality imaging. Quantitative Imaging in Medicine and Surgery. 2019;9(1):63-74.

**Acknowledgement**

SY Park, PhD, department of Clinical Epidemiology and Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, for statistical consult.

## 국문요약

**목적:** 본 연구의 목적은 정량적 확산 강조 영상 (DWI) 의 표준화를 위한 DWI 영상 품질 평가를 위해, 표준 유방 확산 팬텀을 사용하여 국내 여섯 개 기관의 총 아홉 대의 3 테슬라 (T) 자기공명 영상 (MRI) 장비에서 얻은 DWI 의 겉보기 확산 계수 (ADC) 측정, 공간 분해능 및 신호 대 잡음비 (SNR)를 비교하는 것이다.

**연구대상 및 방법:** 총 6 개 기관에서 9 개 MRI 장비로 표준 DWI 팬텀 (Qalibre MD$^{TM}$)을 사용하였다. DWI 스캔 프로토콜은 여러 스캐너 플랫폼의 제약 조건을 고려하여 가능한 한 표준화하였다. 세 개의 *b*-values ($b$ = 0, 800 및 1200 s/mm$^2$)로 단일 샷 (ss) 또는 다중 샷 (rs)을 사용하여 영상을 얻고, 재촬영은 장비 밖으로 스캔 테이블을 뺐다가 다시 넣는 방식으로 시행하였다. 또한, 2 개의 *b*-values ($b$ = 0 및 1000 s/mm$^2$) 를 DWI 의 시험 세트 직전에 수행하여 상이한 *b*-value 조합의 효과를 비교하였다. 인공물 및 SNR 또한 비교하였으며, 공간 해상도의 평가에 관해서는, 팬텀내 구조물의 원형의 구현성 및 면적을 구하여 분석하였다. 반복성 및 재현성은 Bland-Altman 플롯, 개체 내 변동 계수 (wCV) 및 클래스 내 상관 계수 (ICC), 일치 지수 (AI) 및 반복성 계수 (RC)를 사용하여 평가하였다.

**결과:** 측정된 모든 ADC 값은 ADC 기준 측정 범위 내에 있었고, 이는 장비 간에 통계적 차이를 보이지 않았으며 ($p$ = 0.456), 서로 다른 $b$-value 간에는 통계적으로 유의한 차이를 보였다. 보어 온도의 범위는 20 (n = 5) 에서 21 (n = 4)이며 ADC 값은 이에 영향을 받지 않았다 ($p$ = 0.262). SNR 은 각 병원과 장비 간에 통계적 차이 ($p$ < 0.001)를 보였고 높은 $b$-value 영상일수록 낮은 SNR 을 보였다. SNR 또한 온도에 따른 차이는 없었다 ($p$ = 0.802). 모든 ADC 값 재현성은 우수하였으며, ICC = 0.98 (95 %CI 0.97, 0.99), wCV 는 8.1 % 이었다. 그러나 SNR 의 재현성은 ICC = 0.23 (95 %CI 0.13, 0.42) wCV 가 113.2 % 였다. 공간 분해능은 전후 방향 ($p$ < 0.001)과 좌우 방향 ($p$ < 0.05)에 따라 유의 한 차이가 있었다.

**결론:** 확산 강조 영상이 임상에서 이용되는 지표로 사용되기 위해서는 정량적 방법이 충분히 신뢰할 수 있고 장비 간에 차이가 없어야 한다. 이번 다기관 국내 연구에서는 표준 유방 확산강조영상 팬텀을 이용, 표준화된 정량적 확산강조영상을 적용한 연구를 통하여 국내 병원간, 다양한 자기공명 영상 기기간의 ADC 측정에 있어서 반복성과 재현성이 뛰어남을 확인하였다. SNR 은 $b$-value 와 장비 마다 다르지만 이는 ADC 값 측정에 영향을 미치지 않았다. 공간 분해능은 모든 장

비에서 동일하게 전후 방향과 좌우 방향의 차이를 보였다.

**중심단어**

표준 유방 확산 팬텀

확산 강조 영상

품질 평가

다기관 연구