Variability and Bias Assessment in Breast ADC Measurement Across Multiple Systems

Kathryn E. Keenan, PhD,[1]* Adele P. Peskin, PhD,[1] Lisa J. Wilmes, PhD,[2] Sheye O. Aliu, PhD,[2] Ella F. Jones, PhD,[2] Wen Li, PhD,[2] John Kornak, PhD,[3] David C. Newitt, PhD,[2] and Nola M. Hylton, PhD[2]

Purpose: To assess the ability of a recent, anatomically designed breast phantom incorporating T_1 and diffusion elements to serve as a quality control device for quantitative comparison of apparent diffusion coefficient (ADC) measurements calculated from diffusion-weighted MRI (DWI) within and across MRI systems.
Materials and Methods: A bilateral breast phantom incorporating multiple T_1 and diffusion tissue mimics and a geometric distortion array was imaged with DWI on 1.5 Tesla (T) and 3.0T scanners from two different manufacturers, using three different breast coils (three configurations total). Multiple measurements were acquired to assess the bias and variability of different diffusion weighted single-shot echo-planar imaging sequences on the scanner-coil systems.
Results: The repeatability of ADC measurements was mixed: the standard deviation relative to baseline across scanner-coil-sequences ranged from low variability (0.47, 95% confidence interval [CI]: 0.22–1.00) to high variability (1.69, 95% CI: 0.17–17.26), depending on material, with the lowest and highest variability from the same scanner-coil-sequence. Assessment of image distortion showed that right/left measurements of the geometric distortion array were 1 to 16% larger on the left coil side compared with the right coil side independent of scanner-coil systems, diffusion weighting, and phase-encoding direction.
Conclusion: This breast phantom can be used to measure scanner-coil-sequence bias and variability for DWI. When establishing a multisystem study, this breast phantom may be used to minimize protocol differences (e.g., due to available sequences or shimming technique), to correct for bias that cannot be minimized, and to weigh results from each system depending on respective variability.

J. MAGN. RESON. IMAGING 2016;00:000–000.

Quantitative MRI is increasingly used for breast cancer diagnosis, staging, and treatment monitoring. Several clinical trials that use MRI techniques to assess neoadjuvant treatment protocols are under way.[1,2]

Diffusion MRI is used increasingly to provide characterization of breast cancer tumors,[3,4] either in addition to or as a replacement for dynamic contrast enhanced MRI with T_1 measurements and T_1-weighted images.[5,6] Quality control is much needed to ensure accurate quantitation of diffusion-weighted MRI (DWI) measurements for clinical assessments,[4] the importance of which is further substantiated by the bias in diffusion measurements recently observed in a multicenter trial.[7] Similarly, Min et al summarized several studies that determine a threshold apparent diffusion coefficient (ADC) value between benign and malignant breast tumors[8]: the reported thresholds ranged from 0.85×10^{-3} m^2/s to 1.48×10^{-3} m^2/s, which is too broad to be used clinically. To determine a threshold ADC value, an improved study would distinguish the variance in measurements due to the scanner, coil, and sequence from differences between patients' benign and malignant tumor tissue.

A breast phantom is a quality control device that can be used to quantify measurement variability due to system hardware, software, or imaging sequence. A breast phantom

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.25237

Received Jun 19, 2015, Accepted for publication Feb 29, 2016.

*Address reprint requests to: K.E.K., 325 Broadway, Boulder, CO 80305. E-mail: kathryn.keenan@nist.gov

From the [1]Physical Measurement Laboratory, National Institute of Standards and Technology, Boulder, Colorado, USA; [2]Department of Radiology and Biomedical Imaging, University of California, San Francisco, California, USA; and [3]Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA

Additional supporting information may be found in the online version of this article.

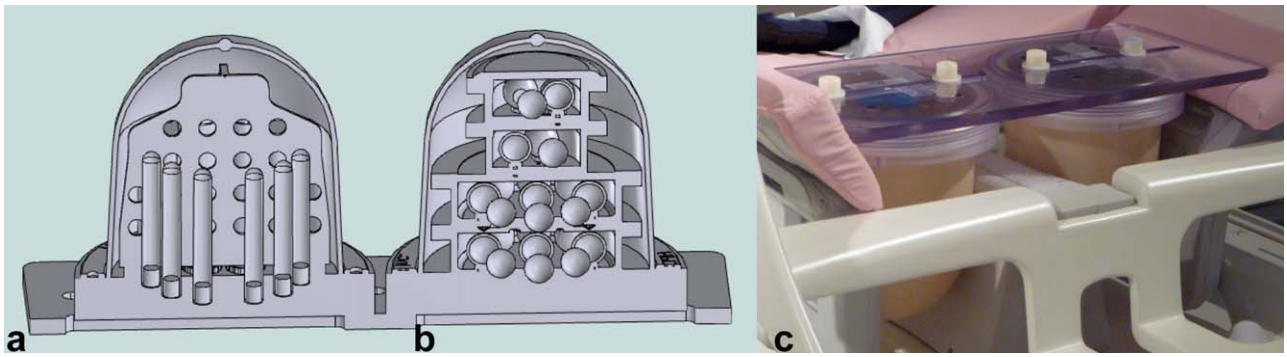

FIGURE 1: Breast phantom CAD model showing the diffusion phantom unit with vertically oriented samples tubes and geometric distortion plate (a) and the T_1 phantom unit with sample spheres arranged on four isolated levels (b), and prototype prepared for imaging in a breast MR coil (c).

was introduced that is compatible with multiple coil styles, contains diffusion mimics for fat and healthy fibroglandular tissue, benign and malignant tumor tissue, and has an anatomically appropriate design,[9] distinguishing it from previous breast phantoms that were compatible with only one coil,[10] lacked diffusion mimics,[11] or lacked fat tissue mimics.[12] The goal of this study is to assess the ability of this breast phantom to serve as a quality control device for comparison of ADC data across scanner-coil-sequence combinations and repeated scanning (e.g., as would be required in a multisystem study or clinical trial).

Materials and Methods

Phantom

The breast phantom evaluated in this study was created as two distinct interchangeable units: one for diffusion and distortion evaluation and one for T_1 relaxation evaluation[9] (Fig. 1). The interchangeable nature allows each phantom unit to be imaged on each side of the coil, thereby allowing examination of differences between the left and right sides of the coil. As described in the phantom design paper,[9] the polyvinylpyrrolidone (PVP) solutions[13,14] in the breast phantom span the range of ADC values for malignant mass to benign lesion reported in the literature.[3,15,16]

The diffusion phantom unit includes a central, axial plate for measuring distortion. The phantom was oriented with the plate in the axial imaging plane. The plate contains a grid of 10 mm diameter holes on 20 mm center-to-center spacing in both A/P and R/L directions, with four A/P rows of four or six holes.

MRI Techniques

Diffusion weighted imaging sets and T_1-weighted (T_1-w) images were collected on three clinical scanner-coil configurations (Table 1). Apparent diffusion coefficient (ADC) measurements were performed using single-shot echo-planar imaging (SS-EPI) as described in Table 2. T_1-w images at 1.5 Tesla (T) used a three-dimensional (3D) fast gradient echo sequence with fat suppression, repetition time/echo time (TR/TE) 7.0/4.2 ms, flip angle 10 degrees, field of view (FOV) 400 mm^2, resolution 0.78 mm^2, and slice thickness 2 mm. T_1-w images at 3.0T used a 3D spoiled gradient echo with fat suppression, TR/TE 7.1/2.2 ms, flip angle 20 degrees, FOV 340 mm^2, resolution 0.76 mm^2, and slice thickness 2 mm. All sequences were acquired in the axial plane.

The diffusion and T_1 relaxation phantom units were swapped between the left and right imaging positions, and the acquisition was repeated to gather data from each side of the breast coil. In some cases, measurements were repeated on the same day and on subsequent days over several months to determine variation in system performance. In addition, components with the same material were in multiple locations in the diffusion phantom unit, allowing evaluation of measurement variation with position from the magnet isocenter.

TABLE 1. Scanner-Coil Configurations

Configuration	MRI scanner manufacturer	Breast coil	Field strength
1.5Ta	Signa, GE Healthcare, Milwaukee, WI	Open, eight-channel coil, Hologic (formerly Sentinelle Medical), Toronto, Ontario	1.5T
1.5Tb	Signa, GE Healthcare, Milwaukee, WI	Open, eight-channel coil, GE Healthcare, Milwaukee, WI	1.5T
3Tc	Verio, Siemens AG Medical Solutions, Erlangen, Germany	Open, 16-channel coil, Hologic (formerly Sentinelle Medical), Toronto, Ontario	3.0T

TABLE 2. ADC Measurement Parameters

Field	1.5T		3.0T
Sequence	2 b-value EPI	4 b-value EPI	4 b-value EPI
Configuration(s)	1.5Ta and 1.5Tb	1.5Ta	3Tc
TR (ms)	7500	7500	14100
TE (ms)	66.4	70.4	88
Flip angle (degrees)	90	90	90
Bandwidth (Hz)	1953.1	1953.1	1935.0
b-value (s/mm^2)	0, 600	0, 100, 600, 800	0, 100, 600, 800
Freq. encode direction	A/P	A/P	A/P
Averages	2, 8	2, 2, 4, 8	8
Slice thickness (mm)	2.0	2.0	2.0
FOV (mm x mm)	400 x 400	400 x 400	399 x 399
Matrix size (pixels)	256 x 192	256 x 160	152 x 152
Imaging plane	Axial	Axial	Axial

Bore temperatures were measured continuously at the time of scanning using an MRI-compatible fiber-optic temperature probe (OTP-M, OpSens, Québec, Canada). Temperature information was used in the interpretation of results; temperature corrections were not applied.

ADC Calculation

ADC measurements required calculation of parametric maps from DWI data and selection of the regions of interest (ROIs). We used software developed in the IDL programming environment (Exelis Visual Information Solutions, Boulder, CO) at the University of California San Francisco.[16] ADC maps were generated by fitting the equation:

$$\ln S_b = \ln S_0 - b\ ADC \quad (1)$$

using a linear least-squares approach, where S_b and S_0 are the diffusion-weighted and non–diffusion-weighted signals, respectively, and b is the diffusion sensitizing factor.

ROIs were created manually. ROIs for the diffusion phantom unit were delineated by rectangles measuring approximately 70 mm by 3 mm on three contiguous axial slices for the four large tubes and 55 mm by 2 mm on two contiguous axial slices for the 12 small tubes. These 2D ROIs were subsequently stacked into 3D multislice ROIs. In some datasets, the geometry of the tubes was distorted due to EPI artifacts. For those datasets, ROIs were rotated in the axial plane up to 5 degrees to ensure that the ROIs were circumscribed within the tubes.

Image Segmentation and Distortion Measurement

Segmentation was performed on the diffusion and T_1-w sets of breast phantom images to assess differences in distortion as a function of b-value, direction of the frequency-encoding gradient (right/left, R/L; or anterior/posterior, A/P) and coil position of the two units (diffusion phantom unit on coil left or coil right). The image segmentation and distortion measurement processes were fully automated and checked against manual measurements.

For each set of images, we selected the center, axial slice of the actual imaging volume, which always contained the grid of holes in the axial distortion plate in the diffusion phantom unit. To locate the rows of holes within the distortion plate, we first located the R/L boundaries of the distortion plate within this unit, and placed the rows of grid points relative to those plate boundaries. The outer boundary of the distortion plate was located by taking a threshold, defining the background to be the lowest 5% of intensity values. By locating the top and bottom A/P position of the boundary, we estimated the R/L vertical locations of the three structural plates in the diffusion phantom that are used for tube alignment from the computer aided design (CAD) model (Fig. 2A). In an A/P area of ± 10 pixels from each vertical estimate, we located pixels whose intensities are well below the neighboring

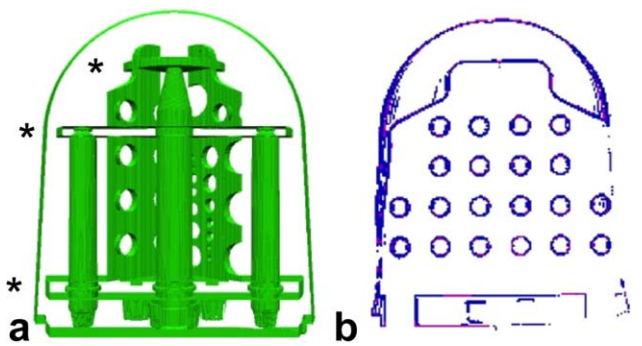

FIGURE 2: a: View of the components in the CAD model on the diffusion phantom unit with the three structural plates used for tube alignment noted with (*). b: Segmentation of holes in the grid using the center slice of the image set. This shows the high gradient/intensity pixels before each cluster is filled.

pixel above or below by more than 5% of the neighboring pixel intensity range. These were clustered and filled to establish the structural plate locations. From the structural plate locations, each of the four rows of grid holes is easily located.

Between the top and bottom large structural plates and inside the diffusion phantom unit outer boundary, we collected and averaged pixel intensity gradients, using a Sobel gradient operator. Sobel gradient magnitudes corresponding with individual pixels in the image are calculated separately in the horizontal (G_x) and vertical (G_y) directions, applying a 3×3 kernel:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \; G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

The Sobel gradient magnitude is then the square root of the sum of the squares of G_x and G_y. For pixels whose Sobel gradient magnitudes were at least 0.1 standard deviations above the slice average, we collected those with at least an intensity difference of more than 5% of the image's pixel intensity range between the pixel and an outside neighboring pixel. These clusters were located and filled, defining the holes. Figure 2B shows an example of these pixel clusters before they are filled. We measured the center of each hole and all center-to-center distances between holes in the R/L and A/P directions. In each set of images, measurements from the diffusion images in both coil positions were compared with those from T_1-w images, in which no distortion occurred, based on comparison with the design specifications.

Statistics

All statistical analysis was performed in R with the nlme package (cran.r-project.org). We used the generalized least squares (GLS) modeling using the nlme package function gls. GLS models are identical to linear models/multiple regression, except that they allow for nonconstant (heterogeneous) variance structures across subjects. Specifically, we allow for different variances across scanner-coil-sequence combinations. This extension to linear modeling is critical to our analysis, because estimation of differences in variability between scanner-coil-sequence combinations is central to our repeatability (test–retest) analyses. In this study design, we efficiently obtain estimates under varying sets of parameters simultaneously, using a modeling approach applied to all of the different combinations, rather than generating an individual dataset for each and every parameter to be estimated. We, therefore, considered each scanner-coil-sequence combination as a separate condition in the model.

The outcomes in all models were the median MR measurement over voxels in a 3D ROI (i.e., for a particular tissue mimic within a tube), and the predictors are the scanner-coil-sequence combination, distance of the ROI from the magnet isocenter (modeled as linear change from isocenter), and left vs. right coil position. All results are given in terms of estimates, 95% confidence intervals (CIs), and *P*-values without adjustment for multiple testing.[17–19] We consider $P < 0.05$ to constitute statistical significance, and we indicate where caution is warranted despite findings with $P < 0.05$.

We examined the relative standard deviation (SD) in measurements for the four different scanner-coil-sequence combinations; a lower relative standard deviation (lower variability) corresponds to a higher repeatability. Rather than providing absolute estimates and CIs of *standard deviations* to estimate *repeatability*, we provide the standard deviations in comparison to a particular baseline scanner-coil-sequence combination. The baseline scanner-coil-sequence combination for ADC was chosen to be that with measured ADC values closest to the reference measurements acquired using a spin-echo sequence with diffusion-weighted gradients,[9] Configuration "1.5Ta – 2 b-value" (Tables 1 and 2). A relative SD less than 1 is more repeatable than the baseline, and above 1 is less repeatable than the baseline.

To assess reproducibility, we compared the difference in estimated means with the baseline configuration mean ADC. GLS model term mean estimates (e.g., for scanner-coil-sequence effects or distance from the magnet isocenter) and CIs relate to lack of *reproducibility*. For example, for scanner-coil-sequence effects or distance from the magnet isocenter, the mean ADC estimate increases or decreases due to scanner-coil-sequence, or linear change in mean ADC due to distance from magnet isocenter. If the estimated mean ADC is significantly different from the baseline mean ADC, there is bias, and the measurement is not reproducible. In addition, we determined the estimated mean ADC as a percent of the baseline mean ADC to assess the clinical impact of the change, e.g., a different scanner-coil-sequence or a change in phantom position.

Results

Temperature

The bore temperatures ranged from 17.53 to 24.10°C across all configurations, and the SD of temperature for each single MRI session ranged from 0.16 to 0.78°C.

ADC Measurement Variability

This breast phantom study found no clear trends for ADC repeatability (Table 3). For both 1.5T scanner-coil-sequence combinations, the relative SD was both lower and higher than the baseline SD, depending on the mimic material and scanner-coil-sequence configuration. The relative SD to baseline across scanner-coil-sequences ranged from low variability (25% PVP: 0.47, 95% CI: 0.22–1.00) to high variability (0% PVP: 1.69, 95% CI: 0.17–17.26), both on 1.5Tb – 2 b-value EPI. The relative SD is higher, demonstrating lower repeatability when compared with baseline for all materials on 3.0Tc – 4 b-value EPI.

Scanner-Coil-Sequence ADC Measurement Bias

In most cases, ADC measures were reproducible; there was little measurement bias. ADC reproducibility for the scanner-coil-sequence model was assessed relative to the baseline, 1.5Ta – 2 b-value EPI, as shown in Table 4. The mean measured ADC for the baseline case, all materials, are given in the second column of Table 4. Only 3 of the 18 scanner-coil-sequence and material combinations were statistically different than baseline and, therefore, not

TABLE 3. Relative SD (Repeatability) on ADC Measurements across Clinical MRI Systems and Sequences

Material	Baseline residual SD (95% CI low, high) (10^{-6} mm^2/s) 1.5Ta, 2 b-value EPI	Relative SD on ADC (95% CI low, high) 1.5Tb, 2 b-value EPI	Relative SD on ADC (95% CI low, high) 1.5Ta, 4 b-value EPI	Relative SD on ADC (95% CI low, high) 3Tc, 4 b-value EPI
Water (PVP 0%)	156.0 (94.1, 258.5)	1.69 (0.17, 17.26)	0.78 (0.33, 1.81)	1.11 (0.69, 1.79)
10% PVP w/w in water	93.8 (74.2, 118.6)	0.63 (0.45, 0.89)	0.83 (0.66, 1.06)	1.45 (1.16, 1.81)
14% PVP	73.8 (42.3, 128.9)	1.41 (0.54, 3.69)	1.06 (0.59, 1.88)	1.08 (0.58, 2.02)
18% PVP	47.3[a]	0.97[a]	1.30[a]	1.73[a]
25% PVP	72.2 (51.9, 100.4)	0.47 (0.22, 1.00)	0.90 (0.63, 1.28)	1.29 (0.94, 1.77)
40% PVP	74.4 (56.5, 97.8)	0.71 (0.50, 1.00)	1.16 (0.87, 1.54)	1.04 (0.76, 1.42)

[a]Model did not converge on confidence interval results.

reproducible: deinonized water, 1.5Ta – 4 b-value EPI (198.8 × 10^{-6} mm^2/s [95% CI: 36.3–361.3], $P = 0.0203$); 10% w/w PVP, 1.5Tb – 2 b-value EPI (75.3 × 10^{-6} mm^2/s [95% CI: 34.0–116.5], $P = 0.0005$); and 10% w/w PVP, 1.5Ta – 4 b-value EPI (56.1 × 10^{-6} mm^2/s [95% CI: 8.5–103.8], $P = 0.0216$).

Dependence on Distance From the Magnet Isocenter

We included distance from the magnet isocenter as a predictor in the GLS model (Table 5). We found a spatial dependence on position from magnet isocenter, as expected from the nonlinear model previously reported,[7,20] and all ADC measurements reported in this study were adjusted by the model to remove the dependence on position.

Phantom Position Within the Coil

We included position of the phantom in the left or right side of the bilateral breast coil as a predictor in the GLS model (Table 6). In the majority of cases, the resulting change in the estimates of the mean ADC was not statistically significant. There was one statistically significant result for phantom position: 40% w/w PVP for the ADC scanner-coil-sequence model ($P = 0.0352$) with an estimated difference in mean ADC measured value of 8.96% (95% CI: 0.64–17.27%).

SS-EPI Distortion Assessment

While the majority of the differences in estimate of the means for ADC measurement with the phantom positioned in the left or right side of the coil were not significant, we did find a difference in the image distortion between the coil left and coil right sides of the SS-EPI axial diffusion images across all three scanner-coil configurations, independent of phase-encoding gradient direction (Fig. 3).

Distortion was assessed on both coil sides for T_1-w and SS-EPI diffusion sequences, comparing grid hole separation measurements in the R/L and A/P directions on the geometric distortion plate to known measurements from the phantom design CAD model. No measureable distortion effects were evident in either direction in the T_1-w images, nor in the A/P direction in the SS-EPI diffusion images. However, we did find a difference in geometric distortion in the R/L direction between the coil left and coil right sides in the SS-EPI axial diffusion images across all three scanner-coil configurations (Fig. 3). The image distortion was clearly visible as seen in Figure 4, illustrating the difference in relative apparent size of the two phantom units, and in Figure 5 showing the geometric distortion plate imaged with different sequences. Measured grid hole spacing was consistently larger on the coil left than the coil right across configurations, with mean differences across the four rows of 1.8 mm (9%), 1.5 mm (7.5%), and 0.7 mm (3.5%) for configurations 1.5Ta, 1.5Tb, and 3Tc, respectively. The direction of the phase-encoding gradient in the image acquisition did cause the expected SS-EPI image distortions, but those distortions did not change the R/L grid hole spacing distortion. Distortion measurements were the same for all b-value images (b = 0, 100, 600, and 800 s/mm^2), ruling out effects of diffusion gradients on the observed grid hole spacing distortion. To verify that this was not an effect of the phantom itself, we retrospectively examined data from human subjects, which agreed with the phantom findings (details are provided in the Supplementary Information, which is available online).

Discussion

Using the breast phantom, we completed an initial assessment of ADC measurement bias and variability on clinical

TABLE 4. Effect of Scanner-Coil-Sequence Configuration on ADC for Scanner-Coil-Sequence Model

Material	Baseline mean ADC (SE) (10^{-6} mm^2/s) 1.5Ta, 2 b-value EPI	Estimated difference in mean ADC for 1.5Tb, 2 b-value EPI (10^{-6} mm^2/s) and 95% CI (low, high)	Estimated difference in mean ADC for 1.5Ta, 4 b-value EPI (10^{-6} mm^2/s) and 95% CI (low, high)	Estimated difference in mean ADC for 3Tc, 4 b-value EPI (10^{-6} mm^2/s) and 95% CI (low, high)
Water (PVP 0%)	1464.8 (479.7)	132.3 (−566.1, 830.8) $P = 0.6890$	198.8 (36.3, 361.3) $P = 0.0203$	157.8 (−102.2, 417.8) $P = 0.2124$
10% PVP w/w in water	1365.1 (42.0)	75.3 (34.0, 116.5) $P = 0.0005$	56.1 (8.5, 103.8) $P = 0.0216$	84.7 (−18.6, 188.1) $P = 0.1064$
14% PVP	998.5 (138.5)	57.4 (−174.0, 288.9) $P = 0.6011$	4.6 (−112.1, 121.2) $P = 0.9339$	67.8 (−40.6, 176.1) $P = 0.1996$
18% PVP	1066.6 (48.6)	65.9[a] $P = 0.0852$	-20.1[a] $P = 0.6837$	97.2[a] $P = 0.2043$
25% PVP	749.1 (53.2)	30.9 (−7.5, 69.3) $P = 0.1111$	6.8 (−51.9, 65.5) $P = 0.8156$	38.3 (−55.1, 131.7) $P = 0.4093$
40% PVP	451.0 (40.6)	38.8 (−4.0, 81.5) $P = 0.0745$	47.6 (−24.6, 119.7) $P = 0.1914$	31.1 (−26.4, 89.0) $P = 0.2807$

[a]Model did not converge on confidence interval results.

TABLE 5. Effect of Position from Magnet Isocenter on ADC for Scanner-Coil-Sequence Models

Material	Absolute estimated difference in mean ADC measured value for 1 cm change in position from magnet isocenter (10^{-6} mm^2/s) and 95% CI (low, high)	*P*-value
Water (PVP 0%)	74.2 (3.2, 145.3)	0.0418
10% PVP w/w in water	37.4 (30.0, 44.8)	<0.0001
14% PVP	63.1 (27.9, 98.4)	0.0019
18% PVP[a]	47.1 (20.4, 73.7)	0.0023
25% PVP	24.1 (13.8, 34.4)	<0.0001
40% PVP	16.3 (9.3, 23.4)	<0.0001

[a]Estimate, confidence interval and *P*-value are based on a model with the 1.5Tb data removed to enable confidence interval estimation for the estimated difference in mean ADC for 1 cm lateral change in position from coil center.

MRI scanners, breast coils, and typical clinical imaging sequences. We found that the breast phantom can be used to identify sources of bias and variability to guide quantitative measurement methods in multisystem studies. Comparing left and right sides of the coil, we identified distortions in the SS-EPI diffusion images, likely attributable to B_0 shim inhomogeneity, which means repeatability of the ADC measurements depends on shimming and position in the coil.

In general, ADC measurements were not biased across scanner-coil-sequence combinations. Temperature varied in the experiments conducted on the same configuration across time and in experiments conducted across different configurations, by as much as 5°C. Temperature fluctuations and differences across configurations may have increased bias, especially for the ADC measurements.

ADC repeatability in this study was mixed. In agreement with these results, a study with an ice-water phantom using similar protocols across multiple scanner-coil combinations found ADC variability increased at locations farther away from magnet isocenter,[7] a challenge in breast imaging. Giannotti et al. found high repeatability in ADC measurement of an ice-water phantom assessed weekly over 12 weeks.[12] In that study, using one scanner-coil-sequence configuration, the largest error was interobserver error in patient imaging. Recognizing that interobserver errors can be difficult to mitigate, studies that include multiple scanner-coil-sequence combinations could use this breast phantom during study set-up to minimize errors due to protocol selection or shim technique and increase repeatability of the ADC measurement.

Our results demonstrate that, on the clinical systems we used, ADC values are dependent on position within the magnet. Our generalized linear model indicates a statistically significant effect of ROI position within the magnet on ADC estimates for all of the breast tissue-mimicking

TABLE 6. Effect of Phantom Position in Right Coil Side on ADC for Scanner-Coil-Sequence Models

Material	Estimated difference in mean ADC for phantom in right coil side (10^{-6} mm^2/s) and 95% CI (low, high)	*P*-Value	Est. difference in mean ADC as a percent of baseline mean ADC value
Water (PVP 0%)	88.3 (−502.2, 678.7)	0.7519	6.03
10% PVP w/w in water	1.0 (−35.1, 37.0)	0.9575	0.07
14% PVP	−107.5 (−217.3, 2.4)	0.0544	−10.77
18% PVP[a]	−97.0 (−254.2, 60.2)	0.2036	−9.09
25% PVP	27.9 (−0.6, 56.3)	0.0549	3.72
40% PVP	40.4 (2.9, 77.9)	0.0352	8.96

[a]Estimate, confidence interval and *P*-value are based on a model with the 1.5Tb data removed to enable confidence interval estimation for the distance from the magnet isocenter.

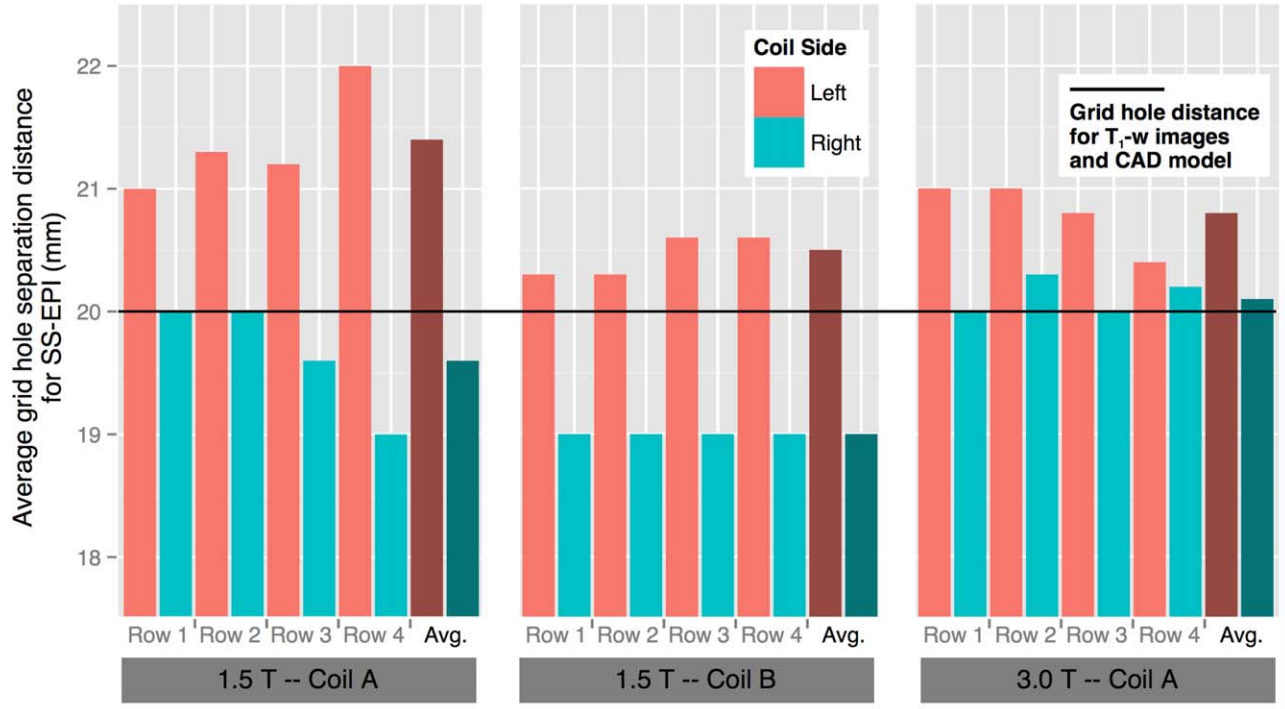

FIGURE 3: Average right–left grid hole separation distance for the SS-EPI images for each configuration (1.5T – coil A; 1.5T – coil B; 3T – coil A) for each distortion plate row, and the overall average distance for a configuration. In all cases, the left coil side distances were greater than the right coil side distances. The black line represents the actual grid hole separation distance of 20 mm from the CAD model. For the T1-w images, no measurable difference was found from the CAD model.

materials. ADC measurement is dependent on the diffusion-sensitizing gradients, which are known to have a nonlinear variation from the magnet isocenter.[7,20] Users should be aware that this dependence exists, and methods to mitigate gradient nonlinearity effects should be applied.[21] This breast phantom enables diffusion sequences to be tested for variation in artifacts with changes in sample position within the magnet, phase encoding direction, spectral bandwidth, or gradient nonlinearity corrections.

Position of the phantom in the left or right coil side was not a statistically significant predictor in the GLS model, but we did find differences in image distortion between the left and right sides of the coil. Specifically, it was found that SS-EPI diffusion weighted imaging method resulted in image distortion differences between the left and right sides of the coil. The distortion was measured across different vendors, breast coils, and field strengths; at the different imaging sessions; and stayed in the left/right

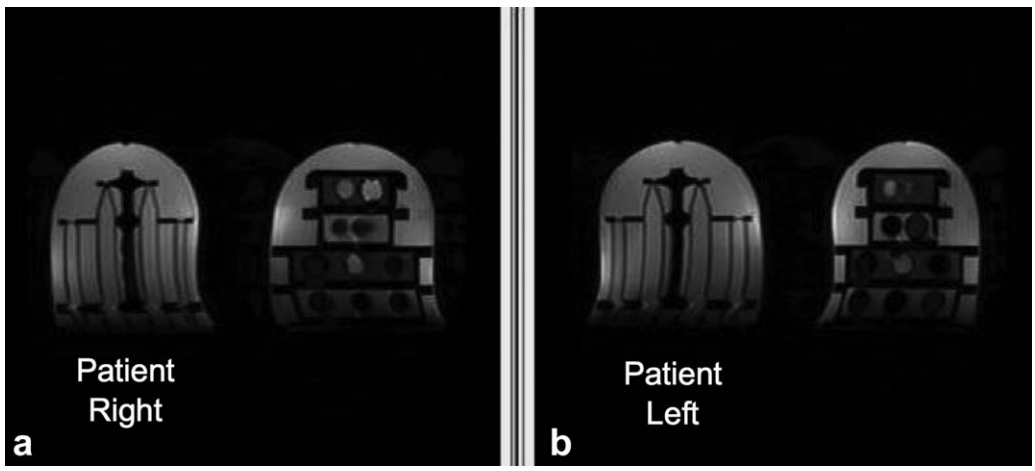

FIGURE 4: Breast phantom slices at the same depth locations, imaged in opposite coil positions from SS-EPI with b = 100 s/mm^2 diffusion weighting. The diffusion side is visibly thinner in the R/L direction (a), while the T_1 side is visibly thinner in the R/L direction (b).

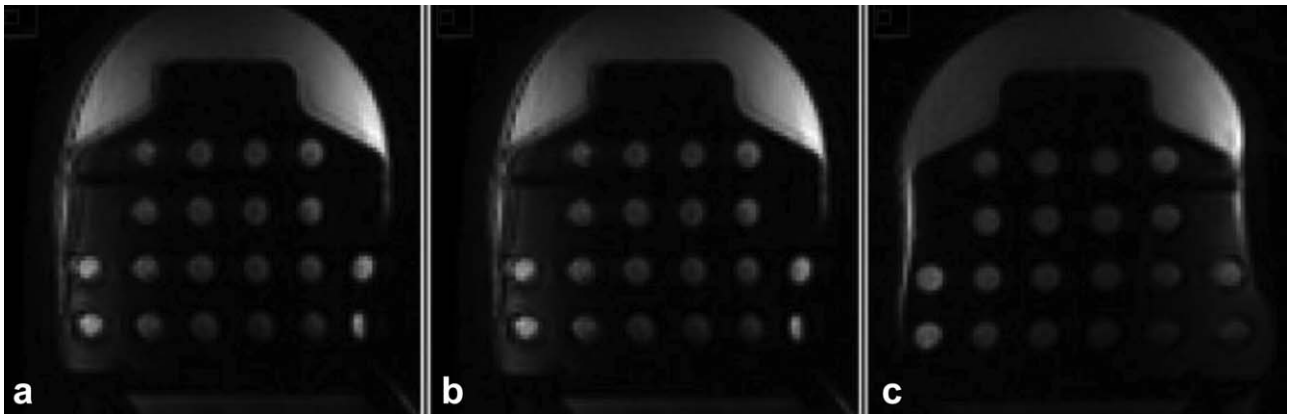

FIGURE 5: Central slice of the: T_1-w image (a), SS-EPI (b), b = 600 s/mm^2 patient right image, and SS = EPI, b = 600 s/mm^2 patient left image (c) of the diffusion side of the phantom.

orientation regardless of phase-encoding direction. There were no cases for which we did not find distortion in the SS-EPI diffusion images, indicating that the effect is likely caused by the SS-EPI imaging itself and not a factor related to specific equipment.

This distortion effect could be due to eddy currents and diffusion-weighted pulses or B_0 shim. If the source of the distortion were from eddy currents and diffusion-weighted pulses, the effect would scale with diffusion-weighting gradient strength (b-value) and would be along the direction of the phase-encoding gradient,[22] which was not the case. A second possible cause is from the B_0 shim,[23] which is known to be important in breast imaging.[24,25] In this study, we used the clinical breast imaging protocol shim procedure. In SS-EPI, background gradients can combine with the series of phase-encoded blipped gradient pulses. On one coil side, the B_0 shim adds to the strength of the blipped positive phase-encoded pulses and subtracts from it on the other coil side. In our study, distance was increased on the left coil side and decreased on the right coil side across two MRI system vendors. In a previous phantom study using a breast coil, geometric distortions were reported in the diffusion weighted images, although no quantitative measurements were reported.[26] In agreement with our findings, the geometric distortions did not change with diffusion weighting (different b-values). These findings indicate that this breast phantom can be used to evaluate distortion correction methods[23,27] and test improved shimming approaches, such as dynamic multislice shimming.[25]

One limitation of this phantom is the lack of temperature control, which increases the uncertainty in ADC measurements. NIST intends to characterize and publish the ADC values of the PVP solutions at a range of bore temperatures. Coupled with an accurate temperature measurement, we can reduce measurement uncertainty and increase reproducibility of ADC measurements. Our statistical methods include the median ADC and standard deviation from each 3D ROI, but the voxel size discrepancy between 1.5T and 3.0T is a limitation that potentially introduced bias in the data. From this initial study, we cannot make statements about manufacturer or product performance; a more substantial study is needed to make such statements.

In conclusion, this study demonstrates that this breast phantom is a tool to identify and correct scanner-coil-sequence bias and variability. This breast phantom may be used when establishing a multisystem study to create a protocol which minimizes differences (e.g., due to available sequences or shimming technique), to correct bias that cannot be minimized, and to weight results from each system depending on respective variability.

Acknowledgments

Contract grant sponsor: National Research Council postdoctoral scholarship; contract grant number: NIH/NCI 1U01CA151235

Contribution of the National Institute of Standards and Technology; not subject to copyright in the United States. Certain commercial instruments and software are identified to specify the experimental study adequately. This does not suggest endorsement by NIST or that the instruments and software are the best available for the purpose. The authors thank Ms. Evelyn Proctor for her assistance scanning the phantom. The authors thank Dr. Tom Chenevert for helpful conversations regarding EPI distortions.

References

1. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther 2009;86:97–100. doi: 10.1038/clpt.2009.68. PubMed PMID: 19440188.

2. Turnbull L, Brown S, Harvey I, et al. Comparative effectiveness of MRI in breast cancer (COMICE) trial: a randomised controlled trial. Lancet 2010;375:563–571. doi: 10.1016/S0140-6736(09)62070-5. PubMed PMID: 20159292.

3. Partridge SC, Mullins CD, Kurland BF, 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; 194:1664–1673. doi: 10.2214/AJR.09.3534. PubMed PMID: 20489111.
4. Galban CJ, Ma B, Malyarenko D, et al. Multi-site clinical evaluation of DW-MRI as a treatment response metric for breast cancer patients undergoing neoadjuvant chemotherapy. PLoS One 2015;10: e0122151. doi: 10.1371/journal.pone.0122151. PubMed PMID: 25816249; PMCID: PMC4376686.
5. Partridge SC, Rahbar H, Murthy R, et al. Improved diagnostic accuracy of breast MRI through combined apparent diffusion coefficients and dynamic contrast-enhanced kinetics. Magn Reson Med 2011;65:1759–1767. doi: 10.1002/mrm.22762. PubMed PMID: 21254208; PMCID: 3201817.
6. Turnbull LW. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. NMR Biomed 2009;22:28–39. doi: 10.1002/nbm.1273. PubMed PMID: 18654999.
7. Malyarenko D, Galban CJ, Londy FJ, et al. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging 2013;37:1238–1246. doi: 10.1002/jmri.23825. PubMed PMID: 23023785; PMCID: 3548033.
8. Min Q, Shao K, Zhai L, et al. Differential diagnosis of benign and malignant breast masses using diffusion-weighted magnetic resonance imaging. World J Surg Oncol 2015;13:32. doi: 10.1186/s12957-014-0431-3. PubMed PMID: 25889380; PMCID: PMC4328303.
9. Keenan KE, Wilmes LJ, Aliu SO, et al. Design of a breast phantom for quantitative MRI. 2016 J Magn Reson Imaging [Epub ahead of print].
10. Tuong B, Gardiner I. Development of a novel breast MRI phantom for quality control. AJR Am J Roentgenol 2013;201:W511–W515. doi: 10.2214/AJR.12.9571. PubMed PMID: 23971483.
11. Freed M, de Zwart JA, Loud JT, et al. An anthropomorphic phantom for quantitative evaluation of breast MRI. Med Phys 2011;38:743–753. PubMed PMID: 21452712; PMCID: 3033878.
12. 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;84:1729–1736. doi: 10.1016/j.ejrad.2015.05.032. PubMed PMID: 26078100.
13. Horkay F, Pierpaoli C, Basser PJ, inventors. Phantom for diffusion MRI imaging. USA2012. US 20120068699 A1.
14. Pierpaoli C, Sarlls J, Nevo U, Basser PJ, Horkay F. Polyvinylpyrrolidone (PVP) water solutions as isotropic phantoms for diffusion MRI studies. In: Proceedings of the 17th Annual Meeting of the ISMRM, Honolulu, 2009.
15. 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;31:562–570. doi: 10.1002/jmri.22078. PubMed PMID: 20187198.
16. Partridge SC, Heumann EJ, Hylton NM. Semi-automated analysis for MRI of breast tumors. Stud Health Technol Inform 1999;62:259–260. PubMed PMID: 10538368.
17. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998; 316:1236–1238. PubMed PMID: 9553006; PMCID: PMC1112991.
18. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1:43–46. PubMed PMID: 2081237.
19. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. Am J Epidemiol 1995;142:904–908. PubMed PMID: 7572970.
20. Bammer R, Markl M, Barnett A, et al. Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging. Magn Reson Med 2003;50:560–569. doi: 10.1002/mrm.10545. PubMed PMID: 12939764.
21. Tan ET, Marinelli L, Slavens ZW, King KF, Hardy CJ. Improved correction for gradient nonlinearity effects in diffusion-weighted imaging. J Magn Reson Imaging 2013;38:448–453. doi: 10.1002/jmri.23942. PubMed PMID: 23172675.
22. Jones DK, Cercignani M. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed 2010;23:803–820. doi: 10.1002/nbm.1543. PubMed PMID: 20886566.
23. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. Magn Reson Med 1995;34:65–73. PubMed PMID: 7674900.
24. Maril N, Collins CM, Greenman RL, Lenkinski RE. Strategies for shimming the breast. Magn Reson Med 2005;54:1139–1145. doi: 10.1002/mrm.20679. PubMed PMID: 16217775.
25. Lee SK, Tan ET, Govenkar A, Hancu I. Dynamic slice-dependent shim and center frequency update in 3 T breast diffusion weighted imaging. Magn Reson Med 2014;71:1813–1818. doi: 10.1002/mrm.24824. PubMed PMID: 23798360; PMCID: PMC3830606.
26. Sinha S, Lucas-Quesada FA, Sinha U, DeBruhl N, Bassett LW. In vivo diffusion-weighted MRI of the breast: potential for lesion characterization. J Magn Reson Imaging 2002;15:693–704. doi: 10.1002/jmri.10116. PubMed PMID: 12112520.
27. Newitt DC, Tan ET, Wilmes LJ, et al. Gradient nonlinearity correction to improve apparent diffusion coefficient accuracy and standardization in the american college of radiology imaging network 6698 breast cancer trial. J Magn Reson Imaging 2015;42:908–919. doi: 10.1002/jmri.24883. PubMed PMID: 25758543.