Technical Note

Multi-system Repeatability and Reproducibility of Apparent Diffusion Coefficient Measurement Using an Ice-Water Phantom

Dariya Malyarenko, PhD,[1] Craig J. Galbán, PhD,[1] Frank J. Londy, RT(R),[1] Charles R. Meyer, PhD,[1] Timothy D. Johnson, PhD,[2] Alnawaz Rehemtulla, PhD,[3] Brian D. Ross, PhD,[1] and Thomas L. Chenevert, PhD[1]*

Purpose: To determine quantitative quality control procedures to evaluate technical variability in multi-center measurements of the diffusion coefficient of water as a prerequisite to use of the biomarker apparent diffusion coefficient (ADC) in multi-center clinical trials.

Materials and Methods: A uniform data acquisition protocol was developed and shared with 18 participating test sites along with a temperature-controlled diffusion phantom delivered to each site. Usable diffusion weighted imaging data of ice water at five b-values were collected on 35 clinical MRI systems from three vendors at two field strengths (1.5 and 3 Tesla [T]) and analyzed at a central processing site.

Results: Standard deviation of bore-center ADCs measured across 35 scanners was <2%; error range: −2% to +5% from literature value. Day-to-day repeatability of the measurements was within 4.5%. Intra-exam repeatability at the phantom center was within 1%. Excluding one outlier, inter-site reproducibility of ADC at magnet isocenter was within 3%, although variability increased for off-center measurements. Significant (>10%) vendor-specific and system-specific spatial nonuniformity ADC bias was detected for the off-center measurement that was consistent with gradient nonlinearity.

Conclusion: Standardization of DWI protocol has improved reproducibility of ADC measurements and allowed identifying spatial ADC nonuniformity as a source of error in multi-site clinical studies.

Key Words: diffusion; MRI; phantom; ice-water; quality control; gradient nonlinearity
J. Magn. Reson. Imaging 2013;37:1238–1246.
© 2012 Wiley Periodicals, Inc.

DIFFUSION WEIGHTED IMAGING (DWI) is widely used in MR examinations for diagnostics (1–4) and is considered a potential biomarker of therapy response assessment. The desire to use apparent diffusion coefficient (ADC) as a viable quantitative biomarker for cancer diagnosis and treatment response monitoring (4) requires determination of measurement uncertainty and systematic dependencies. The complexity of water diffusion measurement in living systems containing anisotropic tissues, perfusion and other sources of motion (5–8) combined with limited instrumental sensitivity (9–12) confound quantitative ADC measurements. Furthermore, synchronization and standardization of DWI acquisition techniques across multiple MRI platforms (4,8,9,13) is necessary for multicenter clinical trials.

This study focused on MRI system-related factors affecting ADC measurements using a standardized acquisition protocol and a single known diffusing medium as an estimate of baseline repeatability and reproducibility. To detect clinically significant changes in diffusion measurements, the sources of technical variability have to be characterized and controlled relative to anticipated biologic/therapeutic diffusion changes (7,11,12,14). Because molecular mobility is dependent on temperature, knowledge and control of temperature of the diffusing media are essential in the course of measurements used for technical quality evaluation. It was recently shown (15) that water maintained in an ice-water bath can serve as a universal temperature-controlled fluid in ADC phantoms to test instrumental variability. However, in the previously described study (15), the DWI acquisition protocol was somewhat variable and defined by local site

[1]Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA.

[2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA.

[3]Departments of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA.

*Address reprint requests to: T.L.C., University of Michigan Hospitals, 1500 E. Medical Center Drive, UHB2 Room A209, Ann Arbor, MI 48109-5030. E-mail: tlchenev@umich.edu
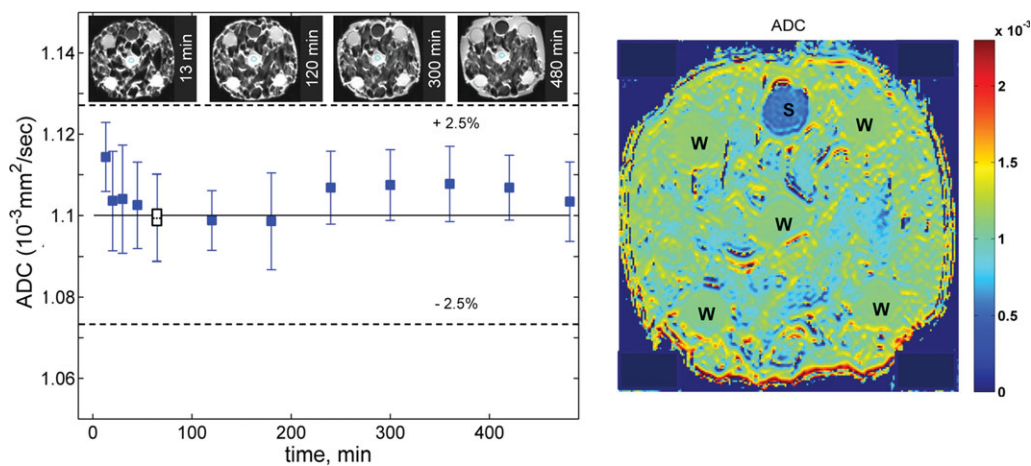
Figure 1. Temporal stability (~8 h) of measured ADC (central tube) of the ice-water phantom in head coil at 3T (b = 1000) is within ±2.5% (dashed lines) of nominal ADC value at 0°C (Ref.(21), solid line). The inserts on top show axial T2wt MR images corresponding to different ADC measurement time points (as indicated by labels) relative to initial phantom filling with ice. Blue circles in the middle of the center tube mark the locations of the 1 cm ROI from which the ADC measurements were performed. Data (blue squares) are presented as mean and standard deviation (error bars) of ADC measurements across an ROI (typically, ~90 pixels). The vertical dimension of the open-rectangle represents the standard deviation of the mean ADC values measured for four consecutive passes of the same exam (~2 min/pass), while the dashed line inside is the corresponding four-pass average ADC. ADC color-map in the right pane corresponds to the first SOP-recommended measurement point (open-rectangle). The labels on the ADC map indicate the positions of water (W, center and four corners) and sucrose (S, top) tubes.

preferences. Protocol variability impact on the measurement was unknown. In contrast, this study uses common acquisition protocol adopted by all sites to achieve an unmixed reproducibility estimate. In addition, a more complex phantom geometry as well as off-center measurements allowed detection and characterization of spatial error contribution. Other compounds (e.g., high concentration sucrose in present study) can be added to phantoms to provide a range in diffusion properties as desired (16,17).

The primary objective of this study was to determine the extent of quantitative agreement for ADC values of a known, temperature-controlled fluid measured on multiple platforms, sites, and field-strength MR systems using a common acquisition protocol. This approach characterizes baseline technical performance of a given system for diffusion measurements (12,14,15) without interference from complex temperature or concentration dependence (13,17,18). The described procedures may be used for site certification/quality control and may serve as a basis for standardization of diffusion imaging protocols across multi-vendor/multi-site for clinical trials that use ADC as a potential biomarker.

MATERIALS AND METHODS

Temperature-Controlled Phantom

A variation of the ice-water-based diffusion weighted imaging (DWI) phantom (15) was devised to measure ADC over a clinically relevant spatial range. As shown in Figure 1, the phantom consisted of an ice water-filled container holding five, 29 mm diameter tubes filled with distilled water (four at phantom corners and one in the center) as the diffusion medium stand-ard, as well as, 1 tube filled with a sucrose solution (9 gm sucrose/30 mL water) to provide some diffusion contrast (see axial MR images of phantom in Fig. 1). Because sucrose solution is not a "universal" fluid, it was not used for absolute ADC measurements. The six tubes were filled and sealed at phantom fabrication, although the interstitial space was filled with ice water at each site and each use per detailed phantom preparation instructions. Recommended ice fill procedure included a two-step preparation: (step 1) prime the phantom with ice water for 10 min for an initial rapid cool-down of the water tubes from room temperature, and (step 2) add ice to replenish the relatively large volume of ice melted during step 1, then let the phantom equilibrate for an hour before scanning. Once filled and equilibrated, the ice-water bath surrounded the measurement tubes with a temperature-controlled, high thermal capacity environment. At thermal equilibrium just above 0°C the diffusion coefficient of water is known to be 1.1×10^{-3} mm²/s (19–21). Thermal mass and an additional foam insulation sleeve allowed use of the phantom over several hours (Fig. 1). If temperature was controlled to within 0–0.5°C, the measured diffusion coefficient should be determined within 1–2% of literature values (15,19,20). To illustrate temporal stability, the central tube of ice-water phantom was monitored on one system over the course of 8 h. The ice melted faster closer to the phantom surface, thus exposing corner tubes to higher temperature gradients than the center tube. The phantom container was semi-flexible to avoid creating a vacuum as the ice melted. As such, the geometrical aspects of the phantom were considered not crucial which kept fabrication costs low and allowed production of 25 phantoms for delivery to 18 institutions and three MRI manufacturers.

MR DWI Data Acquisition Protocol

A core DWI data acquisition protocol was designed for compatibility across several clinical MR imaging platforms (GE, Philips, Siemens) and field strengths (1.5 and 3 Tesla [T]). Key elements of the core protocol were: single spin-echo, single-shot EPI; b-values = 0, 500, 800, 1000, 2000 s/mm^2; DW axes = X, Y, Z in the lab frame; repetition time (TR) = 8000 ms; echo time (TE) = 100 ± 10 ms; Acquisition matrix = 128 × 128; Phase = Anterior/Posterior (A/P), Freq. = Right/Left (R/L); FOV = 240 × 240 mm^2; 25 slices, 6 mm thick, 1 mm gap; Bandwidth = 1.5–2.6 kHz/pixel; NEX = 1; and no parallel imaging to avoid variability in vendor-specific parallel imaging algorithm implementation. To permit compliance across all systems, the appropriate ranges of protocol parameters were established by means of communication with clinical scientists from each of the vendors. The listed parameters were available from DICOM headers of submitted series to insure protocol compliance. Both head and torso phased array coils were used for signal reception. To estimate intra-exam repeatability and confirm thermal equilibrium, sites were instructed to acquire multiple DWI "passes", where all b-values were acquired in each pass (approximately 2 min/pass). Four passes were first acquired in the head coil. Then the phantom was repositioned in the torso coil for two passes near R/L = 0; two passes offset to the left ~110 mm; and two passes offset to the right ~110 mm. Axial measurements along the central tube provided a spatial extent of approximately 140 mm (limited by tube length) in the superior/inferior (S/I) direction.

To independently assess relative gradient amplitude over the range of measured R/L and S/I offsets, local gradient strength was measured using a distortion phantom comprised a three-dimensional (3D) array of point signal sources spaced 15 mm apart. Standard spin-echo MRI was performed without and with the spatial geometric correction routine normally applied on the scanner. Local distances between adjacent point sources were measured on both non corrected and geometrically corrected images as proxies for "actual" and "nominal" local gradient strength. The square of the ratio of these distances was used as a proxy for ratio of [b-actual/b-nominal] for DWI experiments, because b-value scales with the square of gradient amplitude.

System day-to-day repeatability (including phantom preparation) was assessed by comparing complete runs at the center and with right-to-left offset on two different days. Spatial offsets along the center tube, as well as a whole-phantom displacement from the isocenter, were used to sample spatial uniformity of ADC. Sites were instructed to provide a DICOM screenshot of a region-of-interest (ROI) defined on an ADC map generated using scanner software to corroborate with ADC maps produced by centralized processing.

ADC Measurements and Data Analysis

A single Matlab-based package (MathWorks Inc., Natick, MA) was developed for centralized diffusion data analysis to mitigate variability introduced by different software packages. While all images were provided in DICOM format, seven distinct source-dependent image order scenarios were discovered, which were addressed by a custom DICOM import and sort routine. The output of this routine contained relevant system information, acquisition settings, and images in a uniform data structure format regardless of image source. DICOM headers of submitted series from forty scanners were checked for protocol compliance. For each nonzero b-value, apparent diffusion coefficient (ADC) maps were computed by:

$$ADC_b = \frac{1}{b} ln \left[\frac{s_0}{s_b} \right], \qquad [1]$$

where S_O is the b = 0 image and S_b is the isotropically weighted DWI at the given b-value. That is, multiple ADC maps (i.e., ADC_{500}, ADC_{800}, ADC_{1000}, ADC_{2000}) were derived analytically using Eq. [1] as opposed to a single ADC derived by numerical fitting of multiple b-value DWI. Because the diffusing medium is known to be mono-exponential, the array of ADC values should not exhibit any b-value dependence. ADC maps provided by the sites were used to confirm that the centralized processing of raw DWI-data resulted in the same ADC values, as would be measured by system-specific software. This ensured that no software-dependent bias was introduced by preprocessing steps. Signal-to-noise ratio (SNR) measurements for head coil DWI-images were performed on central-tube regions of interest (ROIs) for separate b-values to ensure that detection level was sufficient for unbiased ADC determination. An estimate of noise for SNR measurements was obtained by the temporal standard deviation for each pixel from consecutive passes of the core DWI protocol.

In this study, reproducibility (and repeatability) of a given measure is defined as 2 times the standard deviation (STD) of that measure expressed as percentage of the mean. That is, reproducibility across systems represents the range within which 95% of measurements are expected to fall. The fundamental measurement was the mean of each ADC metric within a 10 mm diameter circular ROI of approximately 90 pixels. ROIs were defined on ADC maps at the center of the middle tube on all usable slices. Typically 20 ROIs were defined along the middle tube (on axial slices center-to-center 7 mm apart) for one phantom position yielding approximately 80 ROIs from all four phantom positions on each exam date. Spatial coordinates of all ROI centers were stored so that subsets of ROIs could be retrospectively graphically selected for analyses detailed below, as well as for automatic application the ROIs to all ADC maps across all like passes. Additional ROIs were defined in the four corner tubes for estimation of intra-exam repeatability for six representative systems (2 field strengths for each of 3 vendors). Corner-tubes were not used for derivation of "baseline" reproducibility statistics for water ADC due to mixed systematic effects of nonlinearity bias and different phantom positioning/rotation by participating sites. Likewise,

off-center position of sucrose tube and its proximity to air resulted in more artifacts limiting the number of useful (homogeneous) ROIs for ADC measurements and precluded generation of sufficient statistics for sucrose ADC across sites. Statistical analyses for multi-system repeatability and reproducibility were performed only on measurements obtained from the central water tube to minimize mixed spatial effects (X-Y-Z offsets), ROI nonuniformity artifacts and ensure best temperature control.

ADC mean and standard deviation statistics derived from central slice ROIs using the head coil and torso coil with the phantom at isocenter were compiled for a fundamental comparison of ADC accuracy and reproducibility across all systems. Statistics from central-slice ROIs of the middle tube with the phantom offset in the torso coil by a nominal ±110 mm in R/L direction were combined to estimate accuracy and reproducibility for R/L off-center ADC measurements. To determine ADC accuracy and reproducibility for offsets in S/I direction, head coil ROIs of the middle tube on peripheral slices at a nominal ±70 mm S/I offset were combined. Intra-exam repeatability was derived from the relative difference in ROI means measured over 2 to 4 passes acquired in immediate succession. Inter-exam repeatability was derived from the relative difference in ROI means measured on a given system over two scan dates.

Intra-exam repeatability was defined as percent ratio of 2·STD to mean ADC value measured for four consecutive passes of the protocol for the same (single) ROI location at the middle or peripheral tubes. Day-to-day repeatability for each system was measured by comparing %difference from the 2-day average ADC. Similar to repeatability, reproducibility of center and off-center measurements between scanners was estimated from 2·STD percent ratio deviation of mean ADC values from the corresponding ROIs for individual b-values. The scanners were grouped by vendor and field strength. The bias error was measured as a percent deviation of system-wise average ADC from the expected value (19–21). Multi-vendor/site/b-value/field-strength comparisons and reproducibility analysis were performed using SPSS (version 19) package to generate statistics and graphical output. Data dispersion was characterized by mean, median, standard deviation, range, and outlier analysis.

RESULTS

Quality assurance inspection of DICOM parameters allowed identification of 5 systems (of the original 40) that did not comply with the acquisition protocol and their corresponding data were discarded from the subsequent analysis. The remaining 35 (protocol compliant) systems represented three vendors, labeled A, B, and C (at 1.5T and 3T, respectively) with the following scanner "population": A - 10 (5 and 5), B - 17 (6 and 11), and C - 8 scanners (4 and 4). Twenty-eight of 35 systems provided 2-day acquisition data to evaluate day-to-day repeatability. For the worst condition at b = 2000, the average SNR of the trace DWI image was found to be 110 at 3T (range, 60 to 175) and 40 at 1.5T (range, 30 to 65). Based on simulation, ADC bias error due to Rician noise at the lowest SNR = 30 would be negligible (10). Thus SNR analysis across specific field and vendor subgroups was not considered informative within the context of ADC measurement.

Figure 1 illustrates temporal stability of the ice-water diffusion phantom as well as within-ROI confidence intervals (error bars defined as ±1 standard deviation of ROI pixels) for one 3T system central tube ADC_{1000} calculated by Eq. [1]. Image inserts illustrate the phantom morphology at different time points along the time-line of the experiment (13, 120, 300, and 480 minutes relative to initial ice filling). The color-map of ADC_{1000} measurement (right pane) corresponds to the protocol-recommended time-point (65 min, un-filled black symbol) for the first pass of the core DWI protocol using head-coil. After initial cool-down (points 1–4, < 1 h), thermal equilibrium was achieved and persisted for over 4 hours as is evident from lower deviation from nominal value (21) for ADC measurements in this time range. The mean ADC measurements during thermal equilibrium were within 0.1% of literature value of 1.1×10^{-3} mm^2/s (21) (marked as a solid line in Fig. 1). The standard deviation of ROI means for 4 passes ($\sim$2 min/pass) of the core DWI protocol for the central tube ROI is illustrated by the vertical extent of the un-filled black symbol on the "65min" time-point (STD for this system = 0.005×10^{-3} mm^2/s). The observed average intra-exam repeatability estimated from four consecutive passes of the same measurement for 70 ROI locations in all five water tubes was within 1%. All measurements (over 8-h period) for this system were within 2.5% of the nominal value (dashed lines in Fig. 1).

Day-to-day repeatability for each system was measured by comparing %difference from 2-day average ADC of the central tube ROI for four b-values of 500, 800, 1000, and 2000 s/mm^2. The summary of measurements pooled over field strength and grouped by vendor is plotted in Figure 2 for (a) head-coil at the center and (b) torso-coil at 110 mm right and left offset (average ADC). For all systems this day-to-day repeatability was within 4.5%. Across all scanners and all b-values the average of median day-to-day difference in ADC values (as illustrated by solid lines in the box-plot of Fig. 2) was less than 1.5%, and the overall standard deviation for all measurements was below 2.3%. A greater degree of variability was observed in vendors A and B for right/left-offset ADC measurements. Several outlier systems (>5% difference between days) were detectible for head coil data at 3T (vendor B) and 1.5T (vendor C).

Figure 3 summarizes results for ADC measurements at the central ROI grouped by vendor (horizontal axis), field strength (left-to-right), and receiver coil type (a, b). The expected ADC for water at 0°C is marked as a solid line with ±5% of this value shown as dashed lines. One clear outlier 3T system generated ADC values as high as 1.8x10^{-3} mm^2/s ($\sim$70% deviation from literature value) for 3 of four b-values. Excluding this one outlier, multi-system reproducibility for the center
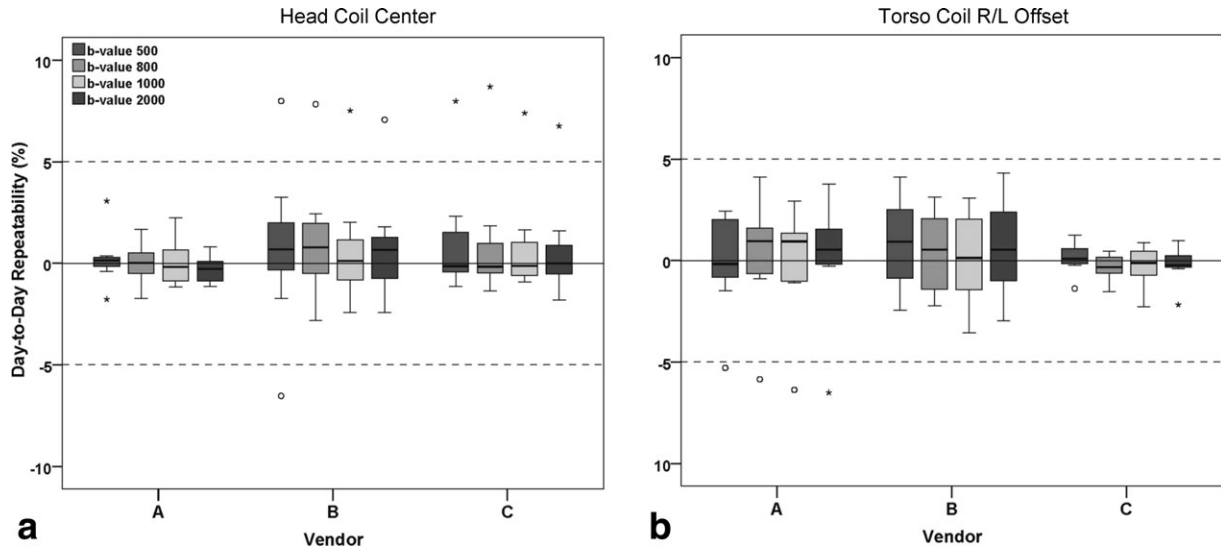
Figure 2. Box-plot summary of day-to-day repeatability for multi-site results pooled over field strength plotted as a percent difference in ADC measurements taken on 2 different days ($\sim$ 30 h apart). On each box, the central mark is the median, the vertical boundaries mark the 25th and 75th percentiles, the whiskers extend to the most extreme data-points excluding outliers, and the outliers are plotted individually. Data for different b-values of 500, 800, 1000, and 2000 s/mm^2 are coded by gray shades left-to-right and grouped by vendor: (**a**) head coil, center; (**b**) torso coil, right-to-left offset 110 mm.

ADC are within 2.8% (Fig. 3a) and 3.1% (Fig. 3b) for head and torso coils, respectively. The bias error of the mean measured ADC in respect to literature value is within 2.5% for head and 3.5% for torso coil at 3T, and approximately half of these values at 1.5T. There is no significant dependence on b-value relative to the typical ROI measurement error (bar size in Fig. 1).

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) and separately the average of right–left (R/L) offsets were used. Off-center ADC measurements were less reproducible across scanners. As depicted in Figure 4, measurement variability

was significantly higher with $\pm$70 mm offset from center in S/I direction for head and $\pm$110 mm R/L offset in torso coil, respectively. As is evident by comparison of data box sizes in Figure 4 versus Figure 3, for each vendor, inclusion of spatial offsets doubled or, in case of vendor A torso coil, tripled the dispersion (standard deviation) of measured values compared with center ROI measurements (Fig. 3). The apparent errors were higher at 3T and show distinct vendor-specific pattern. Vendor-specific standard deviation as well as bias error (with respect to expected ADC value) exceeded 3% at 1.5T and 5% at 3T, and vendor A exceeded by 15% at 3T. Measured ADCs were proportionally underestimated for S/I offsets (Fig. 4a), and
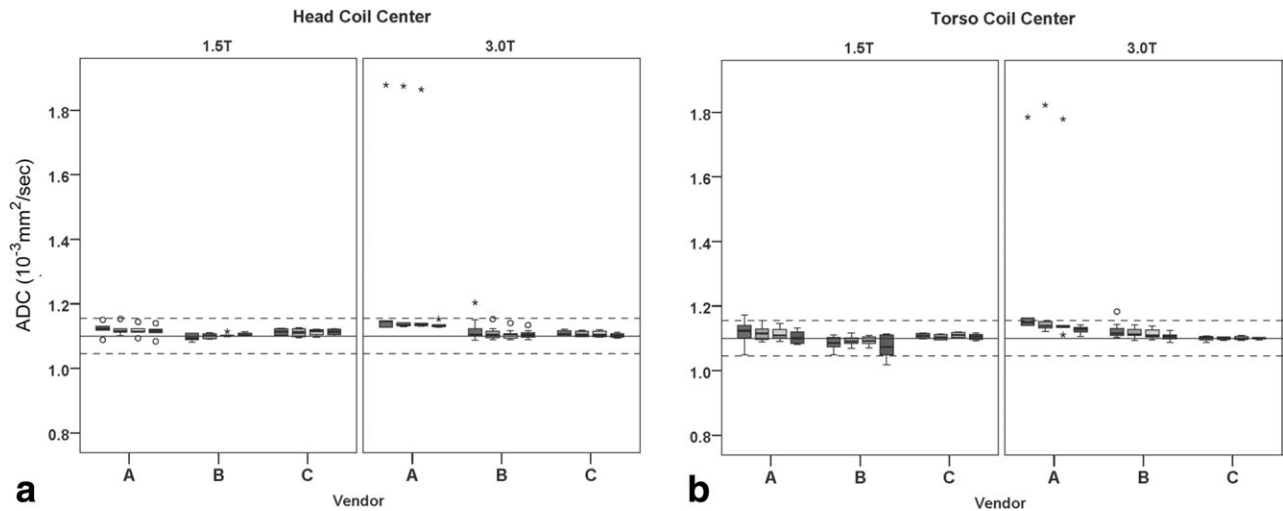

Figure 3. Box-plot summary for between-system reproducibility of ADC measured form central ROI grouped by b-values (500, 800, 1000, and 2000 s/mm^2, coded by shades of gray left-to-right) and by vendor (A,B,C) at two fields (1.5T - left column, 3T - right column) for head coil (**a**) and torso coil (**b**). Dashed lines delineate 5% deviations from reported ADC value of water (21) at 0°C (solid line).
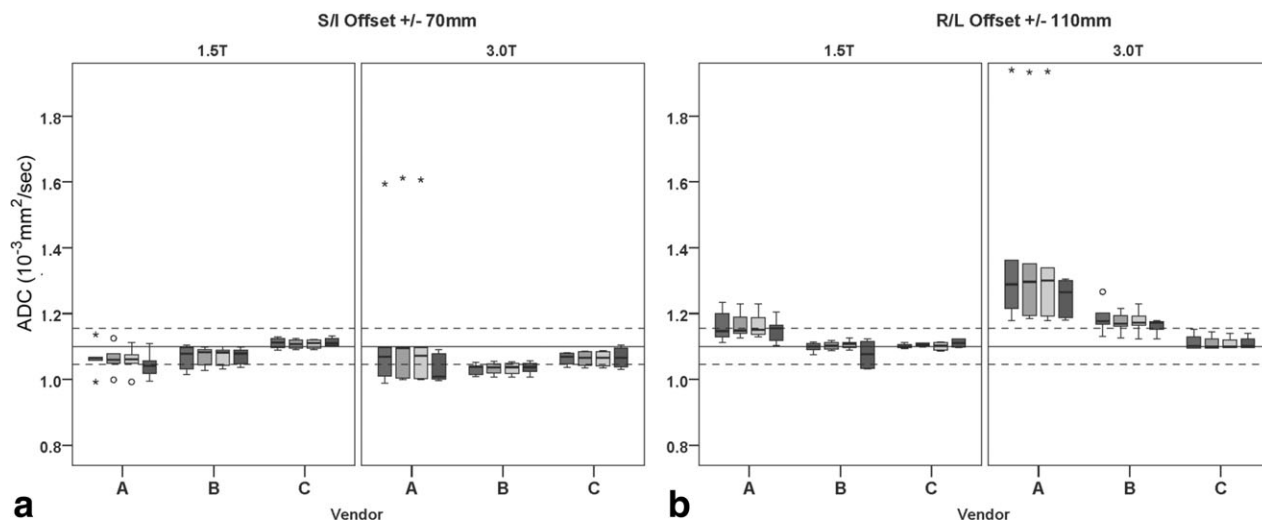
Figure 4. Effect of off-center shift on between-system reproducibility for ADC grouped similar to Figure 3 for head coil S/I-offset (**a**) and torso coil R/L-offset (**b**). Dashed lines mark 5% deviations from reported ADC value of water (21) at 0°C (solid line).

overestimated for R/L offsets (Fig. 4b). Similar to Figure 3, no significance dependence on b-value is observed.

More detailed spatial dependence of measured ADC is depicted in Figure 5a,b for one representative 3T system. These plots are consistent with observation of larger (approximately quadratic) ADC errors for larger offsets, and the opposite sign errors in S/I versus R/L direction (Fig. 4). The results of empiric measurement of local gradient strength are depicted in panes (c,d) of Figure 5. Square of the ratio of "actual" to "nominal" gradient amplitude exhibits parabolic dependence on offset due to gradient nonlinearity. Both the sign and the scale of observed ADC bias (a,b) are consistent with the systematic measurements of gradient nonlinearity (c,d). Combining all systems excluding the one clear outlier, reproducibility (2·STD percent ratio) of the ADC_{1000} measurement was 2.8% at isocenter (head coil), 6.8% for S/I offsets and 11.3% for R/L sampled offsets.

DISCUSSION

The purpose of our multi-center study was to assess repeatability, reproducibility and quantitative quality control of ADC measurements across vendors, field strength and b-values using a standardized acquisition protocol. In a prior study, the DWI acquisition protocol was variable and based on each sites' local "brain DWI" exam, and the issue of spatial variability was not addressed (15). By standardizing data acquisition conditions across participating test sites and using a temperature-controlled fluid phantom/procedure that allowed off-center ADC measures over a wide range of b-values envisioned for clinical trials, we sought to determine the baseline for systematic reproducibility level relevant for brain and body DWI multi-center trials (8,9,13). Our aim was to distinguish between sources of observed technical variabili-

ty and quantify technical deficiencies of individual systems (12,14). Certainly one 3T system was detected as a clear outlier.

Investigation of temporal stability of the ice-water phantom (Fig. 1) confirmed that over the course of ADC measurements no significant variability (STD < 1%) was introduced by temperature fluctuations within phantom and that this design allowed for over 4 h of use. Because all subsequent experimental measurements were performed during thermal equilibrium, ADC variability measurements were not substantially influenced by fluctuations in temperature within the diffusing medium. Note that the intra-exam repeatability error for all ROIs was consistently less than one third of the spatial standard deviation "noise"-error within ROI (<1.5%, Fig. 1). This relation was observed for average repeatability versus average ROI-noise of the ADC measurements in other studied systems, indicating ROI "inhomogeneity" noise (across ROI pixels) as a dominant source of ADC measurement error. This is consistent with observation of slightly larger reproducibility errors and bias (with respect to true value) across multiple systems with greater error in torso coil versus head coil ADC measurements (Figs. 2–4). By inspection of Figure 3 versus Figure 4, higher variability was apparent for off-center measurements at higher field (Fig. 4). The pattern of ADC overestimation for R/L offsets and underestimation for S/I offsets was commonly observed for individual 1.5T and 3T systems, but was more prominent on 3T systems. The spatial pattern appeared "saddle shaped" with greatest ADC uniformity near magnet isocenter and steeper nonuniformity that increases with distance off-center. Therefore, variability in positioning the phantom across sites and days translates to greater variability for offset ADC measurements, particularly for 3T systems that tend to exhibit steeper spatial dependence.

Multiple runs for the same exam were highly repeatable (< 1%), and day-to-day differences did not exceed
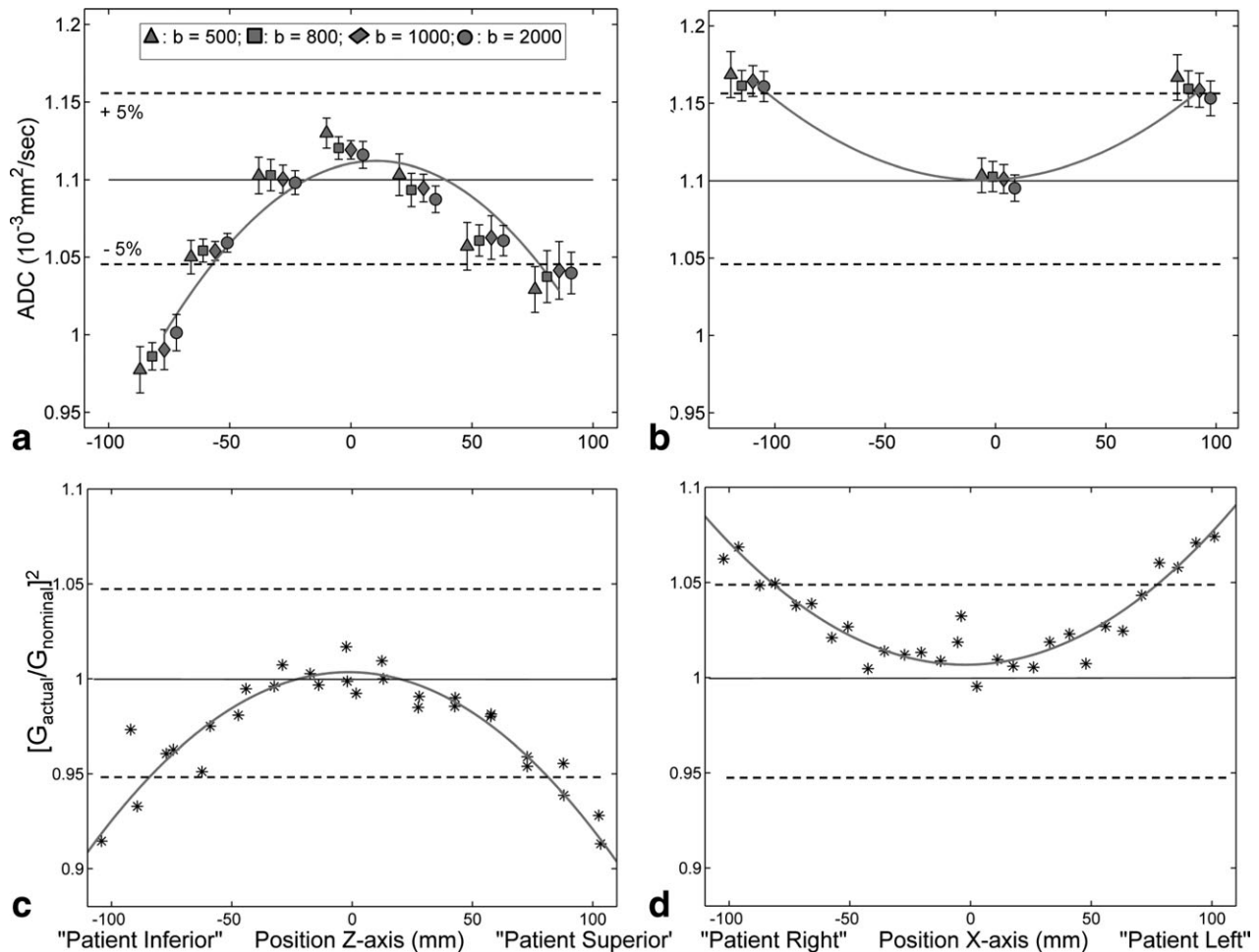
Figure 5. Spatial dependence of measured ADC (**a,b**) and squared gradient deviation from linearity (**c,d**) for a single representative system at 3T for S/I-offset (left column) and R/L-offset (right column). Filled-symbols with error-bars in (a, b)-panes mark ADC measurements for b = 500, 800, 1000, and 2000 (offset left-to-right for clarity). Asterisks in (c,d) panes represent measured squared ratio of true-to-nominal local gradient strength (b-ratio proxy) as a function of offset from isocenter. Data scatter reflects measurement error for gradient strength. Solid gray line represents a quadratic fit for b = 1000 in (a,b) and a fit for true-to-nominal b-ratio in (c,d) panes. Dashed lines mark 5% deviations from reported ADC value of water (21) at 0°C (black solid line) for (a,b), and deviation from nominal gradient (true-to-nominal ratio of 1 marked by solid line) for (c,d)-panes.

3% (Fig. 2) for the majority of systems. Closer examination of images for several outlier systems (> 5% variation) revealed larger areas of melted ice around phantom tubes (see, e.g., morphology around top outer tube in the last insert of Fig. 1). This indicated that observed higher day-to-day variability may be due to not fully achieving thermal equilibrium on one or both exam days at some sites that presumably did not follow recommended ice fill procedure. Note that day-to-day variations for each system included components both from phantom preparation and repositioning. No apparent dependence of day-to-day repeatability on phantom position (Fig. 2) confirmed the absence of temporal bias in system performance.

By first principles, self-diffusion of water is independent of timescale thus should appear mono-exponential as a function of b-value. This is consistent with our observation of no systematic dependence on b-value for all studied systems (Figs. 3 and 4), except the clear 3T outlier by vendor A. Further analysis of

this outlier revealed that ADC maps from two of three orthogonal gradient directions produced exceptionally high ADC values for b-values 500, 800, and 1000; although the b = 2000 directional and trace ADC maps were essentially equivalent with other vendor A systems. For this system, the large bias error was repeated for both measurement days. The source of this error is unknown.

The major contributors to observed higher variability and bias for off-center ADC measurements (Fig. 4) are likely due to spatial dependent error for individual systems (Fig. 5) coupled with slight differences phantom positioning. Several instrumental imperfections (or combinations thereof) may account for the detected spatial nonuniformity of the measured ADC values for individual systems: phantom vibration, concomitant fields, poor shim, and gradient nonlinearity. By exploring observed trends in spatial variations of measured ADC we are able to determine the dominant source of observed spatial nonuniformity

(Fig. 5). Note, that spatial offsets from isocenter were nominal offsets (± 110 mm R/L, and ± 70 mm S/I) defined by protocol. There was a range of actual offsets which contributed to the increased/decreased range of measurable shifts from isocenter combined in Figure 5. If vibrations were major contributors to the ADC error, they would systematically increase apparent ADC with increasing b-value, which does not agree with our observations (Fig. 5). Furthermore, apparent decrease in ADC is observed in superior–inferior direction in contrast to increase right–left. The effect of concomitant fields (22) from pulsed gradients would manifest itself in decreasing ADC with increasing magnetic field (3T versus 1.5T). However, the opposite change is observed in Figures 3 and 4, and its absolute value could account for 2–3%, at most, of the observed spatial ADC variation. Poor shim would be anticipated to produce an asymmetric effect at the opposite edges (decrease versus increase), which would be enhanced by de-shimming. Both of these were ruled out by our experimental observations of symmetric spatial dependence for ADC with respect to isocenter (Fig. 5).

As illustrated in Figure 5, ADC bias scales with the ratio of actual to nominal b-value (measured by the ratio of squared gradient strength). Thus empirically, the observed (approximately) quadratic offset-dependence of ADC is consistent with measured gradient nonlinearity errors. Gradient nonlinearity correction (23) is fixed for a given coil design (independent of a patient). Consequently, the observed spatial nonuniformity of ADC could be corrected by constructing and applying instrument-dependent gradient maps (23). The specifics of the ADC-correction procedure will be described in a future article.

In conclusion, standardization of the DWI protocol for multi-site studies has improved reproducibility of ADC measurements compared with "local-standard" (site-specific) protocol (15). Measured ADC of the ice-water phantom near magnet isocenter was within 3% of literature value for 95% of the systems. Reproducibility between vendors was invariant to b-value and field strength (except for one outlier due to direction dependence discussed above), and 2 STD percent ratio errors ranged from 2.8% to >10% dependent on spatial location of the ADC measurement. Intra-site day-to-day repeatability across all sites was better than 5%, and average intra-exam repeatability (evaluated from several passes through the same exam protocol) within 1%. Observed b-value dependent variability at the phantom center was likewise less than 1%. Large ADC nonuniformity errors (5–15% bias) were present for off-center measurements consistent with gradient nonlinearity. Vendor cooperation is needed to develop viable instrumental correction procedures to control spatial ADC errors in clinical studies. The proposed protocol for standardization and quantitative quality control of ADC measurements is generally applicable for future multi-center clinical trials (8,9,13).

ACKNOWLEDGMENTS

REFERENCES

1. Hamstra DA, Lee KC, Moffat BA, Chenevert TL, Rehemtulla A, Ross BD. Diffusion magnetic resonance imaging: an imaging treatment response biomarker to chemoradiotherapy in a mouse model of squamous cell cancer of the head and neck. Transl Oncol 2008;1:187–194.
2. Galban CJ, Mukherji SK, Chenevert TL, et al. A feasibility study of parametric response map analysis of diffusion-weighted magnetic resonance imaging scans of head and neck cancer patients for providing early detection of therapeutic efficacy. Transl Oncol 2009;2:184–190.
3. Chenevert TL, Ross BD. Diffusion imaging for therapy response assessment of brain tumor. Neuroimaging Clin N Am 2009;19: 559–571.
4. Padhani AR, Liu G, Koh DM, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009;11:102–125.
5. Mulkern RV, Gudbjartsson H, Westin CF, et al. Multi-component apparent diffusion coefficients in human brain. NMR Biomed 1999;12:51–62.
6. Riches SF, Hawtin K, Charles-Edwards EM, de Souza NM. Diffusion-weighted imaging of the prostate and rectal wall: comparison of biexponential and monoexponential modelled diffusion and associated perfusion coefficients. NMR Biomed 2009;22:318–325.
7. Colagrande S, Pasquinelli F, Mazzoni LN, et al. MR-diffusion weighted imaging of healthy liver parenchyma: repeatability and reproducibility of apparent diffusion coefficient measurement. J Magn Reson Imaging 2010;31:912–920.
8. Tiepel SJ, Reuter S, Stieltjes B, et al. Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. Psychiatry Res 2011;194:363–371.
9. Sasaki M, Yamada K, Watanabe Y, et al. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology 2008;249:624–630.
10. Collins DJ, Blackledge M. Techniques and optimization. In: Koh DM, Thoeny HC, editors. Diffusion-weighted MR imaging: applications in the body. New York: Springer-Verlag; 2010. p 19–32.
11. Saritas EU, Lee JH, Nishimura DG. SNR dependence of optimal parameters for apparent diffusion coefficient measurements. IEEE Trans Med Imaging 2011;30:424–437.
12. Ogura A, Hayakawa K, Miyuati T, Maeda F. Imaging parameter effects in apparent diffusion coefficient determination of magnetic resonance imaging. Eur J Radiol 2011;77:185–188.
13. Zhu T, Hu R, Qiu X, Taylor M, et al. Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study. Neuroimage 2011;56: 1398–1411.
14. Delakis I, Moore EM, Leach MO, De Wilde JP. Developing a quality control protocol for diffusion imaging on a clinical MRI system. Phys Med Biol 2004;49:1409–1422.
15. Chenevert TL, Craig JG, Ivancevic MK, et al. Diffusion coefficient measurement using temperature controlled fluid for quality control in multi-center studies. J Magn Reson Imaging 2011;34: 983–987.
16. Tofts PS, Lloyd D, Clark CA, et al. Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. Magn Reson Med 2000;43:368–374.

17. Spees WM, Song S-K, Garbow JR, Neil JJ, Ackerman JH. Use of ethylene glycol to evaluate gradient performance in gradient-intensive diffusion MR sequences. Magn Reson Med 2012;68:319–324.

18. Matsuya R, Kuroda M, Matsumoto Y, et al. A new phantom using polyethylene glycol as an apparent diffusion coefficient standard for MR imaging. Int J Oncol 2009;35:893–900.

19. Krynicki K, Green CD, Sawyer DW. Pressure and temperature dependence of self-diffusion in water. Faraday Discuss Chem Soc 1978;66:199–208.

20. Simpson JH, Carr HY. Diffusion and nuclear spin relaxation in water. Phys Rev 1958;111:1201–1202.

21. Holz M, Heil SR, Sacco A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. Phys Chem Chem Phys 2000;2:4740–4742.

22. Meier C, Zwanger M, Feiweier T, Porter D. Concomitant field terms for asymmetric gradient coils: consequences for diffusion, flow, and echo-planar imaging. Magn Reson Med 2008;60: 128–134.

23. 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.