Repeatability of Quantitative Diffusion-Weighted Imaging Metrics in Phantoms, Head-and-Neck and Thyroid Cancers: Preliminary Findings

Ramesh Paudyal[1], Amaresha Shridhar Konar[1], Nancy A. Obuchowski[2], Vaios Hatzoglou[3], Thomas L. Chenevert[4], Dariya I. Malyarenko[4], Scott D. Swanson[4], Eve LoCastro[1], Sachin Jambawalikar[5], Michael Z. Liu[5], Lawrence H. Schwartz[5], R. Michael Tuttle[6], Nancy Lee[7], and Amita Shukla-Dave[1,3]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY; [2]Department of Quantitative Health Sciences, Cleveland Clinic Foundation, Cleveland, OH; [3]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY; [4]Department of Radiology, University of Michigan, Ann Arbor, MI; [5]Department of Radiology, Columbia University Irving Medical Center, and New York Presbyterian Hospital, New York, NY; [6]Departments of Medicine, and [7]Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY

Corresponding Author:
Amita Shukla-Dave, PhD
Departments of Medical Physics and Radiology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, NY, NY 10065;
E-mail: davea@mskcc.org

Key Words: quantitative Imaging, repeatability, diffusion-weighted imaging, head and neck cancer, thyroid, cancer, within-subject coefficient of variation

Abbreviations: Diffusion-weighted imaging (DWI), head and neck (HN), apparent diffusion coefficient (ADC), isotropic diffusion kurtosis imaging (iDKI), papillary thyroid cancer (PTC), head and neck squamous cell carcinoma (HNSCC), quantitative imaging biomarker (QIB), intra-voxel incoherent motion (IVIM), ceteryl alcohol and behentrimonium (CA-BTAC), repetition time (TR), echo time (TE), number of averages (NA), number of slices (NS), single-shot spin-echo echo planar imaging (SS-SE-EPI), reduced field of view (rFOV), regions of interest (ROIs), confidence interval (CI)

ABSTRACT

The aim of this study was to establish the repeatability measures of quantitative Gaussian and non-Gaussian diffusion metrics using diffusion-weighted imaging (DWI) data from phantoms and patients with head-and-neck and papillary thyroid cancers. The Quantitative Imaging Biomarker Alliance (QIBA) DWI phantom and a novel isotropic diffusion kurtosis imaging phantom were scanned at 3 different sites, on 1.5T and 3T magnetic resonance imaging systems, using standardized multiple b-value DWI acquisition protocol. In the clinical component of this study, a total of 60 multiple b-value DWI data sets were analyzed for test–retest, obtained from 14 patients (9 head-and-neck squamous cell carcinoma and 5 papillary thyroid cancers). Repeatability of quantitative DWI measurements was assessed by within-subject coefficient of variation (wCV%) and Bland–Altman analysis. In isotropic diffusion kurtosis imaging phantom vial with 2% ceteryl alcohol and behentrimonium chloride solution, the mean apparent diffusion ($D_{app} \times 10^{-3}$ mm^2/s) and kurtosis (K_{app}, unitless) coefficient values were 1.02 and 1.68 respectively, capturing in vivo tumor cellularity and tissue microstructure. For the same vial, D_{app} and K_{app} mean wCVs (%) were ≤1.41% and ≤0.43% for 1.5T and 3T across 3 sites. For pretreatment head-and-neck squamous cell carcinoma, apparent diffusion coefficient, D, D*, K, and f mean wCVs (%) were 2.38%, 3.55%, 3.88%, 8.0%, and 9.92%, respectively; wCVs exhibited a higher trend for papillary thyroid cancers. Knowledge of technical precision and bias of quantitative imaging metrics enables investigators to properly design and power clinical trials and better discern between measurement variability versus biological change.

INTRODUCTION

Malignant tumors of the head and neck (HN) region include a diverse group of cancers in the oral cavity, nasopharynx, oropharynx, hypopharynx, larynx, and paranasal sinuses; although salivary and thyroid carcinomas are also located within the HN region, they are typically thought of as separate tumors (1). HN tumors are heterogeneous with complex anatomy ranging between oral cavity to hypopharynx (2, 3). Accurate detection and delineation of tumor extent is critical to optimize treatment planning; patients therefore routinely undergo noninvasive imaging for careful assessment of this complex anatomy by an experienced neuroradiologist (4). Noninvasive magnetic resonance imaging (MRI) has served an important role as a diagnostic test for initial staging and follow-up of tumors in the HN region (5–8).

The quantitative MRI (qMRI) technique, diffusion-weighted imaging (DWI), assesses the Brownian motion of water mole-

ISSN 2379-1381 http://dx.doi.org/10.18383/j.tom.2018.00044

cules at a cellular level (9). Apparent diffusion coefficient (ADC), derived by fitting DWI data to a monoexponential model using ≥2 b-values (ie, diffusion-weighting factor), reflects tumor cellularity (10, 11). Repeatability of ADC has been tested in both phantoms and solid tumors (12-15). In previous studies, ADC has exhibited promise as a quantitative imaging biomarker (QIB) of treatment response in HN cancer (16-20). The use of ADC is helpful in differentiation between malignant and benign solitary thyroid nodules and assessing tumor aggressiveness in papillary thyroid cancer (PTC) (21, 22).

Recent literature reflects interest in acquisition of DWI data using multiple b-values, which allows the measurement of both water diffusion for higher b-values (>200 s/mm^2) and vascular perfusion fraction at lower b-values separately without contrast agent injection (23, 24). Le Bihan et al. developed a biexponential model using multiple b-value DWI data and termed it "intravoxel incoherent motion" (IVIM) (25, 26), which has shown utility for the assessment of treatment response in various cancers, including HN cancer (27, 28). Test–retest studies using IVIM-DWI metrics in normal liver and metastases have a tendency towards better repeatability of measurement of true diffusion coefficient (D), whereas use of perfusion fraction (f) and pseudo-diffusion coefficient (D*) are still exploratory in nature (23, 29).

Underlying biological structures can alter the Gaussian distribution of the water diffusion as assumed in IVIM to be non-Gaussian (NG) in nature (30). This NG behavior has been incorporated in the non-monoexponential diffusion kurtosis imaging (DKI) model which provides the kurtosis coefficient (K) metric, a surrogate QIB of tissue microstructure, in addition to diffusion coefficient (31-33). Lu et al. incorporated the NG diffusion into the IVIM-DWI model (NG IVIM-DWI) and provided estimates for all the aforementioned quantitative imaging metrics (f, D, D*, and K) (34).

QIBs are being used in oncology clinical trials to monitor the effects of treatments, identify subjects likely to benefit from treatment, and as trial endpoints. As compared with other modalities and endpoints, QIBs have the advantage of being noninvasive and requiring little or no subjective interpretation. Furthermore, for disease conditions with multiple treatment options, early detection of nonresponders enables physicians to consult patients about other treatment options earlier, to potentially improve outcomes and limit adverse effects of ineffective treatments.

Before QIBs can be used in clinical trials, their technical performance must be assessed, similarly to how sensitivity and specificity must be established for diagnostic tests (35). Technical performance includes precision, bias, and the property of linearity. Perhaps the most important QIB performance metric is precision, that is, the ability to provide the same, or nearly the same, measurement value on repeated observations (36). Once precision and performance metrics are established, they may be used to formulate a clinical trial's eligibility criteria, to determine the cut-point for defining true change over time, and to compute the sample size required for the trial (37).

There is currently a paucity of repeatability literature for DWI measurements in the clinical setting, particularly for HN cancers and PTC. Hence, it is critical to perform test–retest studies as the fundamental building blocks for QIB discovery and clinical application of these more advanced quantitative imaging methods. The objective of this study was to establish the repeatability measures of quantitative Gaussian and NG diffusion metrics using data from phantoms and from patients with HN cancers and PTC.

MATERIALS AND METHODS

Quantitative DWI Phantom

The quantitative diffusion phantom (High Precision Devices, Inc, Boulder, CO) developed by National Institute of Standards and Technology (NIST)/Radiological Society of North America (RSNA)-Quantitative Imaging Biomarker Alliance (QIBA) consists of 13 vials filled with varying concentrations of polyvinylpyrrolidone (PVP) in aqueous solution (38). The phantom was specifically designed for quantitatively mapping isotropic Gaussian diffusion of water molecules and generating physiologically relevant ADC values. The distribution of PVP concentrations in the phantom is as follows: 0% (vials 1–3), 10% (vials 4–5), 20% (vials 6–7), 30% (vials 8–9), 40% (vials 10–11), and 50% (vials 12–13). The space between the vials within the phantom was filled with an ice-water bath at 0°C to eliminate thermal variability across scanner locations and timepoints in ADC measurements. In this study, we will focus on the measurements obtained from 2 vials, that is, (1) water-only and (2) PVP-20%, as they relate to data from the novel isotropic diffusion kurtosis imaging (iDKI) phantom. Details of the NIST/QIBA DWI phantom have been published previously (38, 39).

The newly developed iDKI phantom used in this study was designed and fabricated by coauthors at the University of Michigan (40). The phantom captures a range of in vivo kurtosis values (K_{app} ranges, 0.4–1.7) (31). Here we report data from 2

Table 1. Summary of Patient Characteristics

Patient	Age (years)	Gender	Primary Cancer
1	63	M	BOT
2	58	M	NPC
3	59	M	Tonsil
4	59	M	Tonsil
5	60	M	BOT
6	68	F	BOT
7	61	F	Hypopharynx
8	75	M	BOT
9	55	M	BOT
10	51	M	PTC
11	44	M	PTC
12	44	M	PTC
13	48	M	PTC
14	44	F	PTC

Abbreviations: BOT, base of tongue; NPC, nasopharyngeal carcinoma; PTC, papillary thyroid cancer.

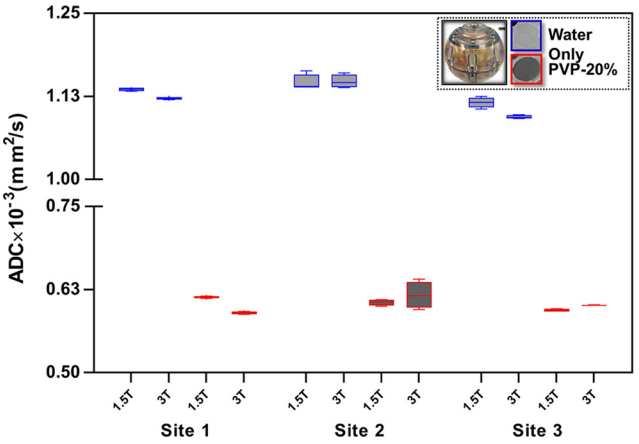

Figure 1. Box-and-whisker plot showing the test–retest mean apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) values obtained from National Institute of Standards and Technology (NIST)/Quantitative Imaging Biomarker Alliance (QIBA) polyvinylpyrrolidone (PVP) diffusion phantom (at 0°C) from the 3 different sites at 1.5T and 3T. The horizontal line inside the box indicates median values. The bottom and top of the boxes indicate 25th and 75th percentiles of the values, respectively. The differences between median values across scanners reflect the differences in gradient designs.

vials in the iDKI phantom: 1 vial containing chemical ceteryl alcohol and behentrimonium (CA-BTAC), a vesicular suspension formed by water solution of 2% CA-BTAC with other (minor) stabilizing ingredients (vial #2 [V2]), and a negative control consisting of a 20% solution of PVP in water (vial #4 [V4],), similar to the vial in NIST/QIBA DWI phantom (41). The iDKI phantom has been detailed in the poster presented at the NCI/Quantitative Imaging Network (QIN) meeting (40), and its full repeatability and long-term stability study is summarized in a research paper by Malyarenko D et al. submitted to this issue of *Tomography*.

The above 2 phantoms were studied to assess the technical performance of the quantitative imaging metrics among the 3 participating sites. There was a need to compare the vials with similar chemical composition for both the standard NIST/QIBA DWI and novel iDKI phantoms to emphasize the differences between the quantitative imaging metrics values for both diffusion and kurtosis coefficients.

Patient Cohort

The institutional review board of Site 1 (Memorial Sloan Kettering Cancer Center [MSKCC]) approved this prospective study for patients with head and neck squamous cell carcinoma (HNSCC) and PTC and was compliant with the Health Insurance Portability and Accountability Act. We obtained written informed consent from all eligible patients. A total of 14 patients were enrolled in the study between December 2016 and August 2017. In total, 30 MRI examinations were performed for these 14 patients, which comprised 60 test–retest MRI data sets. Nine patients with HNSCC were enrolled. All subjects had with metastatic nodes (M/F: 7/2, mean age: 59 years, range = 55–68 years) and underwent standard chemoradiation therapy (dose, 70 Gy). MRI examinations were performed before initiation of the standard chemoradiation treatment (pre-TX) and during treatment (intra-TX weeks 1 and 2) for patients with HNSCC. One patient with pre-TX MRI did not participate in MRI examinations during treatment. Five patients with PTC who underwent surgery (M/F: 4/4, mean age: 47 years, range = 37–61 years) were studied. All patient characteristics are summarized in Table 1.

DWI Data Acquisition

Quantitative DWI Phantom. Diffusion studies were performed using the NIST/QIBA DWI phantom at 0°C on 1.5T and 3T scanners using a 16-channel head coil at all 3 sites (Site 1 [MSKCC], Site 2 [Columbia University Irving Cancer Center; CUMC] and Site 3 [University of Michigan; UMich]). Localizer images were acquired for accurate positioning of the phantom. DWI images were acquired using a single-shot echo planar imaging sequence with 4 b-values (ie, b = 0, 500, 900, 2000 s/mm^2) and the following parameters: repetition time (TR) = 15 000 milliseconds, echo time (TE) = minimum (109–110 milliseconds), number of averages (NA) = 1, acquisition matrix = 128 × 128, field of view (FOV) = 220 mm, number of slices (NS) = 36, slice thickness = 4 mm, all 3 orthogonal directions at both 1.5T and 3.0T scanners. The total acquisition time for the multiple b-value DWI data acquisition was ~2 minutes 30 seconds.

The iDKI phantom, designed and fabricated by Site 3 (UMich), was imaged by all 3 sites at different field strengths of 1.5T

Table 2. Test–Retest Repeatability Measurement of the ADC for NIST/QIBA Phantom

Metrics	Chemical (PVP) Composition	Site 1		Site 2		Site 3	
		1.5T	3T	1.5T	3T	1.5T	3T
ADC × 10^{-3} mm^2/s	0%	1.13 ± 0.008	1.12 ± 0.002	1.14 ± 0.012	1.14 ± 0.01	1.11 ± 0.007	1.09 ± 0.002
	20%	0.61 ± 0.007	0.59 ± 0.005	0.60 ± 0.004	0.61 ± 0.02	0.59 ± 0.003	0.60 ± 0.004
wCV (%)	0%	0.21 (±0.48)	0.15 (±0.34)	1.07 (±2.41)	0.84 (±1.90)	0.67 (±1.48)	0.22 (±0.49)
	20%	0.24 (±0.09)	0.32 (±0.37)	0.71 (±0.85)	3.19 (±3.86)	0.33 (±0.39)	0.10 (±0.11)

wCV data in parentheses are lower and upper 95% confidence intervals.

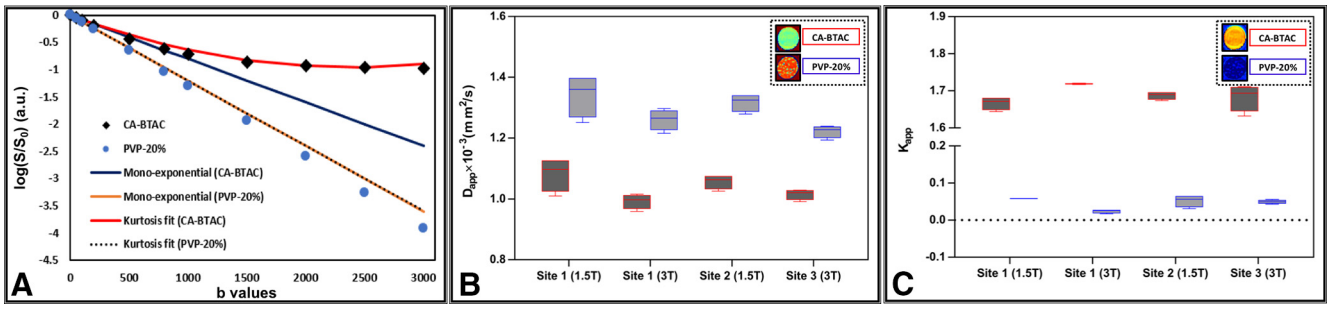

Figure 2. Representative DWI mean signal intensity decay curve vs. b-value obtained from vials of ceteryl alcohol and behentrimonium (CA-BTAC) and PVP-20% in iDKI phantom (scanned at ambient temperature) (A). The diamonds (black) and circles (blue) represent the experimental data, the monoexponential fit is represented by solid blue and yellow lines, and the solid red and dotted black lines are the fitted curves for the diffusion kurtosis model. Box-and-whisker plots show the test–retest for mean values of diffusion coefficient ($D_{app} \times 10^{-3}$ mm^2/s) (B), and kurtosis coefficient (K_{app}, no unit) for the iDKI phantom (C). The horizontal line inside the box indicates median values. The bottom and top of the boxes indicate 25th and 75th percentiles of the values, respectively. The differences between median values across scanners reflect both different scanner room temperatures and system gradient designs.

and/or 3T MRI scanners using a 16-channel head coil at ambient temperature. Localizer images were acquired for accurate positioning of the phantom. DWI images were acquired using a single-shot spin-echo echo planar imaging (SS-SE-EPI) sequence with 11 b-values (ie, b = 0, 50, 100, 200, 500, 800, 1000, 1500, 2000, 2500, 3000 s/mm^2) and parameters on both 1.5T and 3T scanners were kept similar as follows: TR = 10 000 milliseconds, TE = minimum (93–107 milliseconds), NA = 1, matrix = 128 × 128, FOV = 220 mm, NS = 5, slice thickness = 5 mm, all 3 orthogonal directions. The total acquisition time for the multiple b-value DWI data acquisition was ~5 minutes 20 seconds.

Four repeatability experiments for the NIST/QIBA DWI phantom in the study and 2 test–retests for iDKI phantoms with physical repositioning of the phantoms after each diffusion acquisition were performed.

Patient Cohort. MRI examinations were performed at Site 1 for patients with HNSCC on a Philips 3T MRI scanner (Ingenia, Philips Healthcare, The Netherlands) with a neurovascular phased-array coil (maximum number of channels: 20). Standard T1W and T2W imaging was followed by a multiple b-value DWI sequence (28). The DWI data were acquired using a SS-SE-EPI sequence with 10 b-values (ie, b = 0, 20, 50, 80, 200, 300, 500, 800, 1500, 2000 s/mm^2) with TR = 4000 milliseconds, TE = 80 (minimum) milliseconds, NA = 2, matrix = 128 × 128, FOV = 200–240 mm, NS = 8–10, and slice thickness = 5 mm. For patients with HNSCC, DWI was acquired with full field of view as part of the standard clinical imaging protocol. The total acquisition time for the multiple b-value DWI data acquisition was ~5 min. Two multi b-value DWI data sets were acquired at the same MR examination for each patient with HNSCC to test for the repeatability of the derived quantitative imaging metrics. Eighteen multiple b-value DWI data set were acquired at pre-TX (week 0). In addition, 32 multiple b-value DWI data sets were acquired at intra-TX week 1 and week 2 (during chemoradiation therapy). A total of 50 multiple b-value DWI examinations (pre-TX [9 patients], intra-TX week 1 [8 patients], and intra-TX week 2 [8 patients]) were performed (2 MR examinations at each session). As a note, these DWI data sets were acquired with full FOV (phase FOV factor = 1.0).

Table 3. Test–Retest Repeatability Measurement of the D_{app} and K_{app} for Isotropic Diffusion Kurtosis Phantom

Metrics	Chemical Composition	1.5T (Site 1)	3T (Site 1)	1.5T (Site 2)	3T (Site 3)
$D_{app} \times 10^{-3}$ mm^2/s	CA-BTAC	1.06 ± 0.08	0.99 ± 0.029	1.05 ± 0.034	1.01 ± 0.021
	PVP20%	1.32 ± 0.10	1.26 ± 0.041	1.30 ± 0.043	1.22 ±0.024
wCV (%)	CA-BTAC	1.41 (±2.94)	1.18 (±2.32)	1.18 (±2.47)	0.70 (±1.41)
	PVP-20%	1.01 (±2.67)	0.63 (±1.58)	0.31 (±0.79)	0.84 (±1.97)
K_{app}	CA-BTAC	1.66 ± 0.026	1.71 ± 0.001	1.68 ± 0.015	1.68 ± 0.044
	PVP-20%	0.06 ± 0.003	0.03 ± 0.005	0.05 ± 0.023	0.05 ± 0.006
wCV (%)	CA-BTAC	0.35 (±1.17)	0.42 (±1.41)	0.36 (±1.20)	0.43 (±1.43)
	PVP-20%	19.35 (±2.21)	11.12 (±0.57)	7.13 (±0.64)	25.06 (±2.41)

wCV data in parentheses are lower and upper 95% confidence intervals.

MRI examinations were performed at Site 1 for patients with PTC (n = 5) on a 1.5T (n = 2) or 3T (n = 3) scanner (General Electric, Milwaukee, WI), with a neurovascular phased-array coil and consisted of standard T1W and T2W imaging scans followed by multiple b-value DWI data acquisition. This was a feasibility test for the MRI of patients with PTC, which was performed as part of an ongoing research imaging protocol. Data were acquired with reduced field of view (rFOV) DWI technique, using a 2-dimensional spatially selective excitation (42). The acquisition parameters of rFOV DWI scans with the SS-SE-EPI sequence were as follows: 10 b-values (ie, b = 0, 20, 50, 80, 200, 300, 500, 800, 1500, 2000 s/mm^2), TR = 4000 milliseconds, TE = 80 (minimum) milliseconds, NA = 2, matrix = 128 × 64, FOV = 200–240 cm, NS = 8–10, slice thickness = 5 mm, and phase FOV factor = 0.5. The total time for rFOV DWI data acquisition was ~5 min.

Repeatability measures were tested on the multiple b-value DWI data sets obtained from patients with HNSCC at pre-TX, and during intra-TX weeks 1 and 2 of standard chemoradiation therapy. Pretreatment DWI repeatability data were obtained for patients with PTC who underwent surgery.

DWI Data Analysis

All DWI data postprocessing and quantitative metrics map generation, detailed below, were performed using in-house–developed software entitled MRI-QAMPER (MRI Quantitative Analysis of Multi-Parametric Evaluation Routines). The MRI-QAMPER package includes the algorithm routines for DWI data analyses (ADC, diffusion kurtosis, IVIM, and NG-IVIM), implemented in MATLAB (The MathWorks, Natick, MA). The MRI-QAMPER tool is approved by National Cancer Institute/Quantitative Imaging Network (QIN) with pre-benchmark status, which facilitates its use by other QIN site colleagues for analysis of multiple b-value DWI data.

For NIST/QIBA DWI phantom data analysis, 3 distinct circular regions of interest (ROIs) were manually placed (9 mm in diameter) on the selected vials, with water only and PVP-20%, in ADC maps avoiding boundaries; the mean pixel value across the ROIs in each vial was used to measure repeatability.

For iDKI phantom data analysis, 2 distinct circular ROIs (12 mm in diameter, single-plane) were placed on vials with CA-BTAC solution and PVP-20% in the phantom images; the mean pixel value across the ROIs in each vial was used for the test–retest study. To guarantee model convergence, a bmax constraint value for fitting the kurtosis expression in the CA-BTAC phantom vial was set to 1500 s/mm^2 (bmax × D_{app} × K_{app} <3) (43).

For DWI patient data, ROIs were manually delineated on the DWI images (b = 0 s/mm^2) on the metastatic neck node in HNSCC, normal thyroid gland, and PTC. ROIs were placed on thyroid glands avoiding obvious cystic, hemorrhagic, or calcified portions, whereas for normal thyroid tissue, ROIs were placed on the selected contralateral side to the PTC. ROIs were contoured by an experienced neuroradiologist based on the clinical information and T1W/T2W images using ImageJ (44).

Multiple b-value DWI data sets were analyzed using the following models:

1. Mono-exponential (ADC): All b-value DWI signal intensity data obtained from each voxel in the ROI were fitted to a

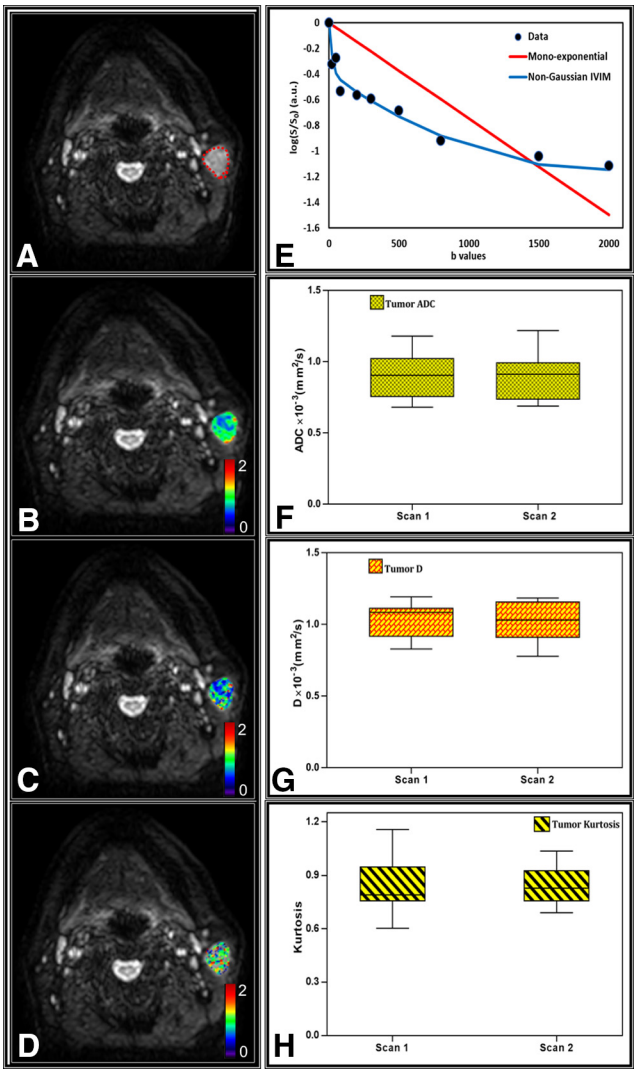

Figure 3. Representative intra-TX week 1 magnetic images (MR) images of a patient with head and neck squamous cell carcinoma (HNSCC) (76 years, male). Diffusion-weighted (b = 0 s/mm^2) image (A), apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) (B), diffusion coefficient (C) (D × 10^{-3} mm^2/s), and kurtosis metric maps overlaid on DWI (b = 0 s/mm^2) image (D). Representative plot of the logarithm of signal intensity vs. b-values (E). The circle (black) represents the experimental data, and the solid lines are the fitted curves with the monoexponential (red) and extended non-Gaussian (NG) intravoxel incoherent motion (NG-IVIM) model (blue). Box-and-whisker plot shows the mean value of (F) apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) and the NG-IVIM model derived metrics: (G) diffusion coefficient (D × 10^{-3} mm^2/s and (H) kurtosis coefficient ("K"). The bottom and top of the boxes indicate 25th and 75th percentiles of the values, respectively. The horizontal line inside the box indicates median values. Note: Data were acquired using standard full FOV DWI sequence.

monoexponential model to calculate ADC (mm^2/s) as follows (45):

$$S(b) = S_0 e^{-bADC} \quad (1)$$

where S(b) and S_0 are the signal intensities with and without diffusion weighting, and the quantity b is the diffusion-weighting factor (s/mm^2).

2. DKI: The signal intensity versus b-value DWI data were fitted to non-monoexponential diffusion kurtosis imaging model (DKI) of the following form (43):

$$S(b) = S_0\left[e^{-bD_{app}+\frac{1}{6}K_{app}b^2D^2}\right] \quad (2)$$

where D_{app} is the ADC (mm^2/s) and K_{app} (no unit) is a dimensionless apparent kurtosis coefficient. D_{app} and K_{app} are associated with the NG behavior of a signal in tissue. As a note, $K_{app} = 0$ is equivalent to equation (1).

3. NG-IVIM: The signal intensity versus b-value DWI signal were fitted to biexponential NG-IVIM DWI model as follows (34, 46):

$$S(b) = S_0\left[fe^{-bD^*} + (1-f)e^{-bD+\frac{1}{6}Kb^2D^2}\right] \quad (3)$$

Where D is the diffusion coefficient (mm^2/s), perfusion fraction (f), and D* is the pseudo-diffusion coefficient (mm^2/s), and K is the kurtosis coefficient.

The NIST/QIBA DWI phantom was analyzed using monoexponential diffusion model equation (1), the iDKI phantom using DKI model [equation (2)], and HNSCC (tumor), and PTC (tumor and normal) using DKI model [equation (2)] and extended NG-IVIM model [equation (3)]. Mean metric values of ADC, DKI-derived metrics (D_{app} and K_{app}), and NG-IVIM-derived metrics (D, D*, f, and K) calculated from each ROI were compared between repeated measurements.

Statistical Analysis

Technical precision of QIBs was evaluated based on the framework proposed by the RSNA/QIBA (https://www.rsna.org/uploadedFiles/RSNA/Content/Science_and_Education/QIBA/QIBA_Process_05Jan2015.pdf). The within-subject coefficient of variation (wCV, %) was used as the measure of precision; it was estimated from the phantom and clinical data as follows (22, 47-49):

$$wCV\,(\%) = \frac{\sigma_w}{\mu} \times 100 \quad (4)$$

where σ_w is the within-subject standard deviation and μ is the mean. A 95% confidence interval (CI) for the wCV was constructed using χ^2 as the pivotal statistic as follows:

$$CI\,(95\,\%) = \sqrt{\frac{N \times wCV^2}{\chi^2_{N,\alpha}}} \quad (5)$$

where N is the number of patients, each having 2 replicate observations and $\chi^2_{N,\alpha}$ is the αth percentile of the chi-square distribution with N degrees of freedom. For the lower bound, α is 0.975, and for the upper bound, α is 0.025. Bland–Altman plots were constructed to measure the repeatability of the quantitative imaging metrics.

Statistical analysis for the data was conducted in R (50) and MATLAB (The MathWorks, Inc., Natick, MA).

RESULTS

Quantitative DWI Phantom

Mean ADC values obtained from the NIST/QIBA DWI phantom (scanned at 0°C) at all 3 different sites on 1.5T and 3T MRI scanners are displayed in a box-and-whisker plot (Figure 1). ADC values are reported for 2 vials only (water-only and PVP-20%). The mean wCV (%) for vial with water-only were ≤1.07% and ≤0.84% and that for vial with PVP-20% were ≤0.71% and ≤3.19% at 1.5T and 3T MRI across the 3 sites, respectively. Results of ADC wCV and 95% CIs are summarized in Table 2.

Figure 2A shows the representative plot of the DWI logarithmic signal intensity versus b-value, fitted by both monoexponential and DKI models obtained from the iDKI phantom ROI for the vials with CA-BTAC (V2) and PVP-20% (V4). The box-and-whisker plots show the mean values of $D_{app} \times 10^{-3}$ mm^2/s (Figure 2B) and K_{app} (no unit) (Figure 2C) obtained from V2 (captures both in vivo tumor cellularity and tissue microstructure) and V4 (captures in vivo tumor cellularity but negative control for kurtosis).

The wCV (%) mean values of D_{app} and K_{app} for V2 were ≤1.41% and ≤0.43% on both 1.5T and 3T MRI. The wCV (%) mean values of D_{app} and K_{app} for V4 were ≤1.01% and ≤25.06% respectively, on both 1.5T and 3T MRI. Table 3 summarizes the

Table 4. Test–Retest Repeatability Measurement of Diffusion Kurtosis Model-Derived Metrics for Patients With HNSCC

Treatment	Measurement	D_{app}	K_{app}
Pre-TX	Mean	(1.54 ± 0.02) × 10^{-3} mm^2/s	(0.94 ± 0.01)
	wCV (%)	5.62 (3.87, 10.30)	5.18 (3.59, 9.47)
Intra-TX Week 1	Mean	(1.56 ± 0.02) × 10^{-3} mm^2/s	0.96 (±0.01)
	wCV (%)	2.99 (2.10, 5.72)	8.12 (3.50, 15.56)
Intra-TX Week 2	Mean	(1.68 ± 0.06) × 10^{-3} mm^2/s	(0.85 ± 0.01)
	wCV (%)	4.29 (2.90, 8.22)	6.01 (4.06, 11.51)

wCV data in parentheses are lower and upper 95% confidence intervals.

Table 5. Test–Retest Repeatability Measurement of the ADC- and NG-IVIM DWI-Derived Metrics for Patients With HNSCC

Treatment	Measurement	ADC	D	D*	K	f
Pre-TX	Mean	$(0.90 \pm 0.04) \times 10^{-3}$ mm^2/s	$(1.03 \pm 0.07) \times 10^{-3}$ mm^2/s	$(2.51 \pm 0.19) \times 10^{-3}$ mm^2/s	(0.84 ± 0.13)	(0 .19 ± 0.04)
	wCV (%)	2.38 (1.67, 4.34)	3.55 (2.44, 6.48)	3.88 (2.67, 7.10)	8.00 (5.57, 14.61)	9.92 (6.8, 18.12)
Intra-TX Week 1	Mean	$(0.93 \pm 0.02) \times 10^{-3}$ mm^2/s	$(1.09 \pm 0.07) \times 10^{-3}$ mm^2/s	$(2.46 \pm 0.11) \times 10^{-3}$ mm^2/s	(0.87 ± 0.08)	(0.18 ± 0.02)
	wCV (%)	0.86 (0.58, 1.66)	3.46 (2.39, 6.63)	2.24 (2.62, 4.28)	4.74 (5.40, 9.09)	9.92 (67.0, 10.96)
Intra-TX Week 2	Mean	$(0.96 \pm 0.04) \times 10^{-3}$ mm^2/s	$(1.16 \pm 0.08) \times 10^{-3}$ mm^2/s	$(2.47 \pm 0.19) \times 10^{-3}$ mm^2/s	(0.86 ± 0.14)	(0.19 ± 0.04)
	wCV (%)	1.18 (0.79, 2.26)	5.57 (3.76, 10.67)	1.20 (0.81, 2.28)	8.36 (5.64, 16.01)	3.01 (2.02, 5.74)

wCV data in parentheses are lower and upper 95% confidence intervals.

D_{app} and K_{app} mean wCV and 95% CIs values for vials with CA-BTAC and PVP-20%. The absolute $K_{app} < 0.05$ value observed for ROI in vial with PVP-20% samples indicates minor bias of the NG model for this monoexponential material.

Patient Cohort. The pre-TX tumor volume (mean ± SD) in patients with HNSSC and PTC were 9.13 ± 6.22 cm^3 and 0.35 ± 0.39 cm^3, respectively.

Figure 3, A–D shows a representative DWI (b = 0 s/mm^2) image, ADC × 10^{-3} mm^2/s, D × 10^{-3} mm^2/s, and K metric maps for a patient with HNSCC. Figure 3E depicts a representative logarithmic DWI signal as a function of the b-value obtained from the metastatic node of the HNSCC patient. The DWI signal was fitted to the monoexponential and NG IVIM model. Figure 3, F–H also displays the box-and-whisker plots for pre-TX test–retest mean values of the same quantitative imaging metrics detailed above.

The wCV (%) mean values of D_{app} and K_{app} at Pre-TX were 5.62% and 5.18%, respectively. Table 4 summarizes the mean wCV (%) and 95% CIs for D_{app} and K_{app} at pre-TX and intra-TX weeks in patients with HNSCC.

The mean wCV (%) values for pre-TX ADC, D, D*, K, and f were 2.38%, 3.55%, 3.88%, 8.0%, and 9.92%, respectively. Table 5 summarizes mean wCV (%) and 95% CIs for ADC- and NG-IVIM-derived metrics (D, D*, K, and f) at pre-TX and intra-TX weeks in patients with HNSCC.

Bland–Altman plots are shown for selective quantitative imaging metrics, ADC, D, and K, obtained from the pre-TX neck nodal metastases of patients with HNSCC (Figure 4). In each panel, the differences in mean values of ADC, D, and K were plotted between the repeated MRI examinations against the combined mean values of ADC, D, and K.

The results from patients with PTC are part of ongoing feasibility testing in the research setting for thyroid MRI imaging using rFOV multiple b-value DWI. Figure 5, A–D displays a representative DWI (b = 0 s/mm^2) image, ADC × 10^{-3} mm^2/s, D × 10^{-3} mm^2/s, and K metric maps for a patient with PTC. Figure 5E shows a representative logarithmic DWI signal as a function of the b-value obtained from the normal thyroid tissue and tumor of the patient with PTC.

The wCV (%) mean values of D_{app} and K_{app} for normal tissue were 12.87% and 17.46%, respectively, whereas these metric values in tumor tissue were 22.42% and 25.94% in patients with PTC. Table 6 summarizes mean D_{app} and K_{app} wCV (%) and 95% CIs for normal and tumor region in patients with PTC.

ADC mean wCV (%) were 11.86% and 10.04%, respectively, for tumor and normal thyroid tissue ROIs. The wCV (%) for NG IVIM-derived metrics (D, D*, K, and f) from tumors were 14.98%, 4.31%, 11.09%, and 13.31%, respectively. Preliminary mean values for ADC, D, D*, K, and f are summarized in Table 7 for normal and tumor tissue in patients with PTC.

Bland–Altman plots are shown for ADC, D, and K, obtained from normal and tumor regions in the PTC patients (Figure 6).

DISCUSSION

In this preliminary study, we measured the repeatability of the quantitative diffusion imaging metrics for Gaussian and NG models using 2 phantoms (the temperature-controlled NIST/QIBA DWI phantom and a novel iDKI phantom at ambient

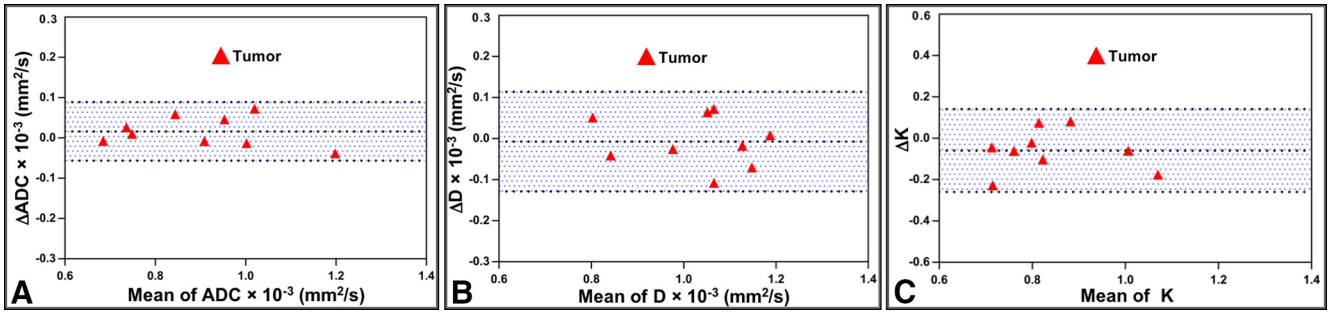

Figure 4. Bland–Altman plots of apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) (A), diffusion coefficient (D × 10^{-3} mm^2/s) (B), and kurtosis coefficient (K) obtained from the metastatic neck node in patients with HNSCC on pre-treatment (C). The solid lines correspond to the mean differences between 2 estimates and the dashed lines show the 95% limits of agreement. Note: Δ represent the change in mean difference between 2 scans.

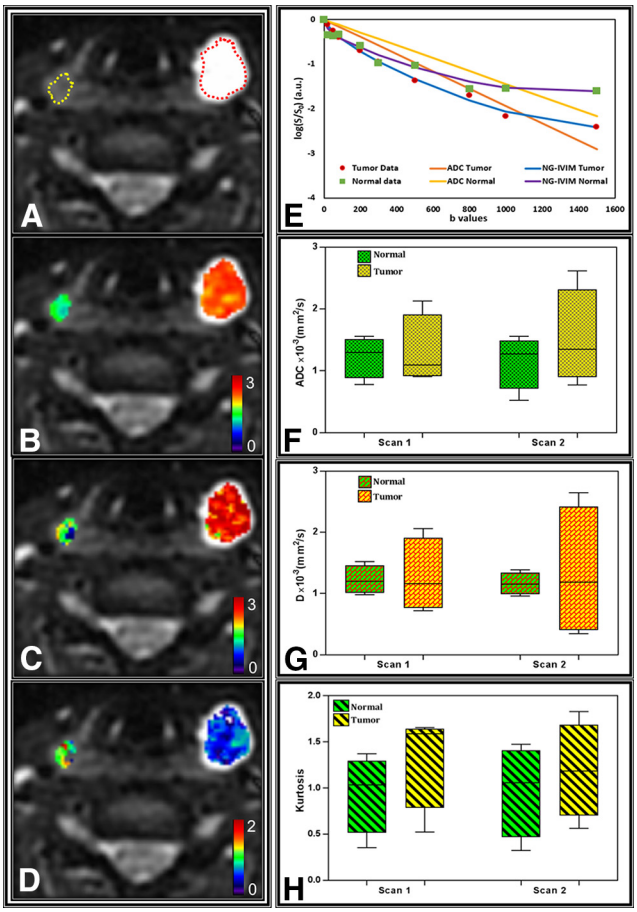

Figure 5. Representative MR images of a patient with papillary thyroid cancer (PTC) (76 years, male). Diffusion-weighted (b = 0 s/mm^2) image (A), apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) (B), diffusion coefficient (D × 10^{-3} mm^2/s) (C), kurtosis metric maps overlaid on DW (b = 0 s/mm^2) images (D), representative plot of the logarithm of DWI signal intensity vs. b values (E). The squares (green) and circles (red) represent the experimental data in normal and tumor; the solid lines are the fitted curves with the monoexponential (yellow and orange) and NG intravoxel incoherent motion (NG-IVIM) (purple and blue). Box-and-whisker plot shows the mean value in normal tissue and in tumor for (F) apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) and the NG-IVIM model-derived metrics: diffusion coefficient (D × 10^{-3} mm^2/s) (G) and kurtosis coefficient (K, unitless) (H). The horizontal line inside the box indicates median values. The bottom and top of the boxes indicate 25th and 75th percentiles of the values, respectively. Note: The DWI images were acquired with reduced FOV DWI sequence.

Table 6. Test–Retest Repeatability Measurement of Diffusion Kurtosis Model-Derived Metrics for Patients With PTC

Treatment	Measurement	D_{app}	K_{app}
Normal	Mean	(2.51 ± 0.32) × 10^{-3} mm^2/s	(1.08 ± 0.19)
	wCV (%)	12.87 (7.71, 37.00)	17.46 (10.46, 50.19)
Tumor	Mean	(2.52 ± 0.57) × 10^{-3} mm^2/s	(1.14 ± 0.29)
	wCV (%)	22.42 (13.43, 64.46)	25.94 (15.54, 74.57)

wCV data in parentheses are lower and upper 95% confidence intervals.

temperature) in a multisite setting, as well as for a small cohort of patients with HNSCC and PTC using the DKI model and the extended NG IVIM model.

For the NIST/QIBA DWI phantom, repeatability of mean ADC wCV (%) and 95% CIs values was excellent for the studied phantom vials with water-only and PVP-20% (≤3.19% and ≤4.0% respectively), for all 3 sites. The results reported herein are comparable to results from similar test–retest repeatability studies (51, 52). The D_{app} and K_{app} wCVs (%) and 95% CIs from all 3 sites were comparable at both 1.5T and 3T MRI. The novel iDKI phantom has been designed and fabricated with the purpose to better understand the performance of the quantitative diffusion metric kurtosis (K) as a surrogate of tissue microstructure and the stability of K over time. Performing appropriate phantom testing is a prerequisite for the QIB pipeline for clinical trials that use quantitative NG diffusion imaging metrics (37). Our phantom results confirmed adequate baseline technical performance of the MRI scanner systems and multiple b-value DWI protocols used for the quantitative DWI studies for patients with HNSCC and PTC.

There is currently paucity of repeatability measures for quantitative Gaussian and NG DWI in cancers of the HN region, despite availability of ADC test–retest data for organs such as brain (wCV = 3.97%), liver (wCV = 9.38%), and prostate (wCV = 16.97%) (13-15, 53). Only a few studies have reported test–retest data for IVIM in organs such as liver (23).

The preliminary findings for test–retest data in HNSCC showed that the mean wCV (%) for ADC-derived metric, DKI-derived metric (D_{app}), and NG IVIM-derived metrics (D and D*) were ≤6% for pre-TX, intra-TX weeks 1 and 2. For f, K_{app}, and K, the mean wCV(%) were ≤10%. Both ADC and D, quantitative imaging metrics, are surrogate biomarkers of tumor cellularity, while f and D* are still exploratory in nature (23, 29, 54). There is keen interest in furthering description of tissue microstructure using the quantitative imaging metric K (31, 34, 55, 56). The uncertainties from clinical HNSCC data slightly exceeded baseline repeatability achieved for phantoms due to additional patient-related variability.

Our clinical repeatability measurements for normal thyroid tissue and PTC are preliminary findings. Lu et al. reported that the ADC mean wCV (%) for the normal thyroid tissue in healthy volunteers is ≤10% using rFOV DWI at 3T (42). The present study found consistent results for the normal thyroid tissue (ADC mean wCV (%) = ≤10%) acquired with rFOV DWI at 1.5T and 3T MRI

Table 7. Test–Retest Repeatability Measurement of the ADC and NG-IVIM DWI-Derived Metrics for Patients With PTC

ROI	Measurement	ADC	D	D*	K	f
Normal	Mean	(1.23 ± 0.24) ×10^{-3} mm^2/s	(1.16 ± 0.58) ×10^{-3} mm^2/s	(2.89 ± 0.43) ×10^{-3} mm^2/s	(0.96 ± 0.36)	(0.26 ± 0.08)
	wCV (%)	10.05 (6.02, 28.90)	25.80 (15.46, 74.17)	7.63 (4.57, 21.94)	19.25 (11.53, 55.34)	16.54 (9.91, 47.57)
Tumor	Mean	(1.31 ± 0.30) ×10^{-3} mm^2/s	(1.54 ± 0.45) ×10^{-3} mm^2/s	(2.87 ± 0.24) ×10^{-3} mm^2/s	(1.21 ± 0.26)	(0.22 ± 0.06)
	wCV (%)	11.86 (7.11, 34.10)	14.98 (8.97, 43.06)	4.31 (2.58, 12.38)	11.09 (6.65, 31.89)	13.31 (7.97, 38.26)

wCV data in parentheses are lower and upper 95% confidence intervals.

(42). Kim et al. reported that mean ADC values obtained at 2 different MRI field strengths (1.5T and 3T) were not significantly different (19). A relatively high wCV was observed for DKI- and NG IVIM-derived metrics that may likely be related to the limited sample size and the biology of the tumors in the thyroid gland.

Establishing the technical performance of a QIB allows us to better understand a patient's measurement at a single time point, especially the changes in measurements over time, by constructing a CI for the true value or the true change. For example, suppose we measure ADC of 1.22 × 10^{-3} mm^2/s for PTC. If we know from our technical performance studies that the measurements are made with negligible bias and precision of wCV (%) = 11.86%, then a 95% CI for the patient's true ADC value is (1.22 ± 1.96 × (0.1186 × 1.22) × 10^{-3} mm^2/s or (0.94 to 1.50) × 10^{-3} mm^2/s. The CI helps differentiate between the true change of the parameter value versus the measurement uncertainty. Now suppose that on a second visit, the patient's tumor has an ADC of 1.31 × 10^{-3} mm^2/s. Has the ADC value truly increased or is the observed change attributable to measurement error? The 95% CI for the true change is $[(1.31-1.22) \pm 1.96 \times \sqrt{(0.1186 \times 1.22)^2 + (0.1186 \times 1.31)^2}] \times 10^{-3}$ mm^2/s or [−0.32, 0.50] × 10^{-3} mm^2/s. Thus, given the known imprecision in the ADC measurements, we cannot conclude that a true change has occurred with 95% confidence.

Once the technical performance of a QIB is known, investigators are better able to design their clinical trials effectively. For example, a measured change in a patient's quantitative imaging metrics (eg, D or K) must exceed 10% (ie, 2.77 × wCV) to be 95% confident that a true change has occurred (37). Thus, in a drug trial using changes in D or K (QIBs) as a measure of therapeutic effect, a ≥10% cut-point should be used to define whether a treatment effect should be used to define treatment success and determine when a change in treatment is warranted. Similarly, in planning a clinical trial where D or K values will be compared across treatment arms, the imprecision in the QIB values affects trial sample size by increasing it relative to its magnitude and the magnitude of the between-subject variability.

There are a few known limitations to this study. This is the first feasibility test–retest study of Gaussian- and NG diffusion-derived metrics from multisite phantom and single-site clinical data testing. A larger cohort of patients (>30) is necessary to confirm statistical significance of the preliminary findings (57). Susceptibility artifacts caused by SS-SE-EPI, voluntary and involuntary bulk motion, are still an issue in the HN region, limiting repeatability. Thus, rFOV DWI for incremental improvement may be an option, exciting only a limited FOV and not surrounding regions that potentially cause interference (42). For the test–retest data set, technically the patients should be scanned, removed from the scanner for a few minutes and scanned again, referred to as a "coffee break" study. Here, the patients were repositioned between scans on the table but not removed from the scanner ("coffee break") owing to practical reasons relating to patient comfort and workflow at the MR scanner. The results reported here provide insights into what is needed and must be paid attention to in test–retest studies in clinical oncology trials. For example, the test–retest studies for ADC in brain tumors derived from monoexponential mod-

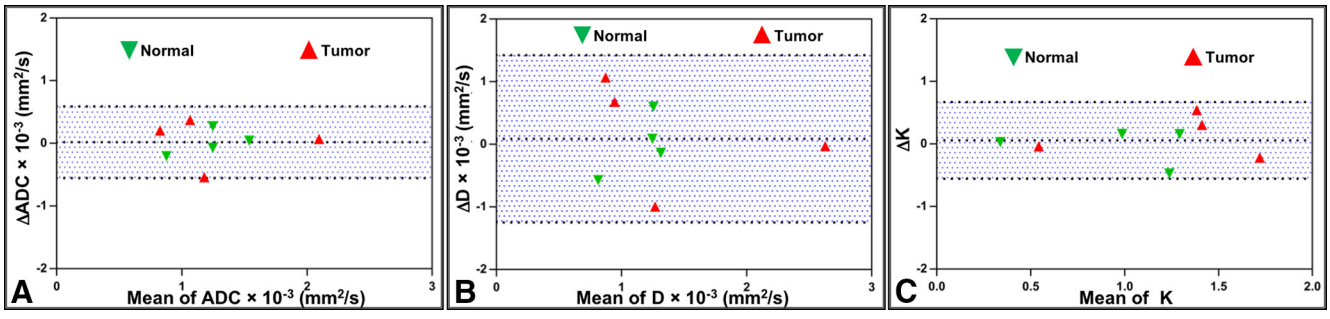

Figure 6. Bland–Altman plots of apparent diffusion coefficient (ADC × 10^{-3} mm^2/s) (A), diffusion coefficient (D × 10^{-3} mm^2/s) (B), and kurtosis coefficient (K) obtained from the papillary thyroid cancer (C). The solid lines correspond to the mean differences between 2 estimates, and the dashed lines show the 95% limits of agreement. Note: Δ represent the change in mean difference between 2 scans.

eling of DWI data reports a wCV of 3.97% (53, 58, 59). A smaller wCV value (<5%) indicates less variation in repeatability measurements.

CONCLUSION

In conclusion, we have shown repeatability of measurements for quantitative Gaussian and NG diffusion imaging metrics using multiple b-value acquisitions for NIST/QIBA DWI phantom and iDKI phantom, across multisite MRI systems, and used in HNSCC and PTC clinical trials. The preliminary results for the repeatability measurement of NG IVIM-derived metrics in HNSCC and PTC show promise and need additional validation with a larger subject cohort. In short, the precision of QIBs must be established for oncology clinical trials to noninvasively monitor the effects of treatment, to identify subjects likely to benefit from treatment and define trial endpoints.

ACKNOWLEDGMENTS

We acknowledge funding from NIH U01 CA211205 (ASD, LHS) and NIH/NCI Cancer Center Support Grant P30 CA008748, and U01CA166104 (TLC, DIM, SDS).

Disclosure: S.D. Swanson, D.I. Malyarenko, and T.L. Chenevert are coinventors on intellectual property assigned to and managed by the University of Michigan for the technology underlying the manufacturing of the quantitative isotropic diffusion kurtosis imaging (iDKI) phantoms.

Conflict of Interest: The authors have no conflict of interest to declare.

REFERENCES

1. Vogel DWT, Thoeny HC. Cross-sectional imaging in cancers of the head and neck: how we review and report. Cancer Imaging. 2016;16:20.
2. King AD, Vlantis AC, Tsang RK, Gary TM, Au AK, Chan CY, Kok SY, Kwok WT, Lui HK, Ahuja AT. Magnetic resonance imaging for the detection of nasopharyngeal carcinoma. AJNR Am J Neuroradiol. 2006;27:1288–1291.
3. Klussmann JP. Head and neck cancer—new insights into a heterogeneous disease. Oncol Res Treat. 2017;40:318–319.
4. Vairaktaris E, Yapijakis C, Psyrri A, Spyridonidou S, Yannopoulos A, Lazaris A, Vassiliou S, Ferekidis E, Vylliotis A, Nkenke E, Patsouris E. Loss of tumour suppressor p16 expression in initial stages of oral oncogenesis. Anticancer Res. 2007;27:979–984.
5. King AD, Mo FK, Yu KH, Yeung DK, Zhou H, Bhatia KS, Tse GM, Vlantis AC, Wong JK, Ahuja AT. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010;20:2213–2220.
6. Lenz M, Greess H, Baum U, Dobritz M, Kersting-Sommerhoff B. Oropharynx, oral cavity, floor of the mouth: CT and MRI. Eur J Radiol. 2000;33:203–215.
7. Zima AJ, Wesolowski JR, Ibrahim M, Lassig AA, Lassig J, Mukherji SK. Magnetic resonance imaging of oropharyngeal cancer. Top Magn Reson Imaging. 2007;18:237–242.
8. Noda Y, Kanematsu M, Goshima S, Kondo H, Watanabe H, Kawada H, Bae KT. MRI of the thyroid for differential diagnosis of benign thyroid nodules and papillary carcinomas. Am J Roentgenol. 2015;204:W332–W335.
9. Le Bihan D. From Brownian motion to mind imaging: diffusion MRI. Bull Acad Natl Med. 2006;190:1605–1627; discussion 1627. [Article in French
10. Chenevert TL, McKeever PE, Ross BD. Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging. Clin Cancer Res. 1997;3:1457–1466.
11. Partridge SC, Nissan N, Rahbar H, Kitsch AE, Sigmund EE. Diffusion-weighted breast MRI: clinical applications and emerging techniques. J Magn Reson Imaging. 2017;45:337–355.
12. Boss M, Chenevert T, Waterton J, Morris D, Ragheb H, Jackson A, de Souza N, Collins DJ, van Beers BV, Garteiser P, Doblas S, Persigehl T, Hedderich D, Martin A, Mukherjee P, Keenan K, Russek S, Jackson E, Zahlmann G. Thermally-stabilized isotropic diffusion phantom for multisite assessment of apparent diffusion coefficient reproducibilty. Med Phys. 2014;41:464.
13. Braithwaite AC, Dale BM, Boll DT, Merkle EM. Short- and midterm reproducibility of apparent diffusion coefficient measurements at 3.0-T diffusion-weighted imaging of the abdomen. Radiology. 2009;250:459–465.
14. Gibbs P, Pickles MD, Turnbull LW. Repeatability of echo-planar-based diffusion measurements of the human prostate at 3 T. Magn Reson Imaging. 2007;25:1423–1429.
15. Jambor I, Merisaari H, Aronen HJ, Jarvinen J, Saunavaara J, Kauko T, Borra R, Pesola M. Optimization of b-value distribution for biexponential diffusion-weighted MR imaging of normal prostate. J Magn Reson Imaging. 2014;39:1213–1222.
16. Vandecaveye V, De Keyzer F, Dirix P, Lambrecht M, Nuyts S, Hermans R. Applications of diffusion-weighted magnetic resonance imaging in head and neck squamous cell carcinoma. Neuroradiology. 2010;52:773–784.
17. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. J Magn Reson Imaging. 2010;32:2–16.
18. Dirix P, Vandecaveye V, De Keyzer F, Op de Beeck K, Poorten VV, Delaere P, Verbeken E, Hermans R, Nuyts S. Diffusion-weighted MRI for nodal staging of head and neck squamous cell carcinoma: impact on radiotherapy planning. Int J Radiat Oncol Biol Phys. 2010;76:761–766.
19. Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, Poptani H. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res. 2009;15:986–994.
20. Galban CJ, Mukherji SK, Chenevert TL, Meyer CR, Hamstra DA, Bland PH, Johnson TD, Moffat BA, Rehemtulla A, Eisbruch A, Ross BD. 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.
21. Lu Y, Moreira AL, Hatzoglou V, Stambuk HE, Gonen M, Mazaheri Y, Deasy JO, Shaha AR, Tuttle RM, Shukla-Dave A. Using diffusion-weighted MRI to predict aggressive histological features in papillary thyroid carcinoma: a novel tool for preoperative risk stratification in thyroid cancer. Thyroid. 2015;25:672–680.
22. Razek AA, Sadek AG, Kombar OR, Elmahdy TE, Nada N. Role of apparent diffusion coefficient values in differentiation between malignant and benign solitary thyroid nodules. Am J Neuroradiol. 2008;29:563–568.
23. Andreou A, Koh DM, Collins DJ, Blackledge M, Wallace T, Leach MO, Orton MR. Measurement reproducibility of perfusion fraction and pseudodiffusion coefficient derived by intravoxel incoherent motion diffusion-weighted MR imaging in normal liver and metastases. Eur Radiol. 2013;23:428–434.
24. Hauser T, Essig M, Jensen A, Laun FB, Munter M, Maier-Hein KH, Stieltjes B. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: evaluation of lymph node metastasis. Eur J Radiol. 2014;83:783–787.
25. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168:497–505.
26. 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;161:401–407.
27. Ding Y, Hazle JD, Mohamed AS, Frank SJ, Hobbs BP, Colen RR, Gunn GB, Wang J, Kalpathy-Cramer J, Garden AS, Lai SY, Rosenthal DI, Fuller CD. Intravoxel incoherent motion imaging kinetics during chemoradiotherapy for human papillomavirus-associated squamous cell carcinoma of the oropharynx: preliminary results from a prospective pilot study. NMR Biomed. 2015;28:1645–1654.
28. Paudyal R, OH JH, Riaz N, Venigalla P, Li J, Hatzoglou V, Leeman J, Nunez DA, Lu Y, Deasy JO, Lee N, Shukla-Dave A. Intravoxel incoherent motion diffusion-weighted MRI during chemoradiation therapy to characterize and monitor treatment response in human papillomavirus head and neck squamous cell carcinoma. J Magn Reson Imaging. 2017;45:1013–1023.
29. Dyvornea H, Jajamovicha G, Kakitea S, Kuehn B, Taoulia B. Intravoxel incoherent motion diffusion imaging of the liver:Optimal b-value subsampling and impact on parameterprecision and reproducibility. Eur J Radiol. 2014;83:2109–2113.
30. Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. Radiology. 2013;268:318–322.
31. Jansen JF, Stambuk HE, Koutcher JA, Shukla-Dave A. Non-Gaussian analysis of diffusion-weighted MR imaging in head and neck squamous cell carcinoma: a feasibility study. AJNR Am J Neuroradiol. 2010;31:741–748.
32. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed. 2010;23:698–710.

33. Yuan J, Yeung DKW, Mok GSP, Bhatia KS, Wang YXJ, Ahuja AT, King AD. Non-gaussian analysis of diffusion weighted imaging in head and neck at 3T: a pilot study in patients with nasopharyngeal carcinoma. PLoS One. 2014;9:e87024.
34. Lu Y, Jansen JF, Mazaheri Y, Stambuk HE, Koutcher JA, Shukla-Dave A. Extension of the intravoxel incoherent motion model to non-Gaussian diffusion in head and neck cancer. J Magn Reson Imaging. 2012;36:1088–1096.
35. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M, Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC; QIBA Technical Performance Working Group. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24:27–67.
36. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, Guimaraes AR, Filice R, Zhang Z, Sullivan DC; QIBA Terminology Working Group. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res. 2015;24:9–26.
37. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan DC; Algorithm Comparison Working Group. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res. 2015;24:68–106.
38. Boss MA, editor. Multicenter Study of Reproducibility of Wide Range of ADC at 0°C. Chicago IL: RSNA; 2015.
39. Palacios EM, Martin AJ, Boss MA, Ezekiel F, Chang YS, Yuh EL, Vassar MJ, Schnyer DM, MacDonald CL, Crawford KL, Irimia A, Toga AW, Mukherjee P; TRACK-TBI Investigators. Toward precision and reproducibility of diffusion tensor imaging: a multicenter diffusion phantom and traveling volunteer study. Am J Neuroradiol. 2017;38:537–545.
40. Swanson SD, Malyarenko DI, Fabiilli ML, Paudyal R, LoCastro E, Jambawalikar SR, Liu MZ, Schwartz LH, Shukla-Dave A, Chenevert TL, editors. Design and Development of a Novel Phantom to Assess Quantitative Diffusion Kurtosis Imaging, a Multi-Site Initiative. Bethesda, MD: NCI/Quantitative Imaging Network (QIN) Annual Face-to-Face meeting; 2018.
41. Pierpaoli C, Sarlls J, Nevo U, Basser PJ, Horkay F, editors. Polyvinylpyrrolidone (PVP) Water Solutions as Isotropic Phantoms for Diffusion MRI Studies. Proceedings of the 17th Annual Meeting of the ISMRM; 2009; Honolulu, HW.
42. Lu Y, Hatzoglou V, Banerjee S, Stambuk HE, Gonen M, Shankaranarayanan A, Mazaheri Y, Deasy JO, Shaha AR, Tuttle RM, Shukla-Dave A. Repeatability investigation of reduced field-of-view diffusion-weighted magnetic resonance imaging on thyroid glands. J Comput Assist Tomography. 2015;39:334–339.
43. Jansen JF, Koutcher JA, Shukla-Dave A. Non-invasive imaging of angiogenesis in head and neck squamous cell carcinoma. Angiogenesis. 2010;13:149–160.
44. Rasband WS. ImageJ. Bethesda, MD: U. S. National Institutes of Health; 1997–2016.
45. Chenevert TL, Galban CJ, Ivancevic MK, Rohrer SE, Londy FJ, Kwee TC, Meyer CR, Johnson TD, Rehemtulla A, Ross BD. Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. J Magn Reson Imaging. 2011;34:983–987.
46. Le Bihan D. Intravoxel incoherent motion imaging using steady-state free precession. Magn Reson Med. 1988;7:346–351.
47. Nakahira M, Saito N, Yamaguchi H, Kuba K, Sugasawa M. Use of quantitative diffusion-weighted magnetic resonance imaging to predict human papilloma virus status in patients with oropharyngeal squamous cell carcinoma. Eur Arch Otorhinolaryngol. 2014;271:1219–1225.
48. Arlinghaus LR, Dortch RD, Whisenant JG, Kang H, Abramson RG, Yankeelov TE. Quantitative magnetization transfer imaging of the breast at 3.0 T: reproducibility in healthy volunteers. Tomography. 2016;2:260–266.
49. Galbraith SM, Lodge MA, Taylor NJ, Rustin GJ, Bentzen S, Stirling JJ, Padhani AR. Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. NMR Biomed. 2002;15:132–142.
50. RstudioTeam. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc; 2015.
51. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, Muzi M, Solaiyappan M, Jacobs M, Fung M, Shukla-Dave A, McManus K, Boss M, Taouli B, Yankeelov TE, Quarles CC, Schmainda K, Chenevert TL, Newitt DC. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging. 2018;5:011006.
52. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan PE, Schmainda KM, Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen YF, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging. 2018;5:011003.
53. Bonekamp D, Nagae LM, Degaonkar M, Matson M, Abdalla WM, Barker PB, Mori S, Horská A. Diffusion tensor imaging in children and adolescents: reproducibility, hemispheric, and age-related differences. Neuroimage. 2007;34:733–742.
54. Lee Y, Lee SS, Kim N, Kim E, Kim YJ, Yun SC, Kühn B, Kim IS, Park SH, Kim SY, Lee MG. Intravoxel incoherent motion diffusion-weighted MR imaging of the liver: effect of triggering methods on regional variability and measurement repeatability of quantitative parameters. Radiology. 2015;274:405–415.
55. Le Bihan D. Diffusion MRI: what water tells us about the brain. EMBO Mol Med. 2014;6:569–73.
56. Padhani AR, Koh DM. Diffusion MR imaging for monitoring of treatment response. Magn Reson Imaging Clin North Am. 2011;19:181–209.
57. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials. 1981;2:93–113.
58. Paldino MJ, Barboriak D, Desjardins A, Friedman HS, Vredenburgh JJ. Repeatability of quantitative parameters derived from diffusion tensor imaging in patients with glioblastoma multiforme. J Magn Reson Imaging. 2009;29:1199–1205.
59. Pfefferbaum A, Adalsteinsson E, Sullivan EV. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. J Magn Reson Imaging. 2003;18:427–433.