

OPEN

# Multiparametric Magnetic Resonance Imaging of the Prostate

## Repeatability of Volume and Apparent Diffusion Coefficient Quantification

Andriy Fedorov, PhD,\* Mark G. Vangel, PhD,† Clare M. Tempany, MD,\* and Fiona M. Fennessy, MD, PhD\*‡

**Objectives:** The aim of this study was to evaluate the repeatability of a region of interest (ROI) volume and mean apparent diffusion coefficient (ADC) in standard-of-care 3 T multiparametric magnetic resonance imaging (mpMRI) of the prostate obtained with the use of endorectal coil.

**Materials and Methods:** This prospective study was Health Insurance Portability and Accountability Act compliant, with institutional review board approval and written informed consent. Men with confirmed or suspected treatment-naïve prostate cancer scheduled for mpMRI were offered a repeat mpMRI within 2 weeks. Regions of interest corresponding to the whole prostate gland, the entire peripheral zone (PZ), normal PZ, and suspected tumor ROI (tROI) on axial T2-weighted, dynamic contrast-enhanced subtract, and ADC images were annotated and assessed using Prostate Imaging Reporting and Data System (PI-RADS) v2. Repeatability of the ROI volume for each of the analyzed image types and mean ROI ADC was summarized with repeatability coefficient (RC) and RC%.

**Results:** A total of 189 subjects were approached to participate in the study. Of 40 patients that gave initial agreement, 15 men underwent 2 mpMRI examinations and completed the study. Peripheral zone tROIs were identified in 11 subjects. Tumor ROI volume was less than 0.5 mL in 8 of 11 subjects. PI-RADS categories were identical between baseline-repeat studies in 11/15 subjects and differed by 1 point in 4/15. Peripheral zone tROI volume RC (RC%) was 233 mm<sup>3</sup> (71%) on axial T2-weighted, 422 mm<sup>3</sup> (112%) on ADC, and 488 mm<sup>3</sup> (119%) on dynamic contrast-enhanced subtract. Apparent diffusion coefficient ROI mean RC (RC%) were  $447 \times 10^{-6}$  mm<sup>2</sup>/s (42%) in PZ tROI and  $471 \times 10^{-6}$  mm<sup>2</sup>/s (30%) in normal PZ. Significant difference in repeatability of the tROI volume across series was observed ( $P < 0.005$ ). The mean ADC RC% was lower than volume RC% for tROI ADC ( $P < 0.05$ ).

**Conclusions:** PI-RADS v2 overall assessment was highly repeatable. Multiparametric magnetic resonance imaging sequences differ in volume measurement repeatability. The mean tROI ADC is more repeatable compared with tROI volume in ADC. Repeatability of prostate ADC is comparable with that in other abdominal organs.

**Key Words:** magnetic resonance imaging, prostate, apparent diffusion coefficient, dynamic contrast-enhanced imaging, repeatability, quantitative imaging

(*Invest Radiol* 2017;52: 538–546)

Multiparametric magnetic resonance imaging (mpMRI) has emerged in the past decade as the most promising imaging tool for detection of suspicious lesions, characterization, staging, selection of treatment, guiding targeted biopsies, and even screening for prostate cancer (PCa).<sup>1–5</sup> Multiparametric magnetic resonance imaging has been used for evaluating lesion volume<sup>6</sup> and for extracting image-based quantitative metrics that correlate with the functional characteristics of the tissue, such as angiogenesis, tumor cell density,<sup>7</sup> and cell membrane disruption.<sup>8</sup> Increasingly, mpMRI is being explored in a longitudinal setting to assess the effect of therapy<sup>9,10</sup> or evaluate for disease progression in patients under active surveillance (AS).<sup>11,12</sup> As such, mpMRI is well positioned as a promising exploratory imaging biomarker of PCa.

Knowledge of measurement repeatability is critical in longitudinal studies as it enables differentiation of true change caused by the tissue biology from measurement noise.<sup>13,14</sup> Although mpMRI is already being actively explored as a tool for evaluating longitudinal changes,<sup>10,12,15</sup> little is known about the repeatability of mpMRI quantification. Many of the prior studies evaluating repeatability of prostate mpMRI<sup>16–19</sup> focused on repeatability of ADC alone in patients with pathology-proven PCa who will require definitive therapy. The relationship between repeatability of quantitative mpMRI metrics and consistency of lesion volume has not been investigated in the prostate. This is important considering the increasing role of mpMRI in the AS population, where it may assist in detecting clinically significant PCa that warrants treatment.<sup>20–22</sup> Several definitions of clinically significant PCa exist,<sup>12</sup> one of which being tumor volume greater than 0.5 mL.<sup>23</sup> Little is known about repeatability of standard-of-care mpMRI applied to quantification of lesion volume. Repeatability studies of the quantitative indices based on diffusion-weighted MRI (DWI), especially in the low volume disease, are limited and often do not report on the volumetry of the analyzed regions. In a recent consensus study concerned with developing guidelines for applying mpMRI in the AS setting, the “optimal way of measuring lesion size to allow repeatability over time, and both the change in size and absolute size that should prompt clinical action” were identified as areas most in need of research.<sup>12</sup>

In this study, we aim to bridge this gap in knowledge by investigating the repeatability of mpMRI-based quantitative measurement. We focus specifically on repeatability of the region of interest (ROI) volume and mean apparent diffusion coefficient (ADC). Tumor volume has long been a key determinant in establishing clinical significance of PCa<sup>23</sup> and its interval change is an important indicator of treatment response.<sup>24</sup> Moreover, volumetric localization of the lesion is a prerequisite for its quantitative assessment using functional imaging parameters. Diffusion-weighted MRI, and ADC estimated from DWI studies, has been shown to correlate with underlying histological features of PCa<sup>25</sup> and is also a promising marker of treatment response.<sup>10</sup> Imaging-based determination of clinical significance of the disease or its response to treatment using these measures requires investigation of the uncertainty in their measurements. There are several community efforts that motivate our evaluation. Prostate Imaging Reporting and Data System (PI-RADS) is a scoring system that aims to enable consistent interpretation, communication, and reporting of mpMRI findings.<sup>26</sup> At present, the latest release of the PI-RADS guidelines, version 2, relies exclusively on the qualitative assessment of the mpMRI findings by the radiologist. Inclusion of the quantitative indices into the future revisions of PI-RADS is

Received for publication February 22, 2017; and accepted for publication, after revision, March 23, 2017.

From the \*Brigham and Women's Hospital, and †Massachusetts General Hospital, Harvard Medical School; and ‡Dana Farber Cancer Institute, Boston, MA.

Conflicts of interest and sources of funding: Funding support: NIH U01 CA151261, U24 CA180918, P41 EB015898, R01 CA111288, and R01 CA160902.

Dr. Tempany declares ongoing financial relationship with Profound Medical. All other authors report no conflicts of interest.

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site ([www.investigativeradiology.com](http://www.investigativeradiology.com)).

Correspondence to: Andriy Fedorov, PhD, 1249 Boylston St #344, Boston, MA 02215. E-mail: [fedorov@bwh.harvard.edu](mailto:fedorov@bwh.harvard.edu), [andrey.fedorov@gmail.com](mailto:andrey.fedorov@gmail.com).

Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0020-9996/17/5209–0538

DOI: 10.1097/RLI.0000000000000382

currently being considered and necessitates improved understanding of the repeatability of the quantitative measures. Quantitative Imaging Biomarker Alliance is engaged in developing imaging biomarker profiles that include performance claims for individual imaging modalities.<sup>27</sup> Understanding of the quantitative imaging biomarker repeatability is of utmost importance in developing such technical profiles and in applying such biomarkers for longitudinal evaluation of the disease progression.<sup>28</sup>

The primary goal of our study was to quantify repeatability of prostate and tumor volume, and mean ADC estimation. Our evaluation was done using data collected in a single-center setting from PCa 3 T mpMRI acquisitions obtained with the use of endorectal coil under standard-of-care clinical settings in a treatment-naive population. Imaging was performed within an interval of up to 2 weeks, with no intervening therapy. To aid in collection of the clinical evidence for the evaluation and improvement of PI-RADS, we also evaluated repeatability of the PI-RADS v2 grading.

## MATERIALS AND METHODS

Institutional review board approval was obtained for this Health Insurance Portability and Accountability Act–compliant study. Written informed consent was obtained from the study participants.

### Subjects

A subset of patients referred for prostate mpMRI at our institution between March 2013 and February 2015 was prospectively consented and enrolled into a research mpMRI study. Selection of patients approached for participation in the study was based on the availability of our part-time research coordinator. Enrolled patients, as part of the consent process, were offered a repeat mpMRI examination within 2 weeks of the baseline examination. Inclusion criteria for this study were the ability of the subject to undergo prostate MRI with endorectal coil and completion of the repeat imaging examination. All patients were treatment-naive at the time of imaging. When present, prostate biopsy pathology results and serum prostate-specific antigen (PSA) were reported.

### Magnetic Resonance Protocol

A standard, PI-RADS v2 compliant, prostate mpMRI protocol was used in this study.<sup>1</sup> For a given patient, we aimed to maintain similar protocol settings and used the same scanner hardware and software configurations for both the baseline and repeat examinations that took place within 2 weeks of each other. Due to a scanner hardware upgrade in the middle of the study, 6 of the patients had baseline and repeat study performed on a GE Signa HDxt platform, whereas the remaining 7 patients had their baseline and repeat study on a GE Discovery MR750w (General Electric Healthcare, Milwaukee, WI). An endorectal coil (e-coil) within an air-filled balloon (Medrad Inc, Warrendale, PA) was used in all imaging studies. Our mpMRI protocol<sup>1</sup> included axial T2-weighted (T2AX), diffusion-weighted (DW) (b-values of 0 and 1400 mm<sup>2</sup>/s<sup>2</sup>), and dynamic contrast-enhanced (DCE) sequences (acquisition parameters in Table 1). Diffusion-weighted MRI apparent diffusion coefficient (ADC) and DCE subtract maps (SUB, computed as

the difference between the phase corresponding to the contrast bolus arrival and the baseline phase) were computed using the scanner software.

### Image Annotation and Quantitation

Imaging studies were de-identified and presented in a randomized fashion (reader was blinded to the given study being a baseline or repeat study, and the studies were presented in a permuted order) to an abdominal radiologist (F.F., 12 years in prostate mpMRI interpretation) using 3D Slicer software (<http://slicer.org>).<sup>29</sup> A consistent hanging protocol was applied to present all of the imaging series required for mpMRI assessment, as specified by PI-RADS v2.<sup>26</sup> Results of the prior PSA tests and systematic transrectal ultrasound (TRUS) biopsy (when available as part of the medical record) were presented to the reader as an aid in tumor localization, per standard clinical interpretation workflow.

For each study, the reader confirmed its diagnostic quality, and that SUB was based upon subtraction of baseline precontrast images from first bolus arrival phase. Regions of interest were manually outlined from the T2AX, ADC, and SUB series for the whole gland (WG), peripheral zone (PZ), normal PZ (nPZ), and (when present) tumor-suspicious ROI (tROI). The reader could see the matching slice in all other series for the given study but was blinded to the ROIs contoured on other sequences of that study and to contours of the paired imaging study. Overall PI-RADS v2 assessment category between 1 and 5 was assigned for each study. The sector(s) containing the lesion were noted.<sup>26</sup> Annotated ROIs identifying voxels corresponding to the suspected lesions were saved as rasterized segmentations using 3D Slicer software for subsequent extraction of ROI statistics.

Agreement in locations of tROI areas was assessed by a separate reader by comparing agreement of the noted lesion sector(s) between baseline and repeat studies. We did not attempt to validate the location of the annotated lesion with either biopsy or WG pathology. Because the data were collected prospectively under standard-of-care conditions, neither targeted biopsy sampling nor whole mount prostatectomy pathology were available to us. Spatial voxel-wise correspondence between the contoured regions was not attempted due to expected misregistration between the baseline and repeat examinations. Volumes of the WG, PZ, and tROI regions, and mean ADC value for the WG, PZ, nPZ, and tROI areas were automatically calculated from the ROIs.

### Statistical Analysis

Agreement of the PI-RADS v2 categories was assessed using weighted Cohen kappa statistics. Repeatability of the quantitative measurements (absolute and relative mean and standard deviation [SD] of the measurement was assessed following established guidelines using repeatability coefficient [RC], and RC as a percentage of the mean [RC%]).<sup>28,30–32</sup> We modeled ROI volume for 3 image types (T2AX, SUB, and ADC) and 3 structures (WG, PZ, and tROI) using mixed-model regression analyses, with subject as a random effect. Similarly, mean signal over ROI was modeled for ADC in these 3 structures and also in the nPZ ROI. We examined the differences in measurement mean first due to ROI, for each valid combination of measurement type, structure, and image type using 1-way analysis of variance, and next

**TABLE 1.** Ranges of the Acquisition Parameters for the Analyzed MRI Image Types

	Repetition Time, ms	Echo Time, ms	Field of View, mm	Slice Thickness, mm	Matrix Size
T2AX	3350–5109	84–107	140–200	3	288–384 × 192–224
DCE	3.68–4.1	1.3–1.42	260–280	5–6	256–288 × 140–160
DWI	2500–8150	76.7–80.6	160–280	3–4	96–128 × 96

MRI indicates magnetic resonance imaging; T2AX, T2-weighted axial image; DCE, dynamic contrast-enhanced image; DWI, diffusion-weighted image.

due to ROI and image type, for each structure using 2-way analysis of variance mixed model regression, whereas adjusting for multiple comparisons using the approach of Bretz et al.<sup>33</sup> Standard error and confidence interval for RC and RC% were estimated using the delta method. RC% of ADC volume and mean ADC were compared using 2-tailed z test. Results were tabulated and summarized using Bland-Altman plots.<sup>34</sup> Analysis was performed using R 3.1.0.<sup>35</sup>

## RESULTS

### Population

A total of 189 patients (the 189 patients we approached amounts to approximately 7.5% of the estimated total volume of the prostate MRI collected at our institution over the course of the study enrollment, considering typical volume of 5 prostate MRI patients per day) were offered to undergo repeat mpMRI examination with endorectal coil, out of which 40 patients agreed. The repeat study was performed in 15 of those who agreed. Sequence acquisition parameters are summarized in Table 1, and clinical indications for the imaging examination are listed in Table 2. The remaining 25 of 40 subjects did not complete the repeat study due to conflicts in patient's schedule or withdrawal from the study (see the flow diagram in Fig. 1). The mean ± SD (range) age of the 15 subjects enrolled in the study was 61 ± 7 (47–69) years. Mean interval between the 2 mpMRIs was 10 (3–14) days. The mean ± SD (range) serum PSA in 14 of 15 subjects was 6.4 ± 2.2 (3.15–9.8) ng/mL. Prostate-specific antigen was not available in 1 subject, who was referred for imaging from an outside practice. Twelve of 15 patients had TRUS-guided sextant biopsy, of which 8 had pathology-confirmed PCa (3 + 3 Gleason score [GS]: 4; 3 + 4 GS: 3; 4 + 5 GS: 1), and 4 did not have confirmed PCa. Three of the patients did not undergo prostate biopsy before imaging, and thus no pathological grading was available. None of the patients had prostate biopsy performed between the baseline and repeat imaging examinations. Based on our experience, the population reported is typical of the prostate MR referrals for our institution. It consists

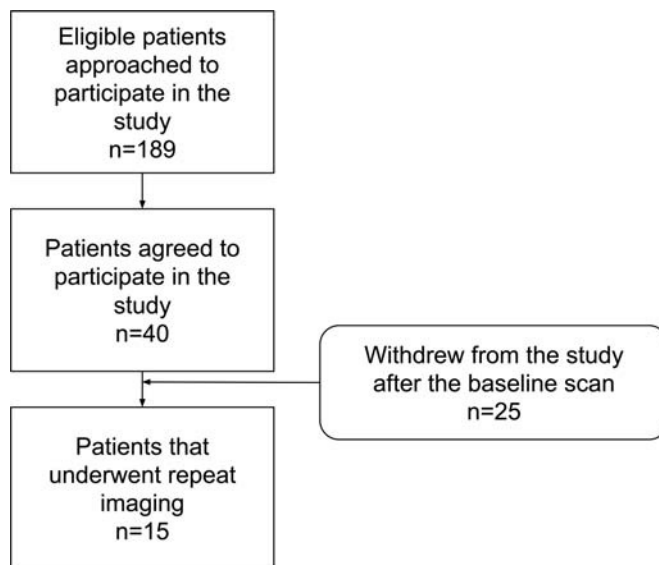


FIGURE 1. Flow diagram of the study population.

primarily of the patients with confirmed PCa for staging or assessment of changes, or referred for imaging assessment due to elevated PSA in absence of histological confirmation of the disease, as shown in Table 2.

### PI-RADS v2 Assessment

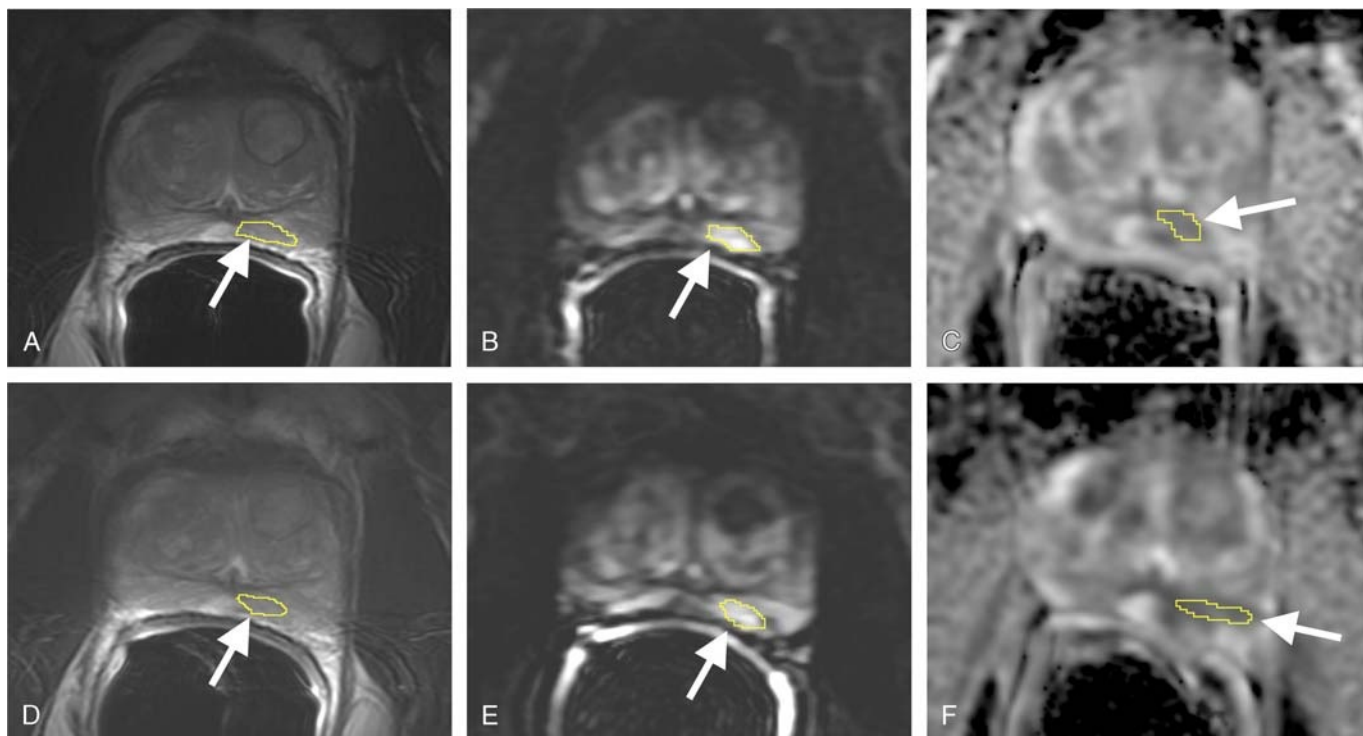
Tumor-suspicious tROIs were localized and contoured in 11 of 15 subjects (PI-RADS v2 > 1 in all 11/15 cases). All focal tROIs were located in the PZ. In 2 subjects, a secondary lesion was identified in the central zone of the gland, which was excluded from the subsequent quantitative analysis. No suspected lesion was identified in 4 of 15 subjects. In the subjects where a tROI was identified, in all cases, tROI was

TABLE 2. Patient-Level Summary of the Clinical Indications for MRI, Histopathology, PSA, and PI-RADS v2 Assessment of the Disease in the Evaluated Cohort

Subject	Indication for the MRI Examination	Maximum Gleason Grade at Biopsy	PSA, ng/mL	PI-RADSV2 Assessment Category, Baseline Study	PI-RADSV2 Assessment Category, Repeat Study
1	Known PCa, staging	3 + 4	5.4	4	4
2	Known PCa, assess change	3 + 4	7.5	2	3
3	Known PCa, staging	3 + 3	8.2	4	4
4	Known PCa, staging	3 + 3	4.3	2	2
5	Suspected PCa, staging	NA*	NA*	2	2
6	Elevated PSA, staging	Benign	5	1	1
7	Elevated PSA, staging	4 + 5	6.2	4	4
8	Known PCa, assess change	3 + 3	4.8	4	4
9	Elevated PSA, staging	Benign	9.4	4	4
10	Known PCa, assess change	3 + 3	3.15	4	4
11	Known PCa, assess change	3 + 3	9.7	4	4
12	Elevated PSA, staging	Benign	5.5	3	4
13	Known PCa, staging	3 + 4	4.16	4	3
14	Elevated PSA, staging	No biopsy performed	7	2	2
15	Repeat negative systematic TRUS biopsy, rising PSA	Benign	9.8	1	2

\*The subject was referred for imaging from an outside practice, and their Gleason and PSA were not present in our institution electronic health records system.

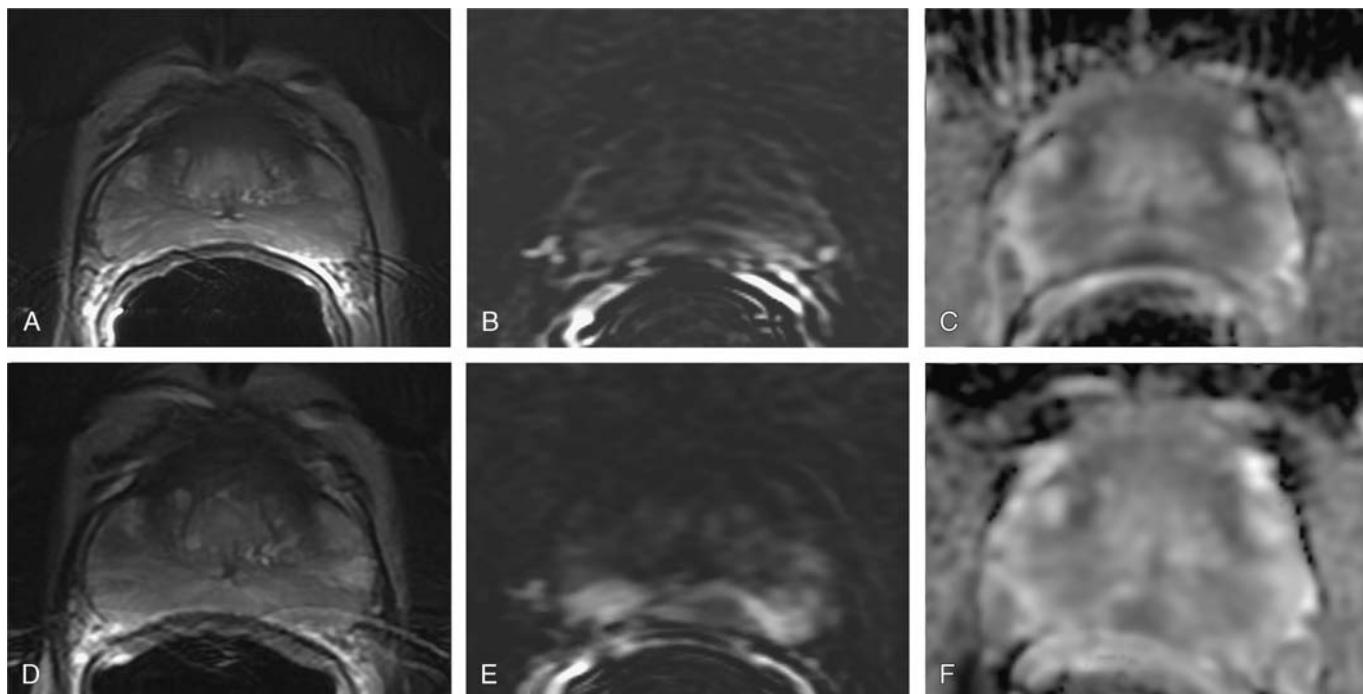
MRI indicates magnetic resonance imaging; PSA, prostate-specific antigen; PI-RADS, prostate imaging reporting and data system; PCa, prostate cancer; TRUS, transrectal ultrasound.



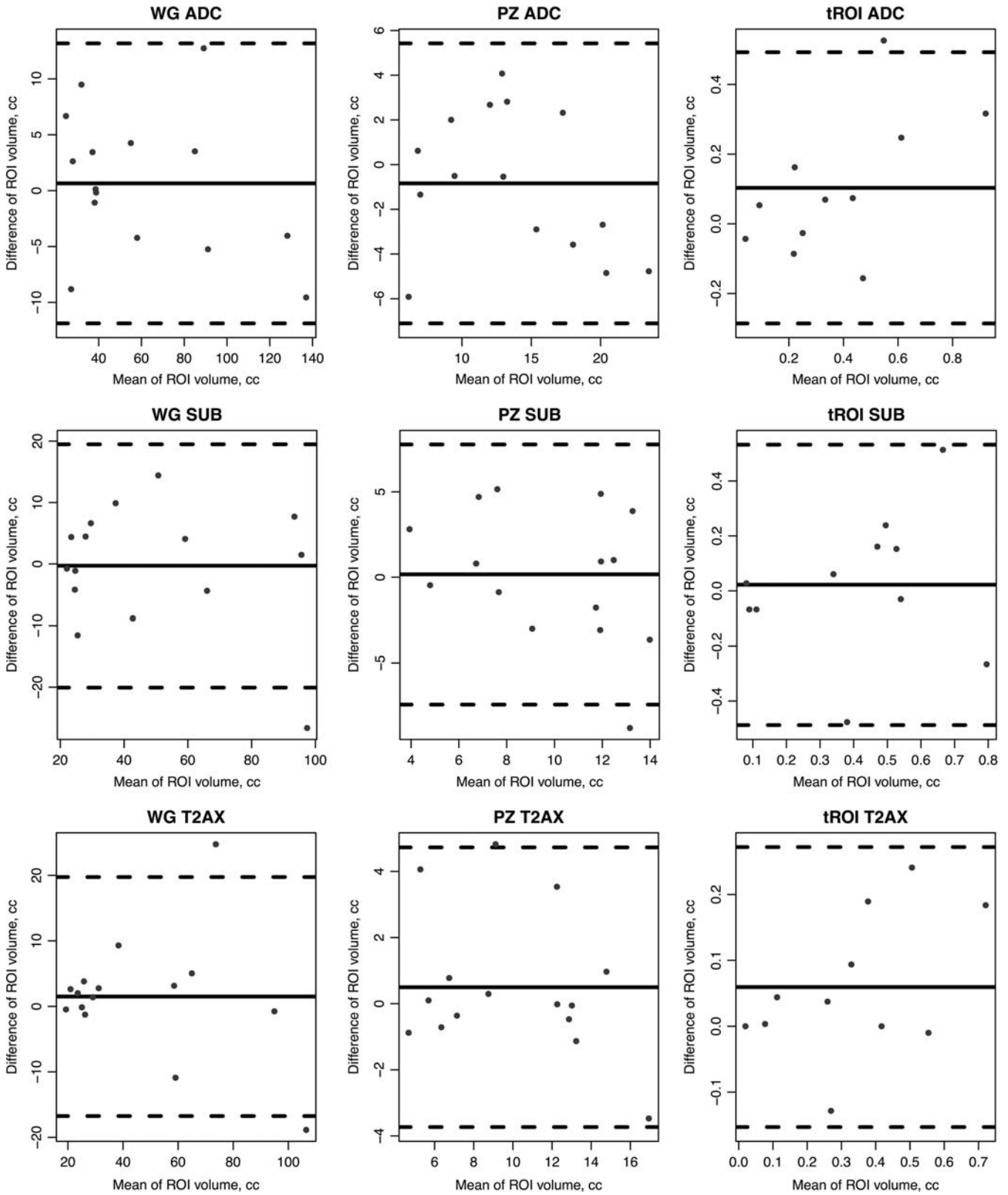
**FIGURE 2.** Midgland level axial T2-weighted (panels A and D), SUB (B and E), and ADC (C and F) images for subject 8 from Table 2. Top row (panels A–C) shows baseline study images, bottom row (panels D–F) is the repeat study. A contour (white arrow) is placed around the tumor on each sequence. At the time of ROI placement for each sequence, previously annotated ROI contours for other sequences within the same study were not visible to the radiologist.

localized in both baseline and repeat studies. There was agreement in the location of the tROI for all 11 cases. The images and annotations of the tROI for 2 of the subjects are shown in Figures 2 and 3.

Agreement in the PI-RADS v2 category was observed in 11 of 15 subjects as follows (also summarized in Table 2): PI-RADS v2 = 4 in 7 of 15 patients whose pathology was GS 4 + 5 ( $n = 1$ ), GS 3 + 4



**FIGURE 3.** Midgland level axial T2-weighted (panels A and D), SUB (B and E), and ADC (C and F) images for subject 14 from Table 2. Top row (panels A–C) shows baseline study images, bottom row (panels D–F) is the repeat study. No lesion was identified for this subject.



**FIGURE 4.** Bland-Altman plots summarizing repeatability of the volumetric measurements for the regions of interest (WG = whole gland, PZ = peripheral zone, tROI = tumor region of interest) across the 3 modalities (T2AX = axial T2-weighted image, SUB = dynamic contrast-enhanced MRI subtract map, ADC = apparent diffusion coefficient map) considered in this study.

**TABLE 3.** Repeatability of the Volumetric Measures for the Segmented Structures on T2-Weighted Axial Images, and ADC and SUB Maps

Structure/Modality	RC% (95% CI)	RC (95% CI), mm <sup>3</sup>	Mean Difference (% Mean Difference), mm <sup>3</sup>	SD of Difference (% SD of Difference), mm <sup>3</sup>
WG				
ADC	20.1 (10.33–29.87)	12166 (6253–18078)	5067 (8.37%)	3711 (6.13%)
SUB	39.8 (20.81–58.79)	19109 (9992–28226)	7364 (15.34%)	6613 (13.77%)
T2AX	38.44 (19.8–57.08)	17869 (9205–26534)	5817 (12.51%)	7266 (15.63%)
PZ				
ADC	45.98 (26.54–65.42)	6270 (3619–8921)	2773 (20.33%)	1651 (12.11%)
SUB	74.98 (44.31–105.65)	7351 (4343–10358)	3049 (31.11%)	2260 (23.05%)
T2AX	42.26 (24.43–60.09)	4201 (2428–5973)	1444 (14.53%)	1639 (16.49%)
PZ tROI				
ADC	112.22 (44.73–179.72)	422 (168–676)	159 (42.48%)	151 (40.27%)
SUB	119.43 (52.34–186.51)	488 (213–762)	187 (42.89%)	171 (42.05%)
T2AX	70.57 (29.05–112.09)	233 (96–371)	84 (25.56%)	88 (26.6%)

Repeatability of the WG and PZ volumetry estimation was obtained using data from all 15 subjects, whereas lesion repeatability was evaluated in 11 patients with identifiable lesions.

T2AX indicates T2-weighted axial image; SUB, DCE MRI subtract image computed as the difference between the phase corresponding to the contrast bolus arrival and the baseline phase; ADC, apparent diffusion coefficient; SD, standard deviation; PZ, peripheral zone; nPZ, normal PZ; tROI, tumor region of interest; RC, repeatability coefficient; CI, confidence interval.

(n = 1), GS 3 + 3 (n = 3), and no tumor (n = 2); PI-RADS v2 = 2 in 3 of 15 patients whose pathology was GS 3 + 3 (n = 1) and no biopsy performed (n = 2); and PI-RADS v2 = 1 in 1 of 15 patients whose prostate biopsy showed no PCa. There was disagreement of 1 point in the overall PI-RADS v2 category in 4 of 15 subjects (see Table 2). Cohen kappa evaluating the agreement between the 2 reads was 0.89 (95% confidence interval [CI], 0.78–1.00).

### ROI Volume and ADC Quantitation

Tumor-suspicious ROI volume, averaged between the measurements obtained in the baseline and repeat T2w scans, was less than 0.5 mL in 8 of the 11 subjects that had an identifiable lesion. Volumes of the considered structures across the baseline/repeat scans, and subjects were as follows (mean ± SD [range]): WG (T2AX, 46.5 ± 28.1 [19.1–115.9] mL; SUB, 48.2 ± 27.9 [19.6–110.7] mL; ADC, 60.5 ± 36.4 [21.3–141.8] mL), PZ (T2AX, 9.9 ± 3.9 [3.2–18.6] mL; SUB, 9.8 ± 3.7 [2.5–17.5] mL; ADC, 13.6 ± 5.5 [3.2–25.8] mL), and tROI (T2AX, 0.3 ± 0.2 [0.02–0.8] mL; SUB, 0.4 ± 0.2 [0.05–0.9] mL; ADC, 0.3 ± 0.2 [0.02–1.1] mL).

Linear mixed-effects model testing showed that volumes of the WG and PZ obtained from ADC maps were significantly larger as compared with those obtained using T2w scans and SUB maps ( $P < 0.00005$ ). However, no statistically significant differences were observed for the tROI volumes.

We observed a gradual increase of measurement variability, and increasingly wider confidence intervals, going from large (WG) to small

(tROI) regions of interest, for both ROI volumes and mean ADC values. Repeatability of the volume measurements are summarized in Figure 4 and Table 3. Unequal variance testing identified no significant difference in the standard deviation of the measurements for either WG or PZ volumes across different sequences (see Table 4), whereas within-subject SD of the tROI volume was significantly different across sequences: 138 mm<sup>3</sup> in ADC, 224 mm<sup>3</sup> in SUB, and 68 mm<sup>3</sup> in T2AX,  $P < 0.0005$ . Repeatability for the mean ADC measurements is summarized in Figure 5 and Table 5. RC% of mean ADC was significantly lower than that of ADC volume in PZ ( $P = 0.03$ ) and tROI ( $P = 0.049$ ).

Coverage probability (CP) curves, intraclass coefficient (ICC), and concordance correlation coefficient (CCC) measures were also calculated and are included in the Supplemental Material, Supplemental Digital Content 1, <http://links.lww.com/RLI/A319>.

### DISCUSSION

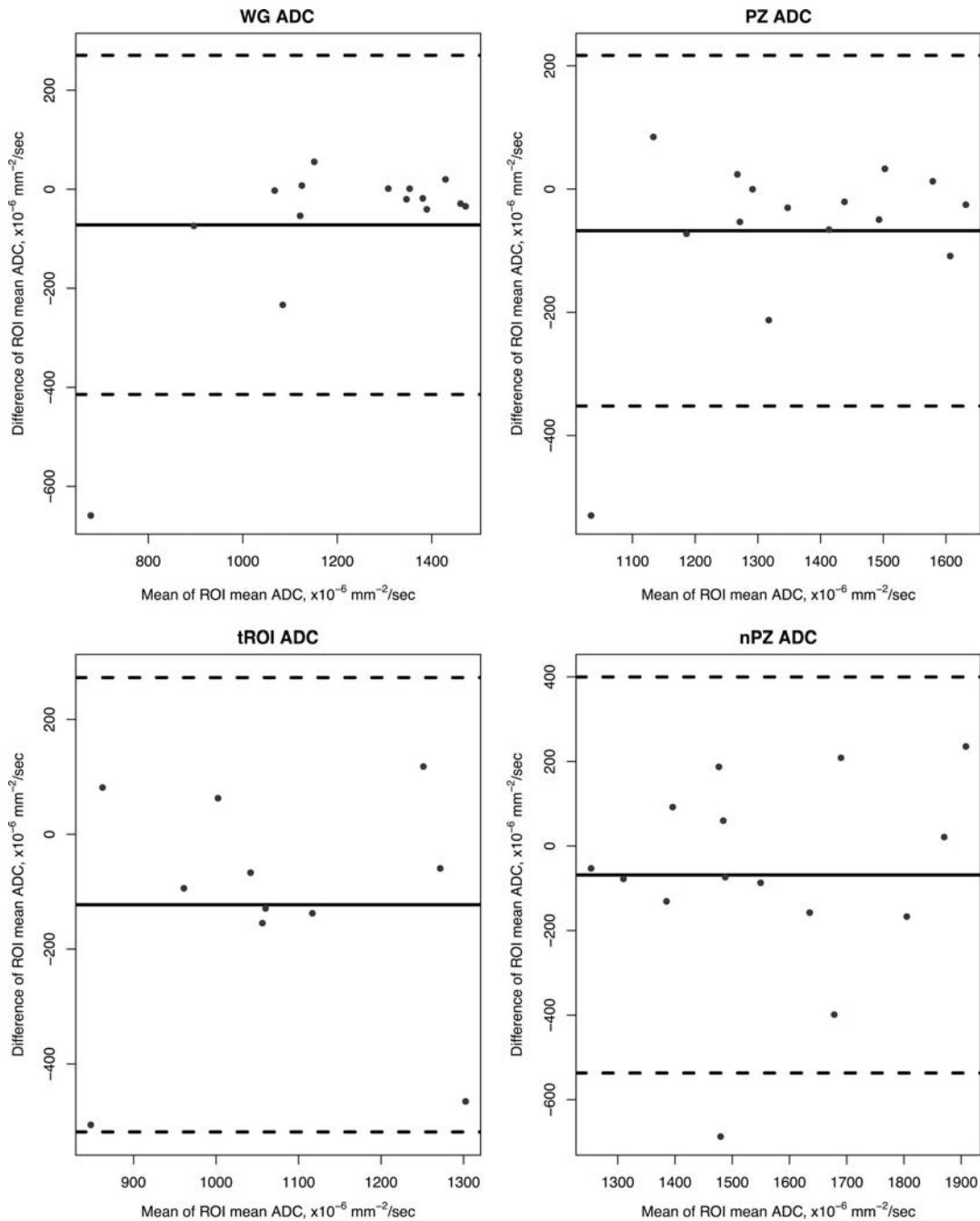
In this study, we evaluated repeatability of e-coil prostate mpMRI at 3 T, in a mixed cohort of 15 consecutive, consenting treatment-naive patients, being evaluated for PCa.

We considered a variety of repeatability measures, including those recommended by the consensus guidelines developed by Quantitative Imaging Biomarker Alliance,<sup>23,25</sup> along with their confidence intervals. RC quantifies the absolute repeatability of the measurement in the same units as the measurement itself. It is defined as a 95% upper confidence bound on the absolute difference between the 2 replicate measurements and is directly related to the limits of agreement

**TABLE 4.** Comparison of the Within-Subject Standard Deviation in ROI Volume Measurements Across Modalities Estimated With Unequal Variance Analysis

Structure	ADC Volume SD, mm <sup>3</sup>	Subtract Volume SD, mm <sup>3</sup>	T2AX Volume SD, mm <sup>3</sup>	P
WG	8458	9030	7722	0.81
PZ	2617	2882	2316	0.70
PZ tROI	138	224	68	0.0004

ROI indicates region of interest; ADC, apparent diffusion coefficient; SD, standard deviation; PZ, peripheral zone; tROI, tumor region of interest; T2AX, T2-weighted axial image.



**FIGURE 5.** Bland-Altman plots summarizing repeatability of the mean apparent diffusion coefficient measurements for the regions of interest (WG = whole gland, PZ = peripheral zone, tROI = tumor region of interest, nPZ = normal peripheral zone).

measure.<sup>27</sup> Based on our review of the literature, repeatability evaluations are often limited to absolute and relative percent differences, and RC is often not reported. This motivated our inclusion of other, more commonly used measures, together with those recommended by the consensus guidelines.

Tumor-suspicious ROI volume seems to be most stable when measured on T2-weighted axial images (~26% as the difference relative to the mean and ~71% RC%). RC and RC% of the tROI from ADC and SUB were about twice as large. This ordering is not unexpected, considering T2AX images have higher resolution, do not suffer from the DWI distortions, and are not dependent on the choice of the contrast uptake

phases and organ motion in DCE analysis. Mean ADC (~16% as the relative difference and ~42% RC% for the PZ tROI) is more stable than volumetric measurements.

Changes in prostate gland volume may be useful preoperatively while evaluating response to hormonal therapy and in determining response of benign prostatic hyperplasia to androgen deprivation.<sup>36</sup> Gland volume is also required for evaluating changes in PSA density and is sometimes considered as a potential marker for evaluating disease progression in patients under AS.<sup>11</sup> These uses justify evaluation of the WG measurements repeatability. The WG and PZ regions can also serve as a measure of reference in evaluating the impact of region size

TABLE 5. Repeatability of the Mean ADC Measurements for the Segmented Structures

Structure	RC% (95% CI)	RC (95% CI), $\times 10^{-6} \text{ mm}^{-2}/\text{s}$	Mean Difference (% Mean Difference), $\times 10^{-6} \text{ mm}^{-2}/\text{s}$	SD of Difference (% SD of Difference), $\times 10^{-6} \text{ mm}^{-2}/\text{s}$
WG	29.53 (18.23–40.83)	359 (221–497)	83 (6.85%)	169 (13.89%)
PZ	22.31 (13.91–30.71)	305 (190–419)	88 (6.45%)	132 (9.71%)
nPZ	30.24 (18.86–41.61)	471 (294–649)	175 (11.27%)	170 (10.91%)
PZ tROI	41.8 (23.12–60.49)	447 (247–647)	170 (15.93%)	159 (14.88%)

ADC indicates apparent diffusion coefficient; SD, standard deviation; PZ, peripheral zone; nPZ, normal PZ; tROI, tumor region of interest; T2AX, T2-weighted axial image; RC, repeatability coefficient; CI, confidence interval.

on repeatability: our results show that measurements become less repeatable for smaller regions, as can be seen from Table 3. In general, we expect that the volume measurement error will be proportional to the surface area of the ROI, which in turn will depend on the shape of the region, because the segmentation error is concerned with the localization of the ROI boundary. This explains the increasing RC % and decreasing absolute RC for the regions we evaluated, going from large size WG ROI to the smaller size tROIs. Given the small number of subjects with lesions ( $n = 11$ ), we could not make any statistically justified statements about the dependence of RC% or RC on the lesion size for tROIs.

We report both the absolute and relative repeatability measures. This choice is motivated by the lack of consensus on what is the best approach for assessing change in PCa lesions; Moore et al<sup>12</sup> suggested that small absolute changes in volume may appear as large percentage changes for small lesions. This is particularly relevant in the present study, because most of the lesions evaluated were below the clinically significant volume threshold.<sup>23</sup> It is important to emphasize that the threshold of 0.5 mL was established based on the volume of the pathology estimated from the surgical specimen.<sup>23</sup> There is evidence of high discordance between the MRI-based and histopathology volumes, especially for small volume disease,<sup>37</sup> which is prevalent in indolent PCa observed in AS patients: a recent AS study reported imaging-based average volume of 0.3 mL,<sup>11</sup> stressing the need for better understanding of repeatability in a low disease burden follow-up setting.

Repeatability of quantitative mpMRI measurements in the prostate has been investigated in the past. Alonzi et al<sup>17</sup> studied repeatability of DCE-derived metrics at 1.5 T without the use of e-coil. We did not evaluate repeatability of the parametric maps derived from multiphase DCE MRI, because it is recognized that DCE MRI does not play a primary role in detection of PCa and differentiation of aggressive cancer, as supported by PI-RADS v2 criteria for detection of PZ lesions.<sup>26</sup> We are not aware of prior studies evaluating the repeatability of mpMRI-based volumetric measurements. However, similar repeatability studies have been performed for PET/CT in advanced gastrointestinal malignancies, reporting RC of ~36%.<sup>38</sup>

There have been prior studies investigating the repeatability of DWI-based measurements in the prostate. Gibbs et al evaluated prostate ADC repeatability within 1 month of acquisition in healthy subjects ( $n = 8$ ), reporting repeatability of 35%.<sup>16</sup> Litjens et al<sup>18</sup> evaluated ADC measurement variability in a cohort of 10 subjects scanned at the interval of 6 to 12 months and observed average difference of ADC in PZ at  $68 \times 10^{-6} \text{ mm}^{-2}/\text{s}$  (neither RC nor %RC was reported), as compared to  $175 \times 10^{-6} \text{ mm}^{-2}/\text{s}$  (see Table 5) observed in our cohort that underwent repeat scan within 2 weeks. Sadiniski et al<sup>19</sup> investigated short-term repeatability of mean ROI ADC in 14 subjects with biopsy-proven PCa, with the same subject scanned twice without repositioning within the scanner bore. Their median tROI ADC variation was  $36 \times 10^{-6} \text{ mm}^{-2}/\text{s}$  (4.2%), lower than the  $170 \times 10^{-6} \text{ mm}^{-2}/\text{s}$  (15.9%) we report (see Table 5). This is not unexpected, considering there was no change in patient positioning and minimal time between the scans. Several

recent studies led by Jambor et al investigated repeatability of DWI-derived parameters using multi-b acquisition without e-coil.<sup>39,40</sup> Those studies evaluated short-term (10–15 minutes) repeatability but did not consider lesion volumetrics and were conducted under clinical trial conditions, which are typically different from clinical standard-of-care settings. Overall, we note that repeatability values we report are comparable with those for other abdominal organs.<sup>41–44</sup>

Our study has several limitations. First, our cohort is small, although we should recognize that we approached 189 patients to partake in the study. Only 40 patients agreed to the study, and a repeat MR was possible in 15 subjects, underscoring the challenges faced in MR repeatability studies enrollment, especially those that utilize e-coil. Of note, the size of our cohort exceeds that reported by Litjens et al<sup>18</sup> and Gibbs et al.<sup>16</sup> Second, most of the lesions analyzed are below the pathology-based clinical significance volume threshold of 0.5 mL.<sup>23</sup> However, we also note that there is no clear consensus on the imaging-based definition of the significant disease,<sup>12</sup> and that quantification of changes is of particular importance in longitudinal follow-up of the low-volume disease in AS patients. Third, histological confirmation with prostatectomy could not be performed considering the disease characteristics of the enrolled patient population and the prospective nature of the study. We do not consider this a major limitation because the main goal was the evaluation of repeatability of the measurements derived from the ROIs consistently identified in mpMRI studies and not their pathological correlation.

In conclusion, we report repeatability measures for standard-of-care single-center e-coil prostate mpMRI at 3 T. Our results indicate that PI-RADS v2 category assignment is highly repeatable. Volumetric measurement of changes in tROI may be considered significant at the 95% confidence level if they exceed 70% ( $233 \text{ mm}^3$ ) on T2AX, or 112% to 119% ( $422\text{--}488 \text{ mm}^3$ ) in subtract and ADC imaging. Peripheral zone tROI mean ADC is more stable than ADC tROI volume, and its changes may be considered significant if they exceed 40% ( $447 \times 10^{-6} \text{ mm}^{-2}/\text{s}$ ). As with any small study, these results should be interpreted cautiously. Further investigation of mpMRI repeatability is warranted.

## REFERENCES

- Hegde JV, Mulkern RV, Panych LP, et al. Multiparametric MRI of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *J Magn Reson Imaging*. 2013;37:1035–1054.
- de Rooij M, Hamoen EH, Fütterer JJ, et al. Accuracy of multiparametric MRI for prostate cancer detection: a meta-analysis. *AJR Am J Roentgenol*. 2014;202:343–351.
- Panbianco V, Barchetti F, Sciarra A, et al. Multiparametric magnetic resonance imaging vs. standard care in men being evaluated for prostate cancer: a randomized study. *Urol Oncol*. 2015;33:17. e1–7.
- Scheenen TW, Rosenkrantz AB, Haider MA, et al. Multiparametric magnetic resonance imaging in prostate cancer management: current status and future perspectives. *Invest Radiol*. 2015;50:594–600.
- Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multiparametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017;389:815–822.



6. Donati OF, Afaq A, Vargas HA, et al. Prostate MRI: evaluating tumor volume and apparent diffusion coefficient as surrogate biomarkers for predicting tumor Gleason score. *Clin Cancer Res*. 2014;20:3705–3711.
7. Glazer DI, Hassanzadeh E, Fedorov A, et al. Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: correlation with tumor cell density and percentage Gleason pattern on whole mount pathology. *Abdom Radiol (NY)*. 2016. Available at: <http://dx.doi.org/10.1007/s00261-016-0942-1>.
8. 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.
9. Barrett T, Gilla B, Kataoka MY, et al. DCE and DW MRI in monitoring response to androgen deprivation therapy in patients with prostate cancer: a feasibility study. *Magn Reson Med*. 2012;67:778–785.
10. Hötker AM, Mazaheri Y, Zheng J, et al. Prostate cancer: assessing the effects of androgen-deprivation therapy using quantitative diffusion-weighted and dynamic contrast-enhanced MRI. *Eur Radiol*. 2015;25:2665–2672.
11. Morgan VA, Riches SF, Thomas K, et al. Diffusion-weighted magnetic resonance imaging for monitoring prostate cancer progression in patients managed by active surveillance. *Br J Radiol*. 2011;84:31–37.
12. Moore CM, Giganti F, Albertsen P, et al. Reporting magnetic resonance imaging in men on active surveillance for prostate cancer: the PRECISE recommendations—a report of a European School of Oncology Task Force. *Eur Urol*. 2017;71:648–655. Available at: <http://dx.doi.org/10.1016/j.eururo.2016.06.011>.
13. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. *Radiology*. 2015;277:813–825.
14. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2016. Available at: <http://dx.doi.org/10.1038/nrclinonc.2016.162>.
15. Padhani AR, MacVicar AD, Gapinski CJ, et al. Effects of androgen deprivation on prostatic morphology and vascular permeability evaluated with MR imaging. *Radiology*. 2001;218:365–374.
16. 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.
17. Alonzi R, Taylor NJ, Stirling JJ, et al. Reproducibility and correlation between quantitative and semiquantitative dynamic and intrinsic susceptibility-weighted MRI parameters in the benign and malignant human prostate. *J Magn Reson Imaging*. 2010;32:155–164.
18. Lijtens GJ, Hambroek T, Hulsbergen-van de Kaa C, et al. Interpatient variation in normal peripheral zone apparent diffusion coefficient: effect on the prediction of prostate cancer aggressiveness. *Radiology*. 2012;265:260–266.
19. Sadinski M, Medved M, Karademir I, et al. Short-term reproducibility of apparent diffusion coefficient estimated from diffusion-weighted MRI of the prostate. *Abdom Imaging*. 2015. Available at: <http://dx.doi.org/10.1007/s00261-015-0396-x>.
20. Somford DM, Hoeks CM, Hulsbergen-van de Kaa C, et al. Evaluation of diffusion-weighted MR imaging at inclusion in an active surveillance protocol for low-risk prostate cancer. *Invest Radiol*. 2013;48:152–157.
21. Hoeks CM, Somford DM, van Oort IM, et al. Value of 3-T multiparametric magnetic resonance imaging and magnetic resonance-guided biopsy for early risk re-stratification in active surveillance of low-risk prostate cancer: a prospective multicenter cohort study. *Invest Radiol*. 2014;49:165.
22. Schoots IG, Petrides N, Giganti F, et al. Magnetic resonance imaging in active surveillance of prostate cancer: a systematic review. *Eur Urol*. 2015;67:627–636.
23. Wolters T, Roobol MJ, van Leeuwen PJ, et al. A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial. *J Urol*. 2011;185:121–125.
24. Chung HT, Noworolski SM, Kurhanewicz J, et al. A pilot study of endorectal magnetic resonance imaging and magnetic resonance spectroscopic imaging changes with dutasteride in patients with low risk prostate cancer. *BJU Int*. 2011;108(8 pt 2):E164–E170.
25. Langer DL, Evans AJ, Plotkin A, et al. Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features. *Radiology*. 2010;255:485–494.
26. Barentsz JO, Weinreb JC, Verma S, et al. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. *Eur Urol*. 2016;69:41–49. Available at: <http://dx.doi.org/10.1016/j.eururo.2015.08.038>.
27. Buckler AJ, Bresolin L, Dunnick NR, et al. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology*. 2011;258:906–914.
28. Obuchowski NA, Buckler A, Kinahan P, et al. Statistical issues in testing conformance with the quantitative imaging biomarker alliance (QIBA) profile claims. *Acad Radiol*. 2016;23:496–506.
29. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30:1323–1341.
30. Barnhart HX, Yow E, Crowley AL, et al. Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting. *Stat Methods Med Res*. 2016;25:2939–2958. Available at: <http://dx.doi.org/10.1177/0962280214534651>.
31. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2015;24:27–67.
32. Vaz S, Falkner T, Passmore AE, et al. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One*. 2013;8:e73990.
33. Bretz F, Hothorn T, Westfall P. *Multiple Comparisons Using R*. CRC Press; 2011.
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
35. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2011. Available at: <http://www.r-project.org>.
36. Rahmouni A, Yang A, Tempny CM, et al. Accuracy of in-vivo assessment of prostatic volume by MRI and transrectal ultrasonography. *J Comput Assist Tomogr*. 1992;16:935–940.
37. Turkbey B, Mani H, Aras O, et al. Correlation of magnetic resonance imaging tumor volume with histopathology. *J Urol*. 2012;188:1157–1163.
38. Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology*. 2014;273:539–548.
39. Toivonen J, Merisaari H, Pesola M, et al. Mathematical models for diffusion-weighted imaging of prostate cancer using b values up to 2000 s/mm<sup>2</sup>: correlation with Gleason score and repeatability of region of interest analysis. *Magn Reson Med*. 2015;74:1116–1124.
40. Merisaari H, Movahedi P, Perez IM, et al. Fitting methods for intravoxel incoherent motion imaging of prostate cancer on region of interest level: repeatability and Gleason score prediction. *Magn Reson Med*. 2017;77:1249–1264. Available at: <http://dx.doi.org/10.1002/mrm.26169>.
41. Braithwaite AC, Dale BM, Boll DT, et al. Short- and midterm reproducibility of apparent diffusion coefficient measurements at 3.0-T diffusion-weighted imaging of the abdomen. *Radiology*. 2009;250:459–465.
42. Kim SY, Lee SS, Byun JH, et al. Malignant hepatic tumors: short-term reproducibility of apparent diffusion coefficients with breath-hold and respiratory-triggered diffusion-weighted MR imaging. *Radiology*. 2010;255:815–823.
43. Jafar MM, Parsai A, Miquel ME. Diffusion-weighted magnetic resonance imaging in cancer: reported apparent diffusion coefficients, in-vitro and in-vivo reproducibility. *World J Radiol*. 2016;8:21–49.
44. Johnston E, Punwani S. Can we improve the reproducibility of quantitative multiparametric prostate MR imaging metrics? *Radiology*. 2016;281:652–653.