Eur Radiol (2017) 27:1547–1555
DOI 10.1007/s00330-016-4449-5

MAGNETIC RESONANCE

Differentiation of prostate cancer lesions with high and with low Gleason score by diffusion-weighted MRI

Sebastiano Barbieri[1] · **Michael Brönnimann**[1] · **Silvan Boxler**[2] · **Peter Vermathen**[1] · **Harriet C. Thoeny**[1]

Received: 13 April 2016 / Revised: 17 May 2016 / Accepted: 23 May 2016 / Published online: 14 June 2016

Abstract

Objectives To differentiate prostate cancer lesions with high and with low Gleason score by diffusion-weighted-MRI (DW-MRI).

Methods This prospective study was approved by the responsible ethics committee. DW-MRI of 84 consenting prostate and/or bladder cancer patients scheduled for radical prostatectomy were acquired and used to compute apparent diffusion coefficient (ADC), intravoxel incoherent motion (IVIM: the pure diffusion coefficient D_t, the pseudo-diffusion fraction F_p and the pseudo-diffusion coefficient D_p), and high b value (as acquired and Hessian filtered) parameters within the index lesion. These parameters (separately and combined in a logistic regression model) were used to differentiate lesions depending on whether whole-prostate histopathological analysis after prostatectomy determined a high (≥7) or low (6) Gleason score.

Results Mean ADC and D_t differed significantly (p of independent two-sample t test < 0.01) between high- and low-grade lesions. The highest classification accuracy was achieved by the mean ADC (AUC 0.74) and D_t (AUC 0.70). A logistic regression model based on mean ADC, mean F_p and mean high b value image led to an AUC of 0.74 following leave-one-out cross-validation.

Conclusions Classification by IVIM parameters was not superior to classification by ADC. DW-MRI parameters correlated with Gleason score but did not provide sufficient information to classify individual patients.

✉ Harriet C. Thoeny
Harriet.Thoeny@insel.ch

[1] Institute of Diagnostic, Pediatric, and Interventional Radiology, Inselspital University Hospital, Inselspital, Freiburgstrasse 10, CH-3010 Bern, Switzerland

[2] Department of Urology, Inselspital, Inselspital University Hospital, Bern, Switzerland

Key Points

- *Mean ADC and diffusion coefficient differ between high- and low-grade prostatic lesions.*
- *Accuracy of trivariate logistic regression is not superior to using ADC alone.*
- *DW-MRI is not a valid substitute for biopsies in clinical routine yet.*

Keywords Prostate cancer · Diffusion-weighted MRI · ADC · IVIM · Logistic regression

Introduction

Prostate cancer is the most common cancer in men in more economically developed countries [1], and it is important to accurately detect those lesions that are clinically significant and require treatment. The current standard at our institution consists in taking 10–12 ultrasound-guided transrectal biopsy samples of the prostate and attributing specific Gleason grades to the cancerous tissue on the basis of the underlying glandular patterns. The sum of the primary grade (assigned to the dominant pattern of the tumour) and the secondary grade (assigned to the next most frequent pattern) constitutes the cancer's Gleason score [2]. Lesions with a Gleason score of 7 or greater or a volume of 0.5 cm^3 or greater are typically considered clinically significant [3]. However, there is no consensus on what constitutes clinically significant prostate cancer and other definitions might be used [4].

If cancer is detected, the level of prostate-specific antigen, clinical staging and Gleason score are used to counsel the patient on further management of the disease. Standard treatment options for patients with significant cancer include radical prostatectomy and radiation therapy. Patients whose cancer is not clinically significant may be assigned to active

surveillance (the lesion is monitored frequently for signs of progression) instead of treatment. Current studies indicate that active surveillance is a safe alternative to immediate intervention for low-risk prostate cancer patients [5, 6].

Unfortunately, besides being an invasive procedure with a considerable risk of infection, ultrasound-guided prostate biopsy is prone to sampling error and associated with a relevant risk of under- and overgrading when compared with results obtained following radical prostatectomy [7–9]. While targeted MR and MR/ultrasound fusion biopsy improves accuracy [10], increased efforts are being undertaken to determine the aggressiveness of prostate cancer noninvasively. Diffusion-weighted magnetic resonance imaging (DW-MRI) has emerged as a promising technique to address this problem, and several studies have shown an inverse correlation between Gleason score and apparent diffusion coefficient (ADC) values [11, 12].

As a result of the considerable overlap in ADC values between high- and low-grade lesions, a recent study combined ADC values with parameters derived from other MRI sequences in a logistic regression model to detect cancers with a high Gleason score; the accuracy of the model was found to be similar to the accuracy of experienced radiologists [13]. Another study found that intravoxel incoherent motion (IVIM) parameters (the pure diffusion coefficient D_t, the pseudo-diffusion fraction F_p and the pseudo-diffusion coefficient D_p) derived from a biexponential fit to DW-MRI [14, 15] correlate with the cancer's aggressiveness [16]. In addition, parameters derived from high *b* value DW-MRI (as acquired or following Hessian focality filtering [17]) were found to be useful to distinguish noncancerous disease from prostatic adenocarcinoma [18]. Although all previous studies showed promising results, the association between lesion aggressiveness and values of DW-MRI parameters is not sufficiently specific to allow application to individual patients. Therefore, the aim of this study was to apply ADC, IVIM and high *b* value parameter values (analysed separately or combined in a logistic regression model) to improve the differentiation of prostate cancer lesions with high (≥7) and with low (6) Gleason score. A secondary objective of the study was to determine whether parameters derived from DW-MRI correlate with Gleason score.

Materials and methods

This prospective clinical trial was approved by the ethics committee of the Kanton Bern and signed informed consent was obtained from all subjects prior to enrolment. Prostate cancer patients included in the present analysis were part of a larger cohort of prostate and/or bladder cancer patients. The study is registered at ClinicalTrials.gov with identification number NCT00622973.

Study population

The patient cohort for this clinical trial comprised 415 consecutive patients at the University Hospital of Bern who were diagnosed with N0 prostate and/or bladder cancer based on cross-sectional imaging (CT and/or MRI) and were scheduled for radical prostatectomy or cystoprostatovesiculectomy, as well as extended pelvic lymph node dissection, between May 2008 and March 2010.

Out of these patients, 295 were excluded for reasons reported in the study flowchart (Fig. 1). Out of the remaining 120 prostate and/or bladder cancer patients, the present study included the 89 patients with prostate cancer (Gleason score ≥ 6). Statistical analysis included the 84 patients with prostate lesions which were visible on DW-MRI.

MR imaging

Pelvic MRI was performed a median of 12 days before radical prostatectomy (range 1–50 days). A 3-T MR scanner (TIM Trio; Siemens Healthcare, Erlangen, Germany), with two phased-array eight-channel coils placed ventrally and on the back, was used for imaging. Patients received 0.5 ml of glucagon (GlucaGen; Novo Nordisk, Kuesnacht, Switzerland) intravenously before morphological MRI and again 0.5 ml before starting the DW-MRI sequence. Morphological MRI included a coronal 3D T1-weighted sequence (voxel size 0.8 × 0.8 × 0.8 mm), a coronal 3D T2-weighted sequence (voxel size 1.0 × 1.0 × 1.0 mm) and a transverse high-resolution T2-weighted sequence (voxel size 0.4 × 0.4 × 4.0 mm). Functional MRI included DW-MRI in the transverse plane, covering the entire pelvis from the aortic bifurcation to the pubic symphysis; sequence parameters are reported in Table 1. Respiratory triggering was applied to the initial 18 % of imaged patients; after a qualitative assessment established that respiratory triggering had little impact on image quality, the remaining patients were imaged without respiratory triggering. The acquisition time was approximately 8, 7, 5 and 5 min for the T1-weighted, T2-weighted, high-resolution T2 and DW-MRI sequence, respectively.

Image analysis

After patient enrolment was closed, all acquired images were anonymised and analysed by three independent readers with less than 1, 1 and 2 years of experience with prostate MRI, respectively. The readers knew that patients were scheduled for radical prostatectomy but were blinded to any other clinical or histopathologic information and were not involved in the selection or care of the patients.

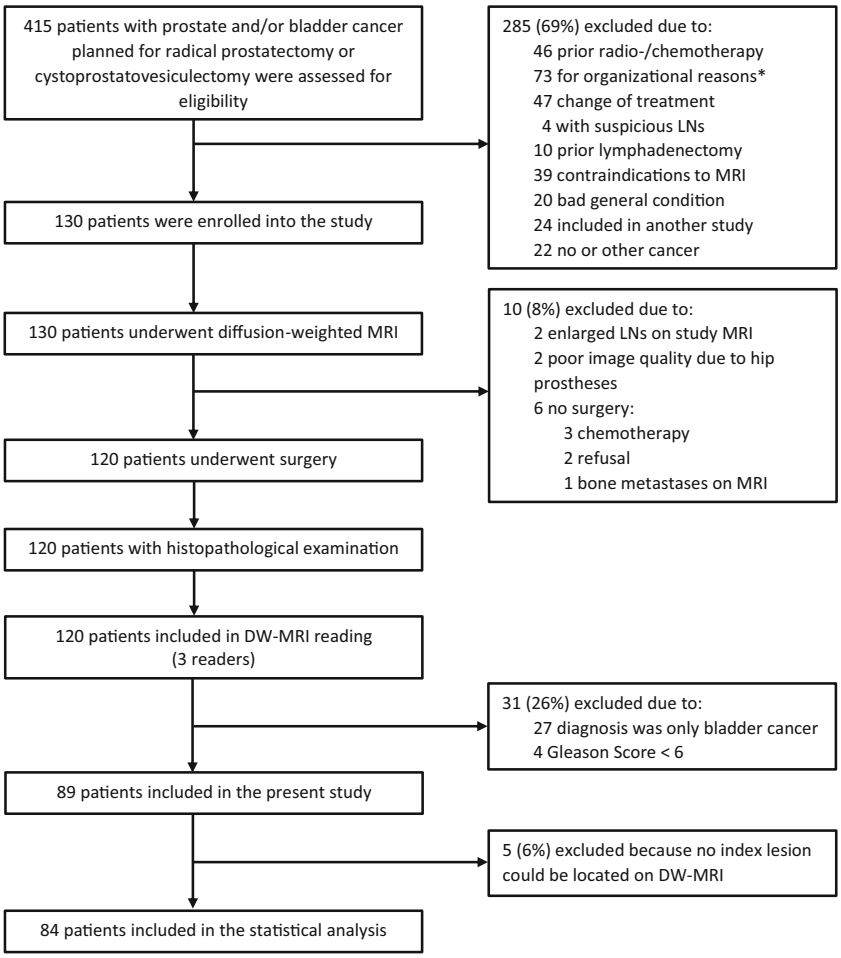

Fig. 1 Flowchart of study population. *LN* lymph node. *Organizational reasons included prostatectomy being performed sooner than expected and non-availability of a drug used in the second part of the trial (not described in the present manuscript)

Table 1 Parameters of the diffusion-weighted magnetic resonance imaging sequence

Sequence	Spin-echo EPI
Repetition time (ms)	2600
Echo time (ms)	58
No. of slices	51
Slice thickness (mm)	4
Slice gap (mm)	0.8
Bandwidth (Hz per pixel)	2300
No. of gradient directions	3
No. of averages	3
b values (s/mm^2)	0, 10, 20, 50, 130, 270, 500, 900
Field of view (mm^2)	330 × 330
Matrix	128 × 128
Parallel acceleration factor	GRAPPA = 3
Acquisition voxel size (mm^3)	2.58 × 2.58 × 4
Reconstruction voxel size (mm^3)	2.58 × 2.58 × 4

GRAPPA generalised autocalibrating partially parallel acquisition

The readers located the index lesion based on the DW-MR images and corresponding ADC maps (as computed by the scanner vendor's software), using the T2-weighted images for morphological reference. The index lesion was defined as a hyperintense region on the high *b* value DW-MR image corresponding to a hypointense region on the ADC map. If a region with such characteristics was found, the readers determined the slice of the ADC map where the axial diameter of the lesion was largest and subsequently delineated the lesion, trying to avoid partial volume artefacts, using ImageJ [19]. If only two out of the three readers agreed on the location of the index lesion, that case was not considered further. If none of the three readers agreed on the location of the index lesion, the most likely region of interest (ROI) was chosen by a fourth reader (2 years of experience with prostate MRI) for further processing.

To minimise the effect of inter-reader variability and to delineate the lesion in three dimensions, a semiautomatic segmentation approach was employed by one of the authors (7 years of experience in medical image processing). Thus, the RegionGrowingMacro module in MeVisLab (MeVis Medical Solutions AG & Fraunhofer MEVIS, Bremen, Germany) was used to place seed points at the centre of the lesion; the interval size was increased sequentially starting at

0 % and at steps of 1 % until its size best matched the average ROI by the initial readers. Successive analysis was based solely on the 3D lesion volumes delineated using MeVisLab.

The following parameters were computed for each voxel within a lesion: the ADC value determined by a least-squares fit to the log-transformed diffusion-weighted signal, the values of IVIM parameters (the pure diffusion coefficient D_t, the pseudo-diffusion fraction F_p and the pseudo-diffusion coefficient D_p), the $b = 900$ s/mm^2 image value and the result of the three dimensional Hessian focality filter applied to the $b = 900$ s/mm^2 image [17]. IVIM parameters were determined voxel-wise within each segmented lesion by using a Bayesian probability-based approach [20] to fit a simplified version of Le Bihan's original model:

$$S(b) = S(0)\left[F_p e^{-bD_p} + \left(1-F_p\right)e^{-bD_t}\right]$$

where $S(b)$ is the diffusion-weighted signal acquired at a specific b value. No additional image processing steps were applied before computing the ADC and IVIM parameters; however, the $b = 900$ s/mm^2 images were normalized to a common mean and variance [21, 22] before determining the corresponding image values and Hessian filtering results.

Logistic regression model

A logistic regression model was derived to detect lesions with a Gleason score of 7 or greater. Potential predictor variables were computed for each lesion on the basis of the histograms of ADC and IVIM parameter values. Thus, the mean, skewness and kurtosis of ADC, D_t, F_p and D_p values were determined. The mean value of the $b = 900$ s/mm^2 image and the maximum value of the Hessian focality filter applied to the $b = 900$ s/mm^2 image were considered as well. To reduce the possibility of overfitting the model, the number of predictor variables was limited to three. Tentative logistic regression models were derived using every possible combination of three predictor variables and evaluated using leave-one-out cross-validation (LOO-CV): each model was trained on all but one patient which was later assigned a predicted score; the process was repeated for all patients; finally sensitivity and specificity values for the tentative model were determined. The model associated with the maximum of min(sensitivity,specificity) across all tentative logistic regression models was deemed the best classifier and evaluated further (in case of a draw, the model associated with the maximum of sensitivity + specificity was chosen).

Reference standard: histopathological analysis

Following surgery, all radical prostatectomy specimens were analysed by two uropathologists blinded to clinical and radiological data. Histopathological analysis was carried out as described previously [23]. Collected information about the carcinoma included the maximum lesion diameter, the primary and secondary Gleason grades, the Gleason score for the whole prostate and indications of extraprostatic extension. When a small amount of cancerous tissue with a higher Gleason grade was found, it was included as a tertiary Gleason grade if its extension was less than 5 % and as a secondary grade otherwise. Pathological stage and grade were assigned according to the TNM classification system, 7th edition [24] and the 2005 modified Gleason grading system [25], respectively.

Statistical analysis

On the basis of a previous study [16, Table 3], a large effect size of more than 0.9 was expected for differences in mean ADC and mean D_t values, and a small effect size of approximately 0.3 was expected for differences in mean F_p and D_p values between high- and low-grade lesions. Assuming a significance level of 0.05, a power of 80 %, and an allocation ratio between groups of 2/3, we determined that 40 and 430 patients were needed to detect the large and the small effect sizes, respectively, by a two-sample t test. Thus, the present study was adequately powered to detect statistically significant differences in mean ADC and mean D_t values but not in mean F_p and D_p values. Nevertheless, it was hypothesised that the mean F_p and D_p values could provide valuable information when deriving the logistic regression model. These power analyses were carried out in G*Power [26].

Welch's independent two-sample t test was used to assess whether histogram parameters differed between lesions with a Gleason score of 6 and those with a Gleason score at least 7 (Q–Q plots did not indicate any strong deviation from the assumption of normal residuals). As these tests were used to explore which predictor variables should possibly be included in the logistic regression model, the significance level was set to $P = 0.05$ without correcting for familywise error rate. For each variable we computed the area under the curve (AUC) of a linear classifier based solely on that variable and Spearman's correlation coefficient ρ between the variable and the lesions' Gleason score (lesions were grouped depending on whether they had a Gleason score of 6, 7, 8, 9 or 10). Furthermore, we qualitatively analysed differences in histogram percentiles of ADC and D_t values between lesions with high and with low Gleason score.

The performance of the logistic regression model was evaluated by computing the AUC following LOO-CV and following leave-pair-out cross-validation (LPO-CV) where each pair of positive–negative patients is sequentially excluded from the training data set. Sample sensitivity and specificity values for the logistic regression model were determined by maximizing the minimum between the sensitivity and the specificity value.

Statistical analyses were carried out in Matlab R2015a (Mathworks, Natick, MA).

Results

Histopathology results for the 89 patients (median age 64 years, range 43–80 years) indicated that 26 had a Gleason score of 6 and that 63 had a score of 7 or greater. A summary of the patients' clinical and pathologic characteristics is reported in Table 2. As no index lesion could be located on DW-MRI in five patients (Gleason scores 3 + 3, 3 + 3, 3 + 3, 3 + 4, 3 + 4), these patients were excluded from the statistical analysis. The volume of the segmented lesions ranged from 0.1 to 10 cm^3 (median 0.7 cm^3).

Exemplary diffusion-weighted images of a high- and of a low-grade lesion, together with the corresponding parametric maps, are presented in Fig. 2.

Differences in variables between Gleason score = 6 and ≥7

The mean of ADC and D_t differed significantly ($p < 0.01$) between lesions with a high and with a low Gleason score.

Table 2 Summary of clinical and pathologic characteristics of patients

Clinical characteristics*	
Age (years)	64 (43–80)
Lesion diameter (mm)	17 (2–40)
pT stage‡	
pT1a	2 (2.2)
pT2a	7 (7.9)
pT2b	4 (4.5)
pT2c	40 (44.9)
pT3a	21 (23.6)
pT3b	10 (11.2)
pT4	5 (5.6)
Gleason score‡	
6 (3 + 3)	24 (27.0)
6 (3 + 3 + 4)	2 (2.2)
7 (3 + 4)	39 (43.8)
7 (3 + 4 + 5)	1 (1.1)
7 (4 + 3)	9 (10.1)
8 (3 + 5)	1 (1.1)
8 (4 + 4)	5 (5.6)
8 (5 + 3)	1 (1.1)
9 (4 + 5)	6 (6.7)
10 (5 + 5)	1 (1.1)

*Data are the medians with the range in parentheses

‡ Data are the number of patients with the percentage in parentheses

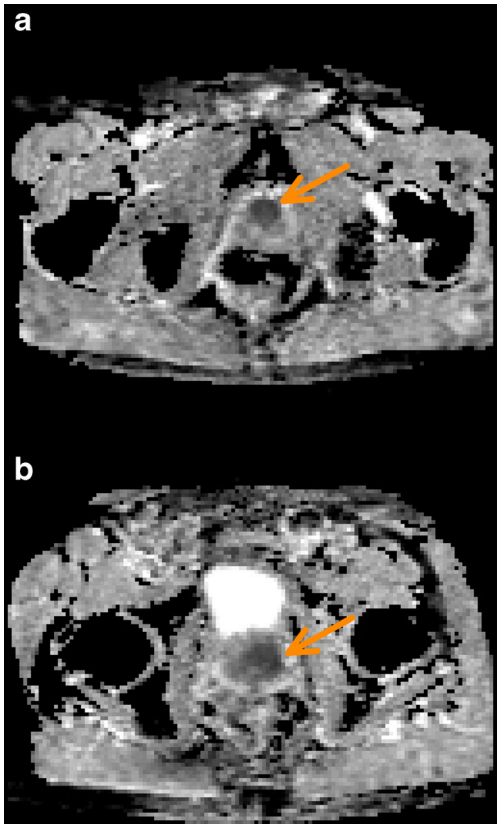

Fig. 2 Exemplary transverse ADC maps of a 67-year-old man with Gleason score 3 + 3 (**a**) and a 57-year-old man with Gleason score 4 + 4 (**b**). The *arrows* point to the index lesions. ADC values in both images range from 0 to 23.5×10^{-4} mm^2/s

Box plots grouped by the lesions' Gleason score are presented in Fig. 3. Results of *t* tests for all variables are reported in Table 3. The highest classification accuracy was also achieved by the mean of ADC (AUC 0.74) and D_t (AUC 0.70). A moderate negative correlation was observed between the lesions' mean ADC ($\rho = -0.47$, $p < 0.01$) and mean D_t ($\rho = -0.40$, $p < 0.01$) values and the Gleason score. The correlation coefficients determined for all other parameters were 0.20 or less. Mean percentiles of ADC and D_t values, grouped by Gleason score, are displayed in Fig. 4. A qualitative analysis of these plots suggests that differences in ADC and D_t values are approximately constant across different histogram percentiles.

Logistic regression model

The mean ADC, mean F_p and mean of the $b = 900$ s/mm^2 image values were determined to be the three predictor variables which maximised the minimum between sensitivity and specificity of the logistic regression classifier. The estimated coefficients of each variable, together with the respective standard error, are reported in Table 4. Prior to leave-one-out analysis, the logistic regression model led to an AUC of 0.79

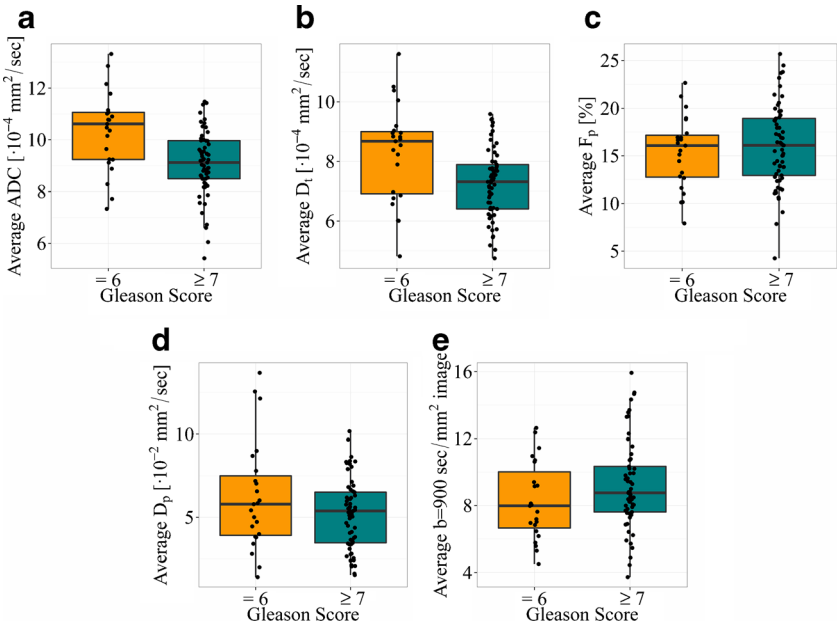

Fig. 3 Data points and box plots of the average ADC (**a**), D_t (**b**), F_p (**c**), D_p (**d**) and $b = 900$ s/mm^2 image (**e**). The *central marks* are the medians, the *horizontal edges of the boxes* are the first (Q_1) and the third (Q_3) data quartiles, and the *whiskers* extend to the most extreme data points not considered outliers i.e. within 1.5 times the distance Q_3–Q_1

(95 % CI 0.64–0.89), a sensitivity of 85 % and a specificity of 74 %. After LOO-CV, the AUC of the classifier was 0.74 (95 % CI 0.58–0.86), the sensitivity 77 % and the specificity 74 %. After LPO-CV, the AUC of the classifier was 0.75 (95 % CI 0.73–0.77), the sensitivity 76 % and the specificity 74 %. Receiver operating characteristic curves of the logistic regression model prior to and with LOO/LPO-CV are presented in Fig. 5.

Discussion

It is well known that parameters derived from DW-MRI (e.g. the ADC) correlate with the aggressiveness of prostate cancer [11, 12]. However, as a result of the considerable overlap in parameter values between high- and low-grade lesions, Gleason scoring in clinical routine still relies on the availability of biopsy samples [27, 28]. This study compared the accuracy of ADC,

Table 3 Quantitative analysis of the individual variables

Variable	95 % CI of expected difference	*p* of *t* test	AUC (95 % CI)	Correlation coefficient
Mean ADC (mm^2/s)	$(5.43, 19.8) \times 10^{-5}$	<0.01‡	0.74 (0.59, 0.85)	−0.47
Mean D_t (mm^2/s)	$(2.90, 18.0) \times 10^{-5}$	<0.01‡	0.70 (0.54, 0.83)	−0.40
Mean F_p (%)	−0.0258, 0.0129	0.51	0.54 (0.40, 0.67)	0.03
Mean D_p (mm^2/s)	−0.00466, 0.0252	0.17	0.58 (0.42, 0.72)	−0.19
Skewness ADC	−0.444, 0.260	0.60	0.50 (0.36, 0.64)	0.11
Skewness D_t	−0.257, 0.401	0.66	0.55 (0.41, 0.68)	−0.08
Skewness F_p	−0.983, 0.237	0.22	0.56 (0.39, 0.71)	0.15
Skewness D_p	−0.309, 0.342	0.92	0.51 (0.37, 0.65)	0.08
Kurtosis ADC	−2.32, 0.539	0.22	0.58 (0.43, 0.72)	0.13
Kurtosis D_t	−1.95, 0.758	0.38	0.59 (0.44, 0.72)	0.14
Kurtosis F_p	−3.09, 2.28	0.76	0.54 (0.39, 0.70)	0.10
Kurtosis D_p	−1.23, 1.33	0.93	0.51 (0.35, 0.65)	0.08
Mean $b = 900$	−21.4, 2.36	0.11	0.61 (0.47, 0.75)	0.19
Max Hessian filter $b = 900$	−12.1, 1.67	0.13	0.56 (0.42, 0.69)	0.20

95 % confidence interval (CI) and *p* value of the *t* test used to assess differences in means between Gleason score = 6 and Gleason score ≥ 7 lesions (‡ indicates a statistically significant difference at the level $P = 0.05$), the area under the curve (AUC) and its 95 % CI of a lesion classifier based solely on one variable, and the correlation coefficient between variable and Gleason score

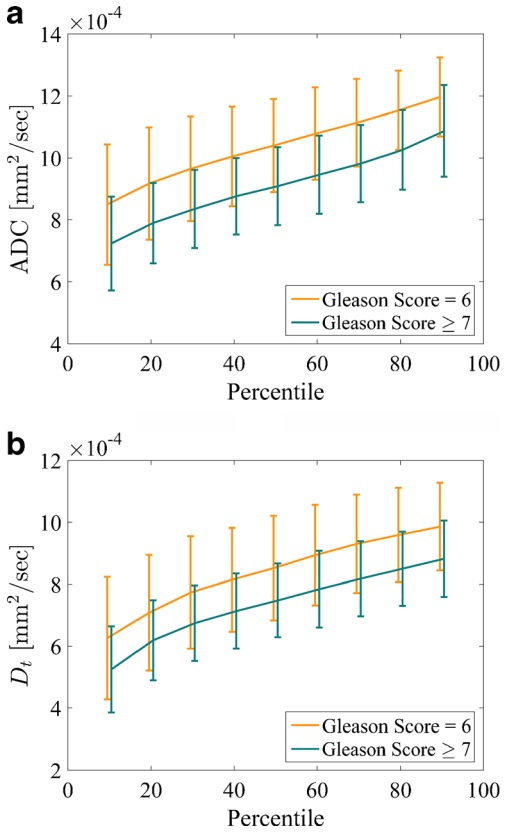

Fig. 4 Comparison of histogram percentiles of **a** ADC and **b** D_t values. The *error bars* represent the mean value ± standard deviation of percentiles across lesions with a Gleason score = 6 and ≥7

IVIM and high b value parameter values and analysed whether results could be improved by deriving a logistic regression model. The only parameters which differed significantly between high- and low-grade lesions were the mean ADC and the mean D_t values ($p < 0.01$). A logistic regression model combining the mean ADC, mean F_p and mean of the $b = 900$ s/mm^2 image values led to an AUC of 0.74 following leave-one-out cross-validation (AUC 0.79 prior to LOO-CV). The classification accuracy of this logistic regression model was comparable to using mean ADC or mean D_t values alone (AUC 0.74 and 0.70, respectively).

A different logistic regression model combining mean ADC value, normalised T2-weighted signal and maximum enhancement measured by dynamic contrast-enhanced (DCE) MRI was recently proposed to classify prostate cancer

Table 4 Estimated coefficients of the logistic regression model (Log. Reg. Coeff.) and corresponding standard errors (Std. Err.)

Variable	Log. Reg. Coeff.	Std. Err.
Intercept β_0	5.61	2.55
Mean ADC	-8.70×10^3	2.54×10^3
Mean F_p	13.2	7.40
Mean $b = 900$	0.0198	0.0129

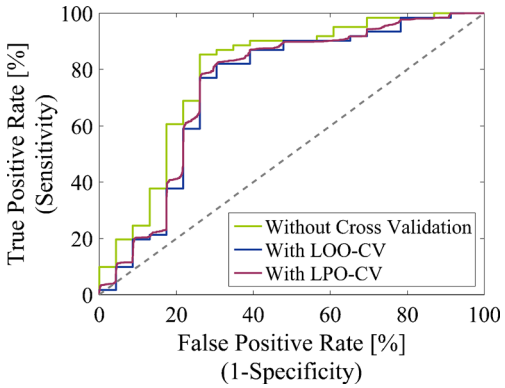

Fig. 5 Receiver operating characteristic (ROC) curves which illustrate the performance of the logistic regression model when distinguishing between Gleason score = 6 and Gleason score ≥7 lesions; without cross-validation, with leave-one-out cross-validation (LOO-CV) and with leave-pair-out cross-validation (LPO-CV)

within the transition zone [13]. When detecting lesions with a Gleason score of 4 + 3 or greater and a diameter greater than 6 mm, the model led to an AUC of 0.71 (95 % CI 0.58–0.84) following leave-one-out regression analysis in 70 patients. While this study focused on lesions in the transition zone, these results appear comparable to the ones obtained in our study. This study suggests that overall the logistic regression model provided little benefit compared with a classifier based on the normalised T2 signal alone. In general, there is currently no consensus about whether bi- or trivariate modelling improves classification performance compared with using e.g. only T2 values in the transition zone and ADC values in the peripheral zone [29, 30]. Another study on 25 patients derived a logistic regression model based on ADC, T2 and volume transfer constant K_{trans} (measured by DCE-MRI) to detect prostate cancer in the peripheral zone (AUC 0.71, 95 % CI 0.69–0.72). The logistic regression model led to a higher AUC compared with classification based solely on measured ADC values but this difference was not significant [31]. A higher AUC of 0.90 (95 % CI 0.83–0.96) was reported by a study combining ADC and DCE-MRI parameters to derive a logistic regression model based on a small sample of 34 prostate cancer patients with lesions in the peripheral and in the transition zone [32]. However, malignant lesions where considered significant irrespective of their Gleason score, which may explain the classifier's relatively high accuracy and limits the clinical value of the study.

A previous study analysing ADC and IVIM parameters in malignant prostate lesions determined significantly different ADC- and D_t-related measurements in high-grade lesions compared with low-grade lesions: lower mean, median, 10th and 75th percentile and higher skewness and kurtosis [16]. In our study we similarly determined significantly lower mean ADC and D_t values in high-grade lesions; however, we found no significant difference in their skewness and kurtosis values. This might be related to the resolution of our diffusion-

weighted images being insufficient to precisely compute higher statistical moments in small lesions or to using the Gleason score of the whole prostate as reference standard. The lack of significant differences between the "shapes" of the histograms was also reflected by relatively constant differences across the computed percentiles of ADC and D_t values. Therefore, our study suggests that measuring the histogram percentiles of ADC and D_t values may not provide any additional information when discriminating between high- and low-grade lesions compared with simple measurements of the mean values of these parameters. Whether mean ADC (as determined in our study) or mean D_t (as suggested by the previous study [16]) is associated with the highest accuracy when detecting high-grade lesions remains an open question. While a previous study suggested the importance of high *b* value images (as acquired or following Hessian focality filtering) to distinguish non-cancerous disease from prostatic adenocarcinoma in our study these parameters were found to provide little additional information when discriminating between high- and low-grade malignant lesions. Nevertheless, it is interesting to note that the mean of the high *b* value images, which have been shown to increase readers' sensitivity [33], was included as a discriminatory var iable in the logistic regression model.

We acknowledge the following limitations. The study was carried out at a single centre which may limit the external reproducibility of results. The number of included patients was comparable to similar studies but still insufficient to analyse additional potential predictor variables, such as the intensity of T2-weighted images, without possibly overfitting the logistic regression model. The number of patients was also insufficient to stratify results by the lesions' location within the prostate. On MRI we analysed the index lesion, defined as the largest most likely cancerous area, while the Gleason score determined by histopathology referred to the whole carcinoma. This might have been a source of bias in our results. Further bias might have been introduced by imaging some patients with and some without respiratory triggering (although no qualitative difference in the acquired images was evident) and by inaccuracies in the delineation of the index lesion.

In conclusion, the present study supports previous research indicating that ADC values differ between high- and low-grade prostate cancer. However, including information provided by IVIM and high *b* value parameters did not lead to improved classification results. Given the large overlap in diffusion coefficients between high- and low-grade prostate cancers, the noninvasive diagnosis of individual patients by DW-MRI in clinical routine does not appear to be feasible yet.

Acknowledgements The scientific guarantor of this publication is Harriet C. Thoeny. The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. This study has received funding by the Swiss National Science Foundation (grant 320000-113512); Nano-Tera (RTD: 20NA21_145919); Carigest (Geneva, Switzerland), representing an anonymous donor; Maiores Foundation; Propter Homines Foundation; Kurt and Senta Herrmann Foundation; and Foundation Fürstlicher Kommerzienrat Guido Feger. One of the authors has significant statistical expertise. Institutional review board approval was obtained. Written informed consent was obtained from all subjects (patients) in this study.

All subjects have been previously reported in Bains LJ et al., J. Urol. (2014); Thoeny HC et al., Radiology (2014); Birkhäuser FD et al., Eur. Urol. (2013); Triantafyllou M et al., Eur. J. Cancer (2013); Froehlich JM. et al., Contrast. Media Mol. Imaging (2012).

Methodology: prospective, diagnostic or prognostic study, performed at one institution.

References

1. GLOBOCAN (2012) Estimated cancer incidence, mortality and prevalence worldwide in 2012. http://globocan.iarc.fr/Pages/fact_sheets_population.aspx. Accessed 17 Mar 2016
2. Epstein JI (2010) An update of the Gleason grading system. J Urol 183:433–440
3. Vargas HA, Hötker AM, Goldman DA et al (2016) Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference. Eur Radiol 26:1606–1612
4. Ploussard G, Epstein JI, Montironi R et al (2011) The contemporary concept of significant versus insignificant prostate cancer. Eur Urol 60:291–303
5. Tosoian JJ, Trock BJ, Landis P et al (2011) Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 29:2185–2190
6. Eggener SE, Mueller A, Berglund RK et al (2013) A multi-institutional evaluation of active surveillance for low risk prostate cancer. J Urol 189:S19–25
7. Shapiro RH, Johnstone PAS (2012) Risk of Gleason grade inaccuracies in prostate cancer patients eligible for active surveillance. Urology 80:661–666
8. Turkbey B, Mani H, Aras O et al (2013) Prostate cancer: can multiparametric MR imaging help identify patients who are candidates for active surveillance? Radiology 268:144–152
9. Gondo T, Hricak H, Sala E et al (2014) Multiparametric 3T MRI for the prediction of pathological downgrading after radical prostatectomy in patients with biopsy-proven Gleason score 3 + 4 prostate cancer. Eur Radiol 24:3161–3170
10. Siddiqui MM, Rais-Bahrami S, Turkbey B et al (2015) Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. JAMA 313:390–397
11. deSouza NM, Riches SF, Vanas NJ et al (2008) Diffusion-weighted magnetic resonance imaging: a potential non-invasive marker of tumour aggressiveness in localized prostate cancer. Clin Radiol 63:774–782
12. Turkbey B, Shah VP, Pang Y et al (2011) Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images? Radiology 258:488–495
13. Dikaios N, Alkalbani J, Sidhu HS et al (2015) Logistic regression model for diagnosis of transition zone prostate cancer on multi-parametric MRI. Eur Radiol 25:523–532
14. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M (1986) MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161:401–407
15. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M (1988) Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 168:497–505

16. Zhang YD, Wang Q, Wu CJ et al (2015) The histogram analysis of diffusion-weighted intravoxel incoherent motion (IVIM) imaging for differentiating the Gleason grade of prostate cancer. Eur Radiol 25:994–1004
17. Li Q, Sone S, Doi K (2003) Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Med Phys 30:2040–2051
18. Litjens GJS, Elliott R, Shih NN et al (2016) Computer-extracted features can distinguish noncancerous confounding disease from prostatic adenocarcinoma at multiparametric MR imaging. Radiology 278:135–145
19. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675
20. Barbieri S, Donati OF, Froehlich JM, Thoeny HC (2016) Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs. Magn Reson Med 75: 2175–2184
21. Crimi A, Commowick O, Ferré JC, Maarouf A, Edan G, Barillot C (2013) Semi-automatic classification of lesion patterns in patients with clinically isolated syndrome. International symposium on biomedical imaging: from nano to macro (San Francisco, United States):1102–1105
22. Nyúl LG, Udupa JK, Zhang X (2000) New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 19:143–150
23. Fleischmann A, Schobinger S, Schumacher M, Thalmann GN, Studer UE (2009) Survival in surgically treated, nodal positive prostate cancer patients is predicted by histopathological characteristics of the primary tumor and its lymph node metastases. Prostate 69:352–362
24. Sobin L, Gospodarowicz M, Wittekind C (2009) TNM classification of malignant tumors, 7th edn. Wiley-Blackwell, Hoboken
25. Epstein JI, Allsbrook WC, Amin MB, Egevad LL, ISUPGC (2005) The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol 29:1228–1242
26. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191
27. Hambrock T, Somford DM, Huisman HJ et al (2011) Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. Radiology 259: 453–461
28. Peng Y, Jiang Y, Yang C et al (2013) Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score-a computer-aided diagnosis development study. Radiology 267:787–796
29. Dikaios N, Alkalbani J, Abd-Alazeez M et al (2015) Zone-specific logistic regression models improve classification of prostate cancer on multi-parametric MRI. Eur Radiol 25:2727–2737
30. Vos EK, Kobus T, Litjens GJS et al (2015) Multiparametric magnetic resonance imaging for discriminating low-grade from high-grade prostate cancer. Invest Radiol 50:490–497
31. Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA (2009) Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. J Magn Reson Imaging 30:327–334
32. Hambrock T, Vos PC, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ (2013) Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging-effect on observer performance. Radiology 266:521–530
33. Rosenkrantz AB, Kim S, Campbell N, Gaing B, Deng FM, Taneja SS (2015) Transition zone prostate cancer: revisiting the role of multiparametric MRI at 3 T. AJR Am J Roentgenol 204: W266–W272