Multiparametric MRI for focal dose escalation in prostate cancer radiotherapy

Schie, M.A. van

Citation

Schie, M. A. van. (2021, June 30). *Multiparametric MRI for focal dose escalation in prostate cancer radiotherapy*. Retrieved from https://hdl.handle.net/1887/3192801

Cover Page

The handle https://hdl.handle.net/1887/3192801 holds various files of this Leiden University dissertation.

Author: Schie, M.A. van
Title: Multiparametric MRI for focal dose escalation in prostate cancer radiotherapy
Issue Date: 2021-06-30

2

Contouring of prostate tumors on multiparametric MRI: evaluation of clinical delineations in a multicenter radiotherapy trial

Marcel A. van Schie

Cuong V. Dinh

Petra J. van Houdt

Floris J. Pos

Stijn W.T.J.P. Heijmink

Linda G.W. Kerkmeijer

Alexis N.T.J. Kotte

Raymond Oyen

Karin Haustermans

Uulke A. van der Heide

Abstract

Purpose

To date no guidelines are available for contouring prostate cancer inside the gland, as visible on multiparametric (mp-) MRI. We assessed inter-institutional differences in interpretation of mp-MRI in the multicenter phase III FLAME trial.

Materials and methods

We analyzed clinical delineations on mp-MRI and clinical characteristics from 260 patients across three institutes. We performed a logistic regression analysis to examine each institute's weighting of T2w, ADC and K^{trans} intensity maps in the delineation of the cancer. As reviewing of all delineations by an expert panel is not feasible, we made a selection based on discrepancies between a published tumor probability (TP) model and each institute's clinical delineations using Areas Under the ROC Curve (AUC) analysis.

Results

Regression coefficients for the three institutes were -0.07, -0.27 and -0.11 for T2w, -1.96, -0.53 and -0.65 for ADC and 0.15, 0.20 and 0.62 for K^{trans}, with significant differences between institutes for ADC and K^{trans}. AUC analysis showed median AUC values of 0.92, 0.80 and 0.79. Five patients with lowest AUC values were reviewed by a uroradiologist.

Conclusions

Regression coefficients revealed considerably different interpretations of mp-MRI in tumor contouring between institutes and demonstrated the need for contouring guidelines. Based on AUC values outlying delineations could efficiently be identified for review.

Introduction

Radiotherapy is one of the standard treatment options for prostate cancer. Although it has been shown that tumor foci are non-uniformly distributed over the prostate (Chen *et al.* 2000), the prostate is usually irradiated with a more or less homogeneous dose distribution. Local recurrence of the disease has been observed at the original location of the tumor, suggesting an insufficient radiation dose at that location (Cellini *et al.* 2002, Pucar *et al.* 2007). As dose escalation to the entire gland would likely increase treatment-related toxicity, a focal dose escalation was proposed (Lips *et al.* 2011). Recently, accrual of patients in the FLAME trial (clinicaltrials.gov identifier NTC01168479) was ended. This large multi-center single-blinded randomized controlled phase III trial aimed to investigate the clinical benefit of focal escalation of the radiation dose to the visible cancer to 95 Gy. This required delineation of the tumor as visible on multiparametric (mp-) MRI, consisting of a T2-weighted (T2w) scan, a diffusion-weighted MRI (DWI) and a dynamic contrast-enhanced (DCE) MRI.

Guidelines on detection, localization, characterization and risk stratification of suspected prostate cancer using recommended mp-MRI were published in the Prostate Imaging – Reporting And Data System (PI-RADS) in 2012, and updated to PI-RADS v2 in 2015 (Barentsz *et al.* 2012, Dickinson *et al.* 2013, Weinreb *et al.* 2016). These guidelines were however not available when the majority of patients were included in the FLAME trial. Moreover, guidelines on contouring of prostate tumors based on mp-MRI are not available to date. Steenbergen *et al.* showed the large inter-observer variability that exists in a prostate tumor delineation study using mp-MRI (Steenbergen *et al.* 2015). Such variability can also be expected in the FLAME trial. In the absence of guidelines, institutional differences in contouring practice caused by differences in interpretation and weighting of the various sequences in mp-MRI scans may have occurred as well.

In this work we investigated the contours of prostate tumors in the FLAME trial, focusing on the weighting of the individual mp-MRI sequences within three institutes. We combined mp-MRI data with the actual clinical delineations to assess the relative contribution of each MRI sequence to the tumor contouring decision. As revision of all contours by a panel of experts is not feasible, we applied a pathology validated model for prostate tumor localization in order to identify cases that showed discrepancies between clinical delineations and MRI data. We selected the patients with the highest inconsistency between predicted tumor location and delineation and reviewed these retrospectively.

2

Materials and methods

Patient characteristics

We analyzed 260 prostate cancer patients who were included in the FLAME trial and randomized in the escalated dose arm. These patients had biopsy-proven prostate cancer, clinically localized intermediate or high-risk disease and no evidence of metastatic disease, according to Ash *et al.* (2000). Institutional review board approval was obtained and all patients provided written informed consent. The patients were treated in three institutes: 160 patients in the University Medical Center Utrecht (UMCU), 54 patients in the Netherlands Cancer Institute (NKI), and 46 patients in the University Hospitals in Leuven (UZL). Thirty-five patients were excluded because they had missing MRI data (15), missing biopsy reports (3), missing delineations (10), registration artifacts (2) or they did not receive the escalated dose (5), which led to analysis of 140, 33 and 52 patients from UMCU, UZL and NKI, respectively.

MRI data

All patients received an mp-MRI exam consisting of a T2w, DWI and DCE sequence. Specifications of the scanner type and sequences for each of the institutes are listed in Table 1. An apparent diffusion coefficient (ADC) map was derived from the b-values of the DWI using a mono-exponential fit. We determined the volume transfer constant (K^{trans}) values with the Tofts model using a population-based arterial input function (Tofts *et al.* 1999, Murase *et al.* 2004). Within each institute a radiation oncologist in consultation with a radiologist had prospectively delineated the prostate and all tumors visible on mp-MRI.

Image processing

We processed the MRI data according to the method of Dinh *et al.* (2017). To minimize the impact of differences between acquisition protocols among the participating institutes, this method applies a normalization of T2w and K^{trans} to the median signal intensity in the peripheral zone (PZ). Since no PZ delineations were made in our cohort, we assumed that 75% of the prostate volume was PZ tissue and 25% central gland (Weinreb *et al.* 2016). The T2w signal in the central gland tends to be lower than in the PZ, while the K^{trans} is higher (Kayhan *et*

Table 1. mp-MRI parameters per institute.

MRI parameters	I-1	I-2	I-3
Scanner type	Siemens 1.5 T	Philips 3.0 T	Philips 3.0 T
Scanner sequence			
T2w			
Pixel size / slice thickness (mm)	0.78 / 4.0	0.40 / 3.0	0.49–1.0 / 2.5–4.0
TE / TR (ms)	124 / 11250	120 / 3126–3828	110–130 / 2698–6717
ADC			
Pixel size / slice thickness (mm)	2.73 / 4.0	1.07–1.11 / 3.0–3.7	1.17–2.38 / 2.5–4.0
TE / TR (ms)	67 / 7110–9900	58–73 / 2712–3500	59–94 / 3119–10036
b-values (s/mm^2)	0, 50, 100, 500, 750, 1000	0, 188, 375, 563, 750 or 200, 400, 600, 800 *	0, 300, 500, 1000 or 0, 300, 1000 *
Ktrans			
Pixel size / slice thickness (mm)	1.37–1.68 / 4.0	1.02–1.36 / 3.0	0.94–2.5 / 2.5–7.0
TE / TR (ms)	1.5 / 4–5	1.9 / 4–5	1.0–1.7 / 4
Dynamic scan time (s)	4.4	2.5	2.5

* b = 0 s/mm^2 was acquired but not used for ADC map calculation.

al. 2010). Considering the upper and lower 75% of the T2w and K^{trans} signal intensity histograms as belonging to the PZ, we normalized the signal to the upper and lower 37.5% respectively.

The data set per patient consisted of normalized T2w and K^{trans} images, ADC, biopsy map and tumor prevalence map, plus the clinically delineated tumor and prostate. From the delineations a labeling mask was derived that contained labels for healthy and tumor tissue within the prostate. The data sets were resampled to an in-plane resolution of 0.49 mm, equal to the resolution of the image data used by Dinh *et al.* (2017), and a slice thickness of 1.0 mm.

Institutional interpretation

We evaluated the institutional differences on interpretation of the mp-MRI with a logistic regression analysis on voxel level of three intensity features, i.e. the T2w, ADC and K^{trans} intensity images. A transformation of each feature *i* to zero mean and unit variance was applied to allow comparison between features. The logistic regression function is:

$$F(x) = \frac{1}{1 + e^{-\left(\beta_0 + \sum \beta_i x_i\right)}} \, , \qquad (1)$$

where $F(x)$ is the probability that voxel x is included in the tumor delineation, x_i is the intensity value of feature i, β_i is the regression coefficient of feature i and represents weight factor, and β_0 is the offset.

Probability model

For the automatic evaluation of the manual delineations we used a published tumor probability (TP) model (Dinh *et al.* 2017), which is a logistic regression model trained on mp-MRI and biopsy data and validated on histology data from 40 patients in two institutes. The coefficients of the TP model are found in Table 2. We combined 29 features from the normalized mp-MRI with biopsy and prevalence information, and applied the TP model to calculate a TP per voxel within the prostate.

For each calculated TP map and labeling mask we derived the Area Under the receiver operating characteristic Curve (AUC). We selected the patients that had a large disagreement between calculated TP map and labeling mask with AUC values below 0.50 and reviewed the

Table 2. *Regression coefficients β_i and offset β_0 of the TP model (Dinh et al. 2017). $G_i(s)$ stands for the Gaussian derivative in direction i, with standard deviation s (in mm).*

Feature description	β	Feature description	β
Offset - β_0	0.1	$T2w - G_y(7.6)$	-18.7
$T2w - Intensity$	-5.3 e-4	$T2w - G_{xx}(7.6)$	-3.0
$T2w - G(3.0)$	-10.9	$T2w - G_{yy}(7.6)$	2.6
$T2w - G_x(3.0)$	2.0	$T2w - G_{xy}(7.6)$	-16.6
$T2w - G_y(3.0)$	-3.3	$T2w - G(12.0)$	39.5
$T2w - G_{xx}(3.0)$	-2.9	$T2w - G_x(12.0)$	-1.4
$T2w - G_{yy}(3.0)$	-0.9	$T2w - G_y(12.0)$	-6.4
$T2w - G_{xy}(3.0)$	-4.0	$T2w - G_{xx}(12.0)$	3.1
$T2w - G(4.8)$	-16.2	$T2w - G_{yy}(12.0)$	-3.5
$T2w - G_x(4.8)$	-3.8	$T2w - G_{xy}(12.0)$	-6.9
$T2w - G_y(4.8)$	-7.8	$ADC - Intensity$	-2753
$T2w - G_{xx}(4.8)$	2.2	$ADC - Blobness$	-376.8
$T2w - G_{yy}(4.8)$	-0.1	$K^{trans} - Intensity$	0.3
$T2w - G_{xy}(4.8)$	-9.1	$K^{trans} - Blobness$	0.2
$T2w - G (7.6)$	-13.2	Biopsy map	1.3
$T2w - G_x(7.6)$	4.1	Prevalence map	0.1

clinical delineations. For each of the review cases we related the TP map with the clinical delineations and described the likely cause of disagreement.

Statistics

We tested patient characteristics and properties of the delineations as well as logistic regression coefficients for statistically significant difference on an institutional level. Median age, iPSA, prostate volume, tumor volume and regression coefficients were pairwise tested with a Mann-Whitney U test at a significance level of $\alpha = 0.05$. Clinical T stage, Gleason Score, number of tumors per patient and tumor location were pairwise tested with a Fisher exact test. Statistical tests were performed with the Statistics and Machine Learning Toolbox in MATLAB (version R2017a, MathWorks, Natick, MA). A Bonferroni correction was applied to correct for multiple testing.

2

Results

Table 3 shows the patient characteristics and delineation properties for the three institutes. We found significant differences in age (between institute (I-) 1 and I-2/I-3), iPSA (between institute I-1 and I-2/I-3), number of delineated tumors per patient (between I-1 and I-3) and tumor volume (between I-1 and I-2/I-3). No significant differences were observed for clinical T stage, Gleason Score, tumor location and prostate volume.

Figure 1 displays the obtained coefficients β_i from equation (1). Mean coefficients for I-1, I-2 and I-3 were -0.07, -0.27 and -0.11 for T2w, -1.96, -0.53 and -0.65 for ADC and 0.15, 0.20 and 0.62 for K^{trans}. The strongest negative coefficient was observed for ADC in I-1, the strongest positive coefficient for K^{trans} in I-3. All institutes weighted the ADC map the most. Comparing within each MRI sequence, T2w, ADC and K^{trans} were weighted the most by I-2, I-1 and I-3, respectively. Statistically significant differences were found between I-1 and I-2/I-3 for ADC and between I-1/I-2 and I-3 for K^{trans} ($p < 0.001$). This demonstrates a different interpretation of MRI sequences among the three institutes.

Median AUC values per institute were 0.92, 0.80 and 0.79, and ranged between $0.50 - 0.99$, $0.47 - 0.98$ and $0.19 - 0.98$, for I-1, I-2 and I-3 respectively. A histogram of the distribution of calculated AUC values is shown in Figure 2. We found a majority of 92% with AUC values above

0.60. Five delineations scored an AUC below 0.50. These were reviewed as cases R-1 to R-5 by a uroradiologist with 14 years of experience. Figure 3 shows the T2w, ADC and K^{trans}, biopsy, prevalence and the obtained TP map of R-1 to R-5 at a representative slice in the prostate. An example patient P-1 with a high AUC is added for comparison.

Table 3. Patient characteristics and delineation properties per institute. Pairwise p-values between institutes are reported, significant values are printed bold (Bonferroni correction for 24 tests). Median values are reported for age, iPSA, prostate volume and tumor volume; mean values for number of tumors per patient. Numbers in parentheses represent range.

|	I-1	I-2	I-3	p-value		
				p_{1-2}	p_{2-3}	p_{1-3}
Patient characteristics						
Age (y)	75 (64–82)	68 (51–78)	71 (56–80)	**< 0.001**	0.008	**0.001**
iPSA (ng/mL)	8.0 (1.8–29.0)	12.7 (3.6–44.7)	12.4 (2.6–100)	**0.002**	0.94	**< 0.001**
% Clinical T stage				0.017	0.22	0.033
T1c	3.0	3.9	12.1			
T2a	12.1	9.6	7.9			
T2b	0.0	11.5	4.3			
T2c	18.2	15.4	15.0			
T3a	48.5	30.8	38.6			
T3b	9.1	28.9	21.4			
T4	9.1	0.0	0.7			
% Gleason score				0.028	0.78	0.004
≤ 6	3.0	21.2	21.4			
7	69.7	46.2	40.7			
≥ 8	27.3	32.7	37.9			
% Tumor location				0.17	0.022	0.80
PZ	61.5	45.6	55.9			
PZ and TZ	23.1	41.2	23.9			
TZ	15.4	13.2	20.3			
Delineation characteristics						
Tumors (n)	1.2 (1–3)	1.3 (1–3)	1.6 (1–5)	0.72	0.057	**0.002**
Prostate volume (cm³)	40.4 (16.9–101)	34.8 (17.9–105)	43.8 (13.8–157)	0.23	0.009	0.27
Tumor volume (cm³)	0.69 (0.05–14.2)	3.3 (0.36–27.2)	3.6 (0.11–50.3)	**< 0.001**	0.97	**< 0.001**
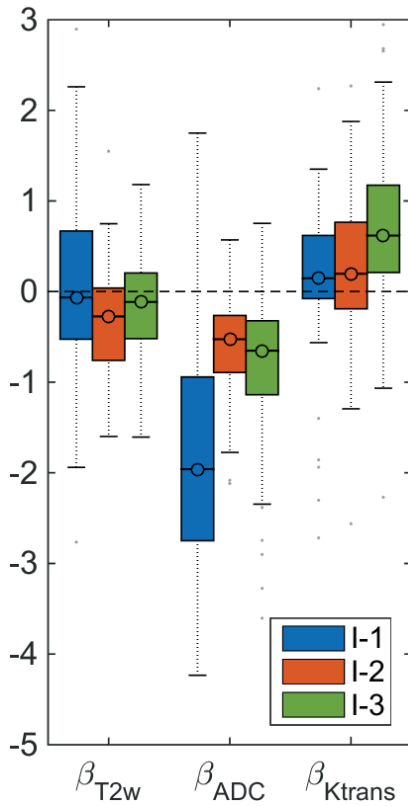
Figure 1. *Box plot of regression coefficients β_i from equation (1) for all patients per institute. Positive regression coefficients represent a direct relation between intensity value and tumor presence, negative coefficients represent an inverse relation.*
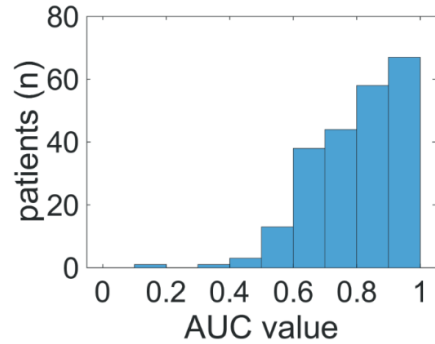

Figure 2. *Histogram of AUC values calculated with the TP model for 225 patients.*

2

R-1 and R-2 showed a transurethral resection of the prostate (TURP) cavity. In the TURP cavity T2w and ADC signal intensities are high and K^{trans} intensities are low, which causes low TP values. In R-1, with an AUC of 0.19, the tumor was delineated around the TURP. Upon review however, suspected tissue was localized in the left PZ, based on low ADC and high K^{trans} signal intensity. The area around the TURP cavity was considered as post-operative tissue response. In R-2 (AUC = 0.31) the clinical tumor delineation enclosed almost the whole prostate, with exception of the dorsolateral left PZ. The reviewer suspected tumor tissue in the latter region as well.
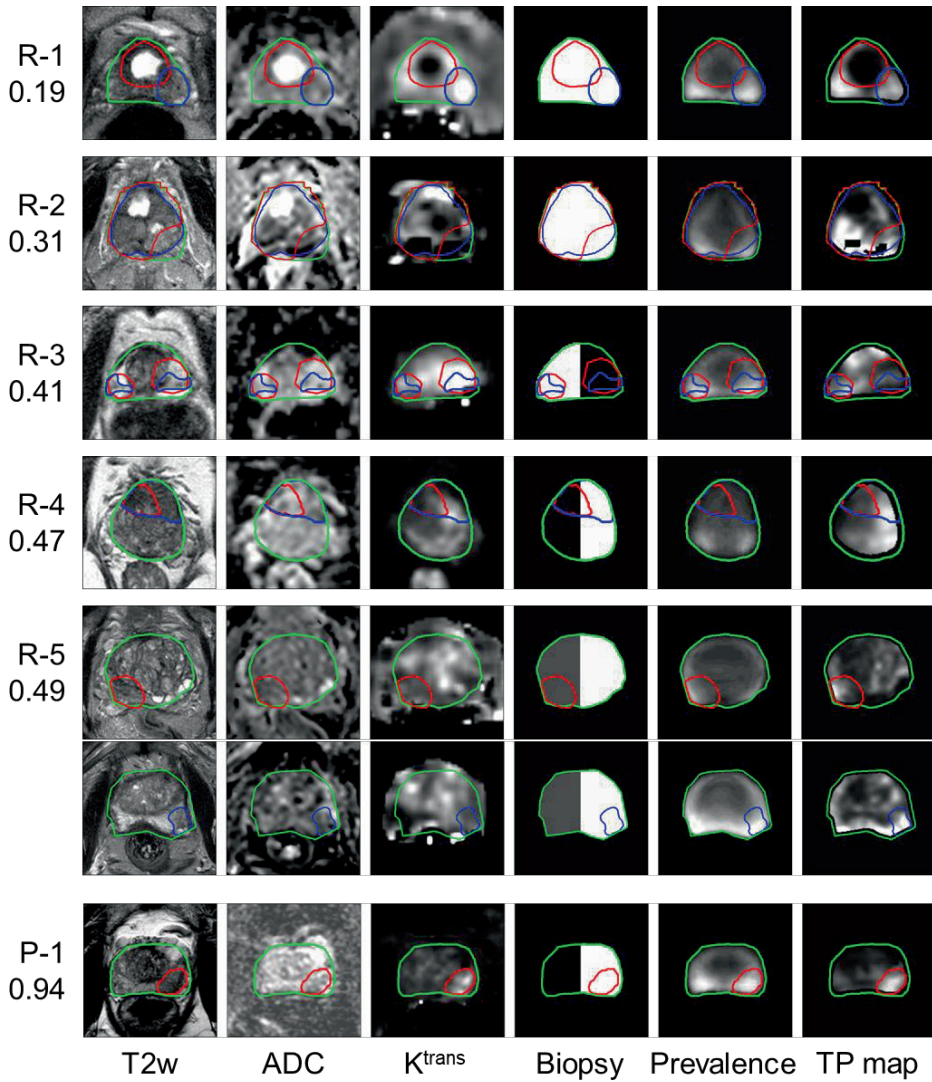
Figure 3. *Overview of T2w, ADC and K^{trans} signal intensity images, biopsy and prevalence map, and the calculated TP map of the reviewed patients R-1 to R-5 and example patient P-1 with corresponding AUC values. Prostates are delineated in green, clinical tumor delineations in red. Tumor delineations after review are blue. The reviewed delineation in R-5 was in a different axial plane and is therefore displayed on two rows.*

The location of the delineations in R-3 and R-4 were consistent with the mp-MRI. In R-3 (AUC = 0.41) the reviewer agreed with the locations of both contours based on T2w and K^{trans}, but the contoured volumes were too large in some parts, while missing parts of the suspected tissue elsewhere. R-4 had an AUC of 0.47. Upon review, the TZ delineation was considered too small and should have included the left ventral part of the TZ as well, based on low ADC and high K^{trans} signal.

In R-5, with an AUC of 0.49, the reviewer regarded the clinical tumor delineation as nonmalignant central gland tissue. Instead, a PI-RADS 2 region on the left apex side on the T2w image was mildly suspect for tumor and supported with positive biopsy map on the left side of the prostate (not indicated in Figure 3). ADC and K^{trans} values did not confirm this finding. However, TP values were moderately high in the clinical tumor delineation, while higher TP values were found in the left apex region, indicated by the reviewer.

2

Discussion

Delineation of the visible cancer on mp-MRI was required for focal dose escalation in the FLAME trial. Because of the absence of contouring guidelines for prostate cancer inside the gland at the time of treatment, the accuracy of these delineations is uncertain. For this reason, we evaluated the delineations of the FLAME trial in the three institutes and observed statistically significant differences between institutes in the weighting of each of the MRI sequences during contouring of the visible cancer.

Guidelines for the detection and localization of prostate cancer were established in 2012 in PI-RADS and have been updated in version 2 in 2015 with a decreased relevance of DCE-MRI (Barentsz *et al.* 2012, Weinreb *et al.* 2016). Several studies have already demonstrated the need for contouring guidelines based on inter-observer studies (Bratan *et al.* 2013, Jung *et al.* 2013, Rischke *et al.* 2013, Anwar *et al.* 2014, Steenbergen *et al.* 2015). With established contouring guidelines, training of radiation oncologists could improve the accuracy of clinical tumor delineations on mp-MRI.

Using logistic regression, we now showed differences in interpretation of mp-MRI on an institutional level. We limited the logistic regression analysis to T2w, ADC and K^{trans} intensity images only. Nonetheless, other MRI features may be considered by the radiation oncologists and radiologists, but these are not made explicit. Structural appearance for example plays an important role as well. However, for the purpose of investigating differences between

institutes, the intensity values allow for a more objective analysis. Therefore, we considered the intensity images as most informative for contouring of the visible cancer in the clinic. Logistic regression performed on an increased number of features would furthermore increase the risk of correlations and compromise the separation between obtained coefficients.

In Table 1 significant differences between institutes are observed for age, iPSA, number and volume of delineated tumors, after Bonferroni correction of the significance level for 24 tests. The significant differences we found for the delineated number and volume of tumors are considered to be of clinical relevance, since they have direct impact on the escalated dose volume and thereby may propagate into tumor control. Age and iPSA were not considered during manual contouring of the tumor, so we regarded the clinical relevance of the significant differences observed for these patient characteristics to be limited.

Retrospective evaluation of 225 clinical delineations by a panel of experts is not feasible in practice. Instead, we applied a published tumor probability model to assess the delineations automatically (Dinh *et al.* 2017). Most models used in other studies are focused on the classification or risk stratification of prostate cancer (Vos *et al.* 2008, Langer *et al.* 2009, Ozer *et al.* 2010, Rouvière *et al.* 2012). Some that do predict tumor probability in the prostate use information from imaging only (Viswanath *et al.* 2012, Groenendaal *et al.* 2012). The TP model that we applied is comparable to the latter, but benefits from the inclusion of clinical information and knowledge of the prevalence of prostate cancer in the gland, that is also available to the radiation oncologist.

As a proof of principle, we reviewed the five patients that showed the largest discrepancies between model predictions and delineations based on their AUC value. We found low AUC values caused by TURP regions for two patients and inaccurate tumor delineations for another two. Also, biopsy information that was contradictive to MRI lowered the AUC for two patients. We found that all the selected patients indeed needed revision of the clinical delineations.

In patients treated with radiotherapy, no whole mount section histology information can be available to validate the delineations. Therefore, discrepancies between the TP model and clinical delineations can be reviewed by an expert, but confirmation with ground truth is not feasible. Discrepancies may be attributed to inaccurate delineations as well as to inaccuracy of the model. Since the model was trained on a multi-institutional dataset with comparable MRI scanners and scanning protocols as used in this study, we attempted to minimize the latter. Limitations in the MRI normalization method could translate into inaccurate model predictions, although the risk is small when the tumor accounts for more than 50% of the PZ area. Because of the absence of ground truth information and possible model inaccuracies we

treated the model as a screening tool to select patients that need further review by experts, rather than a method to improve or even substitute clinical delineations.

The absence of histology information also complicates the assessment of the impact of delineation variation on the delivered focal dose escalation. The high median AUC values suggest however that this impact may be small overall. Still insufficient dose escalation to the tumor tissue and increased dose levels to normal tissue in the prostate may have occurred in some instances.

Conclusions

2

In conclusion, we demonstrated inter-institutional differences in the interpretation of mp-MRI for delineation of the visible cancer. This supports the need for contouring guidelines for prostate cancer inside the gland. AUC values were high in general, suggesting good agreement between TP model predictions and clinical delineations. Observed discrepancies based on low AUC values were a clear indication for further review by experts.