Multiparametric MRI for focal dose escalation in prostate cancer radiotherapy

Schie, M.A. van

4

CHAPTER 4

Repeatability of dose painting by numbers treatment planning in prostate cancer radiotherapy based on multiparametric magnetic resonance imaging

Marcel A. van Schie
Peter Steenbergen
Cuong V. Dinh
Ghazaleh Ghobadi
Petra J. van Houdt
Floris J. Pos
Stijn W.T.J.P. Heijmink
Henk G. van der Poel
Steffen Renisch
Torbjørn Vik
Uulke A. van der Heide

Physics in Medicine & Biology, Volume 62, pp5575–5588 (2017)

Abstract

Purpose

Dose Painting by Numbers (DPbN) refers to a voxel-wise prescription of radiation dose modelled from functional image characteristics, in contrast to dose painting by contours which requires delineations to define the target for dose escalation. The direct relation between functional imaging characteristics and DPbN implies that random variations in images may propagate into the dose distribution. The stability of MR-only prostate cancer treatment planning based on DPbN with respect to these variations is yet unknown. We conducted a test-retest study to investigate the stability of DPbN for prostate cancer in a semi-automated MR-only treatment planning workflow.

Materials and methods

Twelve patients received a multiparametric MRI on two separate days prior to prostatectomy. The tumor probability (TP) within the prostate was derived from image features with a logistic regression model. Dose mapping functions were applied to acquire a DPbN prescription map that served to generate an intensity modulated radiation therapy (IMRT) treatment plan. Dose calculations were done on a pseudo-CT derived from the MRI. The TP and DPbN map and the IMRT dose distribution were compared between both MRI sessions, using the intraclass correlation coefficient (ICC) to quantify repeatability of the planning pipeline. The quality of each treatment plan was measured with a quality factor (QF).

Results

Median ICC values for the TP and DPbN map and the IMRT dose distribution were 0.82, 0.82 and 0.88 for linear dose mapping, and 0.82, 0.84 and 0.94 for square root dose mapping. A median QF of 3.4% was found among all treatment plans.

Conclusions

We demonstrated the stability of DPbN radiotherapy treatment planning in prostate cancer, with excellent overall repeatability and acceptable treatment plan quality. Using validated tumor probability modelling and simple dose mapping techniques it was shown that despite day-to-day variations in imaging data still consistent treatment plans were obtained.

Introduction

In prostate cancer radiotherapy dose escalation to the tumor based on image characteristics is referred to as dose painting. While dose painting is not a standard procedure in prostate cancer treatment (Bauman *et al.* 2013), it has been hypothesized that a clinical benefit without increased toxicity can be achieved with this approach (Pickett *et al.* 1999, van Lin *et al.* 2006, Singh *et al.* 2007, Fonteyne *et al.* 2008). The clinical benefit of focal dose escalation is currently investigated in the FLAME trial, a multicenter phase III randomized clinical trial (Lips *et al.* 2011). In dose painting by contours tumors are manually delineated based on multiparametric (mp-) MRI, consisting of a T2-weighted (T2w), diffusion-weighted (DWI) and a Dynamic Contrast-Enhanced (DCE) sequence (Barentsz *et al.* 2012, Dickinson *et al.* 2013). However, it has been demonstrated by Steenbergen *et al.* (2015) that a large inter-observer variation in manual tumor delineations on mp-MRI exists. Since delineation of the tumor is an inherently binary procedure, variability in tumor delineations may have an impact on the dose coverage of the actual tumor.

As an alternative to manual contouring automated methods to derive a tumor probability (TP) map of the prostate have been developed (Groenendaal *et al.* 2012, Viswanath *et al.* 2012, Vos *et al.* 2012, Dinh *et al.* 2016, Dinh *et al.* 2017). These maps represent the likelihood of tumor presence in each voxel. We showed earlier that this tumor probability correlates with the level of consensus among observers (Dinh *et al.* 2016). Dose Painting by Numbers (DPbN) refers to the concept of deriving a prescription dose from image characteristics directly. For this, dose mapping functions converting characteristics into a dose prescription need to be applied (Bowen *et al.* 2009).

In the current study we incorporated DPbN and tumor probability modelling in the radiotherapy treatment planning pipeline of prostate cancer. Based on dose prescription maps treatment plans were realized with conventional planning objectives and dose-volume constraints. Additional planning objectives allowed us to modulate the heterogeneous dose distribution within the prostate.

While DPbN does not suffer from the inter-observer variation of manual contouring, the MRI data are affected by image noise and day-to-day patient variation. The direct relationship between functional imaging characteristics and dose painting by numbers implies that such variation in images may propagate into the dose distribution. The impact of the image data variations and further processing on the treatment plan quality in DPbN is yet unknown. Therefore, we performed a test-retest mp-MRI study to investigate whether variability due to

imaging noise propagates through the planning pipeline and influences the repeatability of semi-automated DPbN treatment planning in an MR-only workflow.

Materials and methods

The repeatability of the proposed semi-automated MR-only workflow was investigated with a test-retest study, scanning each patient two times on a different day with a mp-MRI. For each mp-MRI exam a radiation treatment plan was simulated via four planning stages. A repeatability analysis between both treatment planning sessions was performed for each of the stages in this workflow. The images of the first and second MRI exams were registered with a B-spline deformable image registration.

Patient inclusion

We included twelve patients with biopsy-proven stage T1-T2 prostate cancer (median age 67, range 54 – 71 years) between October 2014 and March 2016. Each patient underwent two mp-MRI examinations with a median interval of 21 days (range 7 – 37 days), in which no change in tumor appearance was assumed. The second MRI examination was followed by a radical prostatectomy. Institutional review board approval was obtained, and all patients provided written informed consent.

Image acquisition

Images were acquired on a 3.0 T Philips Achieva MR scanner (Philips Healthcare, Best, the Netherlands). We scanned nine patients with a 6-channel phased-array coil in combination with an endorectal coil. After a dStream upgrade of the scanner, we scanned the last three patients with a 16-channel dS anterior coil and a 12-channel dS posterior coil.

In accordance with recent recommendations (Barentsz *et al.* 2012, Dickinson *et al.* 2013), the mp-MRI exam included an axial, sagittal and coronal T2-weighted turbo spin echo sequence, a diffusion weighted single-shot echo-planar imaging sequence, and a DCE scan using a 3D spoiled gradient echo sequence. A pre-contrast T1-weighted (T1w) 3D gradient echo sequence

was used to detect hemorrhage areas as a result of preceding diagnostic biopsies. For dose calculations we scanned an mDIXON sequence prior to contrast, from which a pseudo-CT scan was derived using the MRCAT algorithm (Philips Medical Systems MR Finland, Vantaa, Finland).

T2w scans were acquired with a TE/TR of 12/2000–5000 ms, and a reconstructed voxel size of 0.27 x 0.27 x 3.0 mm^3 before and 0.4 x 0.4 x 3.0 mm^3 after dStream upgrade. T1w scans were obtained with a TE/TR of 1.8/3.7 ms, while the mDIXON sequence had a TE1/TE2/TR of 1.16/2.4/3.7 ms. The DWI sequence had a TE/TR of 59/3500 ms, using b-values of 200, 600 and 1000 s/mm^2 to derive an apparent diffusion coefficient (ADC) map using a mono-exponential model (Bammer 2002). The DCE-MRI had a TE/TR of 1.9/4–5 ms and was acquired at a flip angle of 20°, after injection of 15 ml 0.5 M gadolinium-based contrast agent (DOTAREM®, Guerbet, Paris, France). In total the DCE sequence consisted of 110 acquisitions with an interval of 2.7 – 2.9 s. Signal intensities were converted to concentration values using a baseline T1-map obtained from variable flip angle scans at 3°, 6°, 10°, 20° and 30°. (Schabel *et al*. 2008). Volume transfer constant (K^{trans}) values were estimated with the Tofts model using a population-based arterial input function (Tofts *et al.* 1999, Murase 2004).

Any organ motion that occurred between the different mp-MRI sequences was corrected with a rigid registration to the prostate on the axial T2w scan.

Planning pipeline

Data from each examination was processed separately through the planning pipeline. The prostate gland and organs at risk (OARs) were delineated manually using the T2w and pseudo-CT images. In total 30 image features from T2w, ADC and K^{trans} were combined to derive a per-voxel TP map for the prostate using a logistic regression model (Dinh *et al.* 2016, Dinh *et al.* 2017). Prior knowledge about tumor location was included in the TP model in the form of a tumor prevalence map, derived from radical prostatectomy patients (Ou *et al.* 2009). The TP model was trained previously on data of 17 patients from the same institute, and was validated with pathology data using a leave-one-out approach. After resampling of the TP map to a 2-mm cubic voxel grid compatible with treatment planning software, polynomial dose mapping functions were applied to translate the TP map to a dose prescription map (DPbN map).

The pseudo-CT, OAR structure set and DPbN map were imported in the treatment planning software to generate an intensity modulated radiation therapy (IMRT) treatment plan, with heterogeneous dose prescription to the prostate. A schematic overview of the data flow from mp-MRI examination to treatment plan realization is depicted in Figure 1.

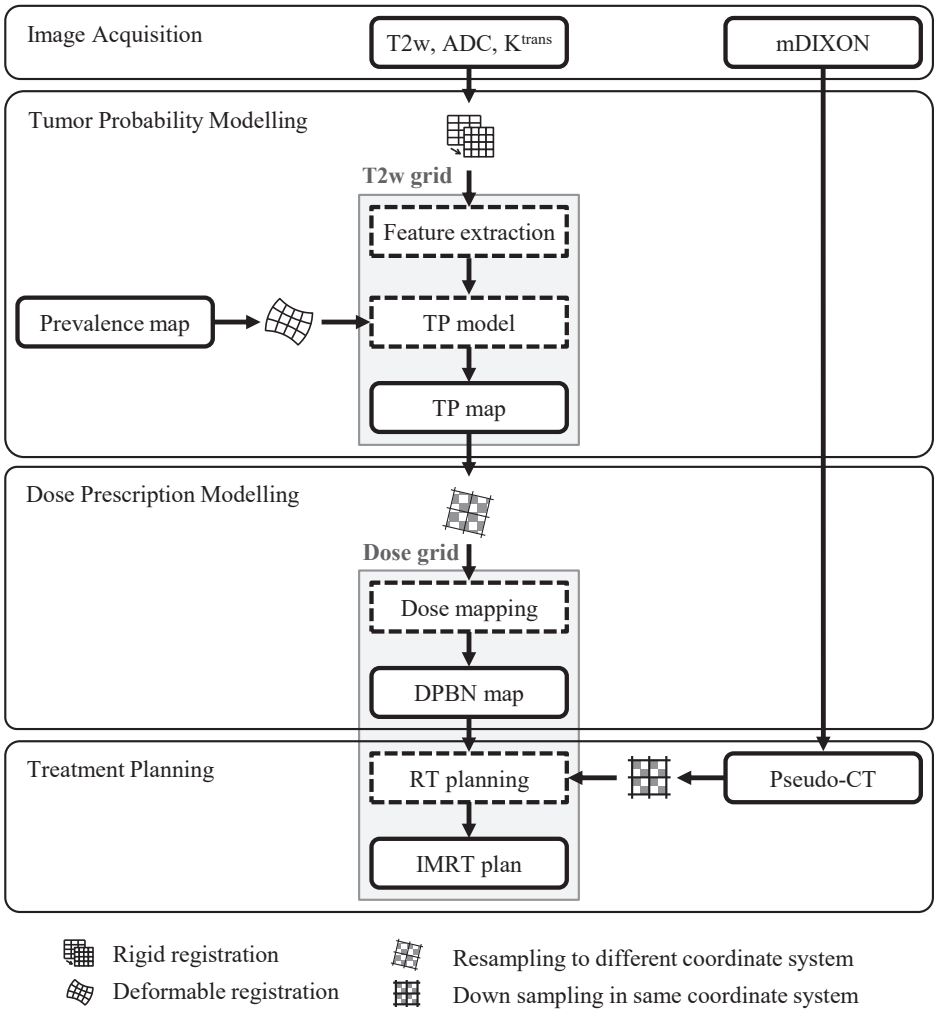

Figure 1. *Schematic overview of the planning pipeline in DPbN prostate cancer treatment. The treatment planning pipeline involves four stages: image acquisition, tumor probability modelling, dose prescription modelling and treatment planning. Processing steps are indicated with dashed contours. Registration and resampling steps to different coordinate systems are indicated with symbols.*

Dose prescription modelling

For each voxel in the prostate the TP was translated into a prescription dose D_{presc} using a polynomial dose mapping function (Bowen *et al.* 2009):

$$D_{presc}^{n} = D_{\min} + \left(D_{\max} - D_{\min}\right) \cdot TP^{n}, \quad (1)$$

with D_{min} and D_{max} the minimum and maximum prescribed dose respectively, and n the polynomial order of the mapping function. DPbN prescription maps were created with values for n of 1 and 0.5, corresponding to linear and square root mappings of TP. D_{min} was set to a safe lower bound of 68 Gy, which reflects the standard treatment in the Dutch dose escalation trial (Peeters *et al.* 2006). A D_{max} of 102 Gy was allowed, which is 107% of 95 Gy and corresponds to the escalated dose to the visible tumor in the FLAME trial (Lips *et al.* 2011). D_{presc} was calculated on a 2-mm cubic voxel grid to be compatible with the treatment planning software. The dose prescription outside the prostate was set to 0 Gy. Modeling of TP and DPbN maps was performed with MATLAB (version R2015b, MathWorks, Natick, MA).

Treatment planning

Apart from the prostate, the following OARs were delineated on the mDIXON images according to the clinical guidelines of our institute: the rectum (up to either the sigmoid or pelvic joints) and anal sphincter, the femoral heads (including articular cartilage) and the bladder. The pseudo-CT, DPbN map and the OAR structure set were imported in a research version of Pinnacle planning software (Pinnacle 9.710, Philips Research, Hamburg, Germany) to establish an IMRT treatment plan. The treatment plan was optimized for a 10 MV step-and-shoot photon beam at seven angles: 210°, 260°, 310°, 0°, 50°, 100°, 150°. The minimum segment area of the multi-leaf collimator was set to 9 cm^2, the minimum number of monitor units per segment to four and the maximum number of segments per beam position to 10. Plan optimization with a maximum of 100 iterations was performed using a standard cost function in combination with two dose painting objective functions f_{min} and $f_{max,}$ summing over all voxels i in the prostate to penalize under- and overdosing of the prescribed dose respectively:

$$f_{\min}\left(D_{plan}\right) = w_{\min} \cdot \sum_i \min\left(\frac{D_{plan,i} - D_{presc,i}}{D_{presc,i}}, 0\right)^2$$
$$f_{\max}\left(D_{plan}\right) = w_{\max} \cdot \sum_i \max\left(\frac{D_{plan,i} - D_{presc,i}}{D_{presc,i}}, 0\right)^2 . \quad (2)$$

Here $w_{\min}$ and $w_{\max}$ are tunable weight factors, and $D_{\text{presc,i}}$ and $D_{\text{plan,i}}$ are the prescribed and planned dose to voxel *i*. The treatment plan was accepted if the dose-volume constraints for OARs in Table 1 were met.

Evaluation of the treatment plan was performed with a quality factor (QF), summing over the normalized absolute differences between planned and prescribed dose to each voxel *i* (Vanderstraeten *et al*. 2006):

$$QF = \frac{1}{n}\sum_i \left|\frac{D_{plan,i} - D_{presc,i}}{D_{presc,i}}\right| \cdot 100\% . \quad (3)$$

To get information about the capability of the treatment planning system to deliver high dose to small regions, a $QF_{95\%}$ was also calculated, which evaluates the treatment plan quality similar to equation (3), but only for voxels that were prescribed at least 95% of the maximum prescribed dose. Within the group of patients we tested both QF and $QF_{95\%}$ for significant difference between linear and square root dose mapping functions and between the first and second planning session with the paired Wilcoxon signed rank test at a 5% significance. We also performed a cross evaluation of the treatment plan quality with xQF and $xQF_{95\%}$, where the dose distribution of the treatment plan from the first planning session was compared with the prescribed dose distribution of the second session and vice versa. This allows to quantify

Table 1. *List of OARs that were delineated prior to treatment planning, together with the dose-volume constraints that were imposed to the treatment planning system to build a clinically acceptable treatment plan.* $V_{X\,Gy}$ *refers to the volume receiving X Gy.*

Organ at Risk	Dose-volume constraint
Rectum + 2 mm	$V_{80\,Gy} \leq 1$ cc
Rectal wall	$V_{64\,Gy} \leq 35$ %
	$V_{75\,Gy} \leq 10$ %
Anal sphincter	$D_{mean} \leq 45$ Gy
Bladder	$V_{80\,Gy} \leq 1$ cc
Femoral head	$D_{max} \leq 50$ Gy
Bowel loop	$D_{max} \leq 68$ Gy

to what extent treatment planning based on a different mp-MRI compromises the plan quality. Treatment plan quality (Q) was visualized with a Q-volume histogram (QVH), displaying Q-values defined as the ratio of planned over prescribed dose.

Between-session registration

For voxel-level evaluation of the TP, DPbN and IMRT map within the prostate a registration between both scan sessions is required. A B-spline deformable registration between the delineated prostates on the axial T2w scans of both mp-MRI sessions was used to match and resample the TP, DPbN and IMRT maps from the first planning session to the second. As a consequence of the registration step the DPbN and IMRT maps were upsampled to the higher resolution T2w grid, thereby introducing correlated data. We therefore resampled all maps again to the lower resolution 2-mm cubic voxel grid used for the treatment planning.

Repeatability analysis

Stability and repeatability of the planning pipeline was investigated on the TP, DPbN and IMRT map with the intraclass correlation coefficient (ICC), to quantify the variability between voxels relative to the measurement error (de Vet *et al.* 2006). In this study a one-way random model was used to measure consistency between two measurements, where single values were calculated for each individual patient (Shrout *et al*. 1979, McGraw *et al*. 1996, Raunig *et al.* 2015). The one-way random model was chosen since it effectively models the between-voxels variation as a fixed effect. This holds for an ICC calculated on a single patient where all voxels of interest of the patient are included instead of a random selection. To satisfy the requirement of normally distributed data, repeatability of the TP map was assessed on the logarithm of the TP odds ratio. Hemorrhage areas as a result of biopsies taken prior to the first MRI examination were identified on T1w images. Since these areas shrink from first to second MRI exam, we excluded these regions from the repeatability analysis. In order to investigate to what extent the variability of the mp-MRI image features propagates through the planning pipeline, the repeatability of the 30 individual imaging features that were input for the TP model was assessed with the ICC as well.

Results

To illustrate the planning pipeline, an example of the tT2w, ADC and K^{trans} intensity maps together with the calculated TP, DPbN and IMRT map is given in Figure 2. Slices are taken in the axial direction of the patient from a location in the prostate involving an area suspected of being tumor. The top row depicts the intensity maps of T2w, ADC and K^{trans}. The T2w map was normalized to the 75th percentile of the intensity values within the prostate. Within the prostate, image features from the intensity maps were combined to create the TP map in Figure 2D. A suspicion of tumor can be identified at the left side of the prostate, characterized by low T2w and ADC, and high K^{trans} values. The increased K^{trans} values at the transition zone of the prostate was disproven to be tumor from histopathology, in accordance with the TP map in panel B.

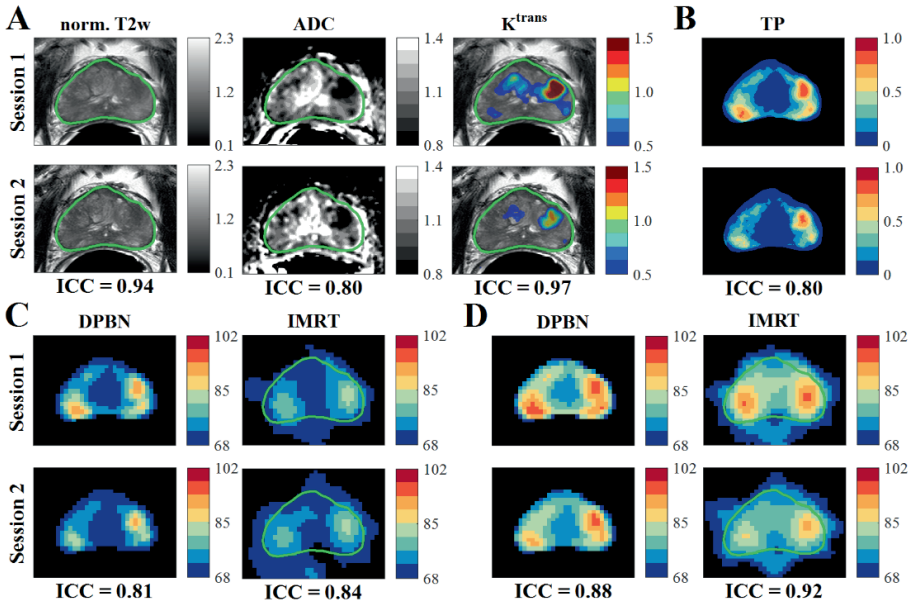

Figure 2. Example of an axial slice through the prostate. In each panel A-D the first (top row) and second planning session (bottom row) are compared. The ICC value is shown for each set of maps. Panel A depicts the mp-MRI: normalized T2w, ADC (in 10^{-3} mm^2/s) and K^{trans} map (in min^{-1}), where K^{trans} is shown as a colorwash on top of the normalized T2w scan. The DPbN map (in Gy) and IMRT plan (in Gy) are shown for linear (panel C) and square root (panel D) dose mapping functions. The delineated prostate is contoured in green.

After resampling of the TP map to the 2-mm cubic dose grid a linear and square root dose mapping function were applied to obtain the DPbN map. TP values of 0 and 1 were mapped to prescription dose levels between 68 and 102 Gy respectively. An example of the two dose mapping functions is shown in Figure 3. The TP histogram at the top shows a majority of voxels with low TP, and only a small fraction of voxels with TP above 0.5. In this example the maximum TP is 0.93, corresponding with 99.0 Gy for linear and 100.5 Gy for square root dose mappings.

An example of IMRT treatment plans in both planning sessions is given for linear dose mapping (Figure 2C) and square root dose mapping (Figure 2D). Smoothing of the prescribed dose due to scatter of photons and dose deposition limitations can be observed in the optimized plan. The quality of the IMRT plan was assessed with QF and $QF_{95\%}$ and visualized with a QVH in Figure 4. The majority of planned dose voxels was between 90% and 110% of the prescribed dose values, as can be observed from the gradient between Q-values of 0.9 and 1.1. For linear and square root dose mappings respectively, 3.4% and 1.5% of the prostate volume received less than 90% of the prescribed dose, while overdosing with more than 110% of the prescribed dose occurred in 2.3% and 6.2% of the voxels. QFs of 3.7% and 4.4% were observed, which is within the 5% treatment planning goal proposed by Duprez *et al.* (2011). $QF_{95\%}$ were 13.8% and 7.5%, indicating a higher agreement of planned with prescribed dose for square root dose mapping.

Comparable QVHs and quality factors were obtained for the other patients in this study, showing good quality of treatment plans based on DPbN. For linear dose mapping, QFs ranged between 2.3% and 4.1% (median 2.9%) and $QF_{95\%}$ values ranged between 5.5% and 17.3% (median 11.6%). For square root dose mapping, QFs between 2.7% and 4.4% (median 3.4%) were observed, whereas $QF_{95\%}$ values were between 4.1% and 11.7% (median 7.2%). Statistical testing with a Wilcoxon signed rank test of both QF and $QF_{95\%}$ revealed a significant difference between linear and square root dose mapping ($p < 0.001$). No significant differences between plans from session 1 and plans from session 2 were observed ($p = 0.80$ for QF and $p = 0.98$ for $QF_{95\%}$). xQF values ranged between 2.1% and 4.4% (median 3.3%) for linear dose mapping, and between 2.9% and 4.4% (median 3.9%) for square root dose mapping. $xQF_{95\%}$ values for linear dose mapping were between 4.8% and 18.1% (median 11.6%), and for square root dose mapping between 3.2% and 15.7% (median 8.0%).

Stability of automated treatment planning based on DPbN was tested with a repeatability analysis and expressed with the ICC. For ICC classification we considered excellent (above 0.75), good (0.6 – 0.75), fair (0.4 – 0.59) and poor (below 0.4) repeatability (Cicchetti *et al*. 1981). The evaluation of TP and DPbN map and IMRT dose distributions of the whole patient

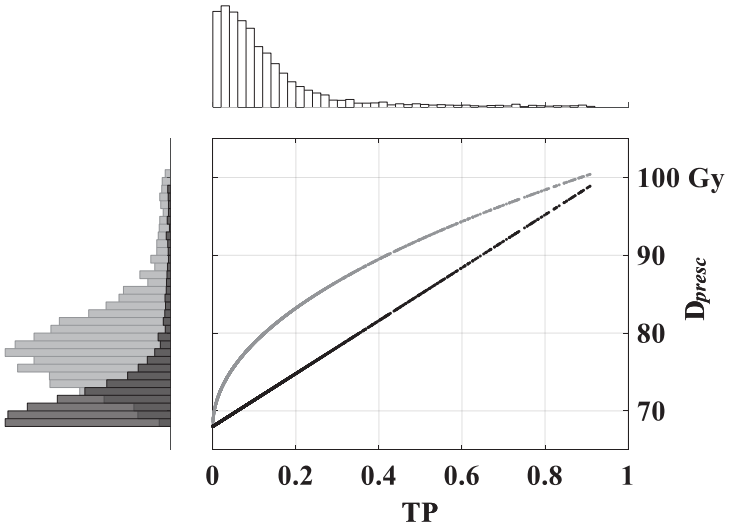

Figure 3. *Example of a linear (black) and square root (grey) dose mapping from tumor probability TP to prescribed dose D_{presc}. Normalized histograms for TP and D_{presc} are shown along the axis indicating the density of data points in the plot. Prescription dose values ranged from 68.0 to 99.0 and 100.5 Gy for linear and square root mappings.*

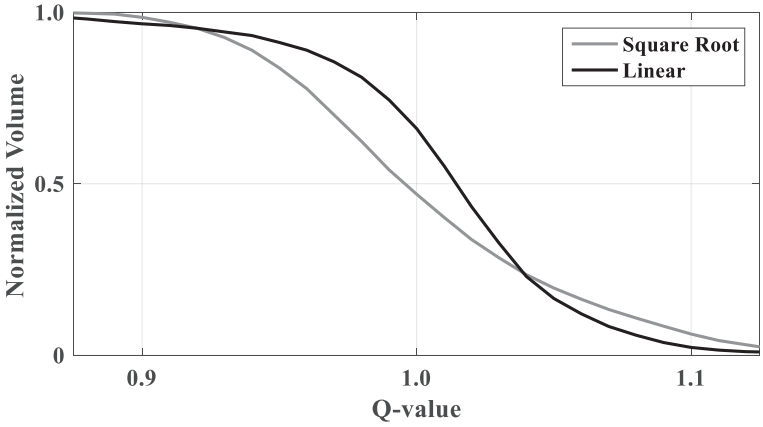

Figure 4. *Example of a QVH. Inverse cumulative histogram of the prostate volume receiving fraction Q of the prescribed dose. For linear dose mapping Q-values below 0.9 are observed in 3.4% of the prostate, and above 1.1 in 2.3% of the prostate. For square root dose mapping 1.5% of the prostate voxels had Q-values below 0.9, while 6.2% of the voxels had Q-values above 1.1.*

group with ICC resulted in excellent agreement at all three stages. Figure 5 shows the ICC values with median values of all patients for linear and square root dose mapping. The median ICC of TP maps was 0.82, ranging from 0.67 to 0.93. Median ICC values of 0.82 and 0.84 were found for the DPbN maps, using linear and square root dose mapping respectively, and ICC values ranged from 0.71 to 0.90. At IMRT stage a median ICC value of 0.88 was observed for linear dose mapping (range 0.71 – 0.93), while for square root dose mapping this was 0.94 (range 0.88 – 0.96). On individual level good to excellent agreement between both planning sessions was shown. Density scatter plots of the IMRT dose values based on both linear and square root dose mapping are provided in Figure 6. Dose values of the first planning session are plotted versus values from the second session and are accumulated over all patients. We observed a symmetric distribution around the diagonal for both dose mapping functions, with 95% of the voxel-to-voxel dose differences below 4.8 and 5.7 Gy for linear and square root dose mapping, respectively. To improve visualization of the low-count pixels in the scatter plot, we log transformed the intensity values.

Median ICC values for the repeatability of the 30 imaging features over the group of patients varied between 0.82 and 0.98. The five most repeatable features were all T2w Gaussian smoothed derivatives of the first (G_y) and second (G_{yy}) order, where y denotes the AP direction of the prostate. The ICC values of these five features ranged between 0.97 – 0.98: $G_y(3.8)$, $G_y(2.4)$, $G_y(6.0)$, $G_{yy}(6.0)$ and $G_{yy}(3.8)$. Here the values between brackets represent the scale (in mm) of the smoothing kernel. The five best performing features according to Dinh *et al.*

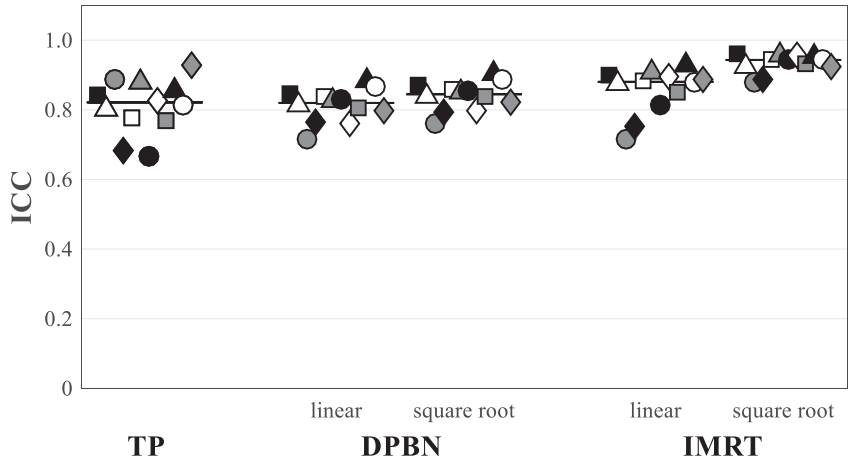

Figure 5. *ICC values of twelve patients at three stages of the treatment planning pipeline for both linear and square root dose mapping. Median ICC per stage is indicated with a horizontal bar. For each patient, identified by symbol-filling combination, repeatability can be traced through the planning pipeline.*

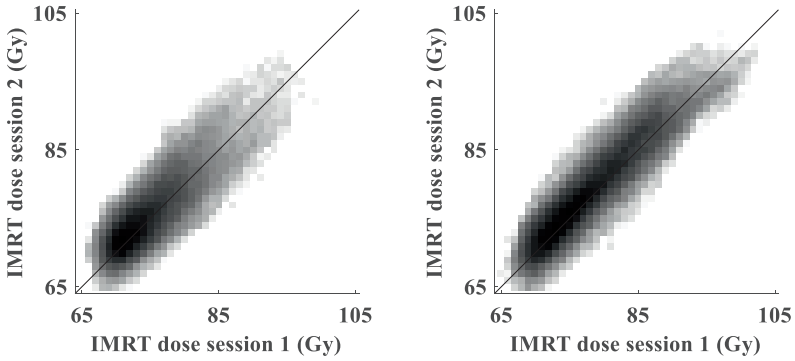

Figure 6. *Density scatter plots of dose values within the IMRT plan based on linear (left) and square root (right) dose mapping. Dose values are accumulated over twelve patients and plotted as session 1 versus session 2. Intensity values were log transformed to improve visualization of the low-count pixels.*

(2017) were ADC intensity, Prevalence map, K^{trans} intensity, $G_{xx}(6.0)$ and $G_{xx}(1.5)$, where x denotes the LR direction of the prostate. ICC values of these features were 0.84, 0.94, 0.85, 0.92 and 0.94, respectively.

Discussion

We investigated the repeatability of mp-MRI-based DPbN treatment planning for prostate cancer. Results from ICC analysis showed excellent repeatability for TP modelling, dose prescription and treatment planning, and QF analysis revealed good agreement between prescribed and planned dose. Simple polynomial dose mapping functions resulted in realistic prescription dose maps, and high repeatability of IMRT planning was observed. These results confirm that prostate cancer radiotherapy based on DPbN leads to stable treatment plans.

Dose prescription maps were derived from the TP maps using a polynomial dose mapping function. In this work we implemented linear and square root dose mappings, and left out squared dose mapping as described by Bowen *et al.* (2009). We calculated a prescription dose map for one patient based on squared dose mapping, but we observed a mapping of more than 90% of the voxels to dose values between 68 and 70 Gy. This almost homogenous dose

distribution can be explained with the skewed TP histogram, which results in an even more skewed prescription dose histogram with a squared dose mapping. For this reason, the squared dose mapping was not included in the test-retest study.

DPbN dose prescriptions with linear and square root dose mappings resulted in comparable repeatability of dose prescription maps. We found that the IMRT treatment plans based on square root dose mapping were more repeatable than the plans based on linear dose mapping. This might be explained by the steeper linear dose mapping functions, particularly at dose levels above 85 Gy, in combination with the limited capability of the treatment planning optimization algorithm to deliver high dose to small, isolated areas. Isolated high dose areas are assumed to give rise to uncertainty in the final treatment plan, which ultimately leads to lower repeatability when compared in a test-retest setting. In these areas also less agreement with the prescribed dose was obtained resulting in lower treatment plan quality. This was confirmed with higher $QF_{95\%}$ and $xQF_{95\%}$ values that were found for linear dose mapping plans compared to square root dose mapping plans. Nevertheless, QF values showed that treatment plans were on average in good agreement with the prescribed dose distributions, and xQF values that were all below 5% indicated that the differences in mp-MRI did not compromise play quality.

Feature repeatability analysis revealed a top five stable features all being Gaussian derivatives in the AP direction of the prostate at different smoothing levels. Feature ranking based on the TP model performance revealed a different top five: ADC intensity, tumor prevalence, K^{trans} normalized intensity, and two second order Gaussian derivatives of the T2w image in the LR direction of the prostate with different smoothing kernel (Dinh *et al.* 2017). The higher ranking of the Gaussian derivative features in the repeatability analysis is explained by smoothening of the stochastic noise which is the main cause of within-patient image variability. ICC values of ADC and K^{trans} intensity maps were 0.84 and 0.85 respectively, indicating high repeatability of these features. Comparable findings on repeatability of ADC and K^{trans} intensity features in MRI prostate imaging are reported in literature, as well as on CT and withß other tissue types. Toivonen *et al.* (2015) reported an ICC of 0.89 for ADC intensity in prostate cancer using MRI, although performed on a ROI basis. Alonzi *et al.* (2010) found an ICC between 0.81 – 0.84 for K^{trans}, and reported a within-patient coefficient of variation (wCV) of 13.9% – 15.8%. Koh *et al.* (2009) reported a high repeatability for ADC measurements in a two-center phase I clinical trial. Padhani *et al.* (2002) used the within-patient standard deviation (wSD) to quantify K^{trans} repeatability in muscle tissue in the pelvic region, and found values between 0.32 – 0.33 min^{-1}. K^{trans} was also shown to be reproducible between CT and MRI with an wSD of 0.03 min^{-1} for median K^{trans} values of 0.10 and 0.08 min^{-1} in MRI and CT respectively (Korporaal *et al.* 2011).

Evaluation of differences between planned and prescribed dose distributions during treatment planning occurred via visual comparison of DVH curves and isodose lines. Unacceptable under- and overdosage was controlled by tuning the weight factors of the dose painting objective functions. Remaining underdosing in treatment plans consistently occurred in high dose prescription voxels that are associated with near-certainty about tumor presence. The underdosing is a limitation of the dose painting objective functions in the sense that they penalized under- and overdosing for the whole prostate volume. The relative small fraction of underdosed high prescription voxels has a relatively minor contribution to the total cost function. As a result, the optimizer is not steered strongly enough to increase the dose to these regions. Figure 2 shows an example of this issue where a small lesion in the peripheral zone with a TP of 0.6 has a planned dose lower than the prescribed dose.

Improvement of the optimization algorithm for IMRT planning may be possible, for example by adding new cost functions. This would result in dose distributions that are more similar to the prescribed dose and would be reflected in smaller values of QF and particularly $QF_{95\%}$. This would also imply that the ICC of the IMRT dose distributions would approach the slightly lower ICC of the DPbN prescriptions. However, taking into account that dose distributions in treatment plans are more blurred compared to prescription dose distributions, higher ICC values for IMRT plans can be expected since dose blurring reduces both the inter-voxel heterogeneity and the part of intra-voxel differences caused by noise, thus resulting in higher repeatability in a test-retest situation.

Although the tumor probability of the last three patients was calculated on a coarser T2w-grid, no apparent differences in the ICC values were observed. We assume that the difference in voxel size at this resolution has no discernable influence on the final repeatability. Instead, resampling artifacts within the planning pipeline and deformable registrations between the planning sessions have a higher impact on the observed ICC values. For a test-retest study resampling and registrations steps are inevitable. Introduction of artifacts in the data could however be minimized with reduction of resampling steps within the planning pipeline.

Throughout the planning pipeline repeatability was above 0.75 and median values indicated that stable treatment plans can be realized. Although median values already show high repeatability, the stability of the planning pipeline on the individual level may be further improved. For example, with selection of image features based on not only performance but also their repeatability, which was not investigated in the current study. Furthermore, only simple polynomial dose mapping functions were implemented to prove the principle of DPbN. Instead of dose mapping functions, tumor control probability models could serve to relate tumor probability to prescribed dose based on radiobiological assumptions.

The pipeline holds the prospect of automation. In this work the prostate and OARs were delineated manually, a process that could be replaced with automatic segmentation software. Also, the treatment planning process used for this study involved manual interventions to ensure that the dose-volume constraints from Table 1 were met. Automatic plan generation software is already available and will be valuable in the development of a fully automatic dose painting by numbers MR-only treatment planning workflow.

Conclusions

Using test-retest mp-MRI, we have shown that DPbN treatment plans for prostate cancer can be realized with excellent repeatability. From tumor probability modelling based on mp-MRI towards treatment plan realization based on voxel-wise dose prescription a stable treatment planning pipeline was demonstrated. Using validated tumor probability modelling and simple dose mapping it was shown that despite day-to-day variations in imaging data consistent treatment plans were still obtained.