



Universiteit  
Leiden  
The Netherlands

## **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>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3192801>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



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



3

# Knowledge-based assessment of focal dose escalation treatment plans in prostate cancer

Marcel A. van Schie

Tomas M. Janssen

Dave Eekhout

Iris Walraven

Floris J. Pos

Peter de Ruiter

Alexis N.T.J. Kotte

Evelyn M. Monninkhof

Linda G.W. Kerkmeijer

Cédric Draulans

Robin de Roover

Karin Haustermans

Martina Kunze-Busch

Robert Jan Smeenk

Uulke A. van der Heide

# Abstract

## Purpose

In a randomized focal dose escalation radiation therapy trial for prostate cancer (FLAME), up to 95 Gy was prescribed to the tumor in the dose-escalated arm, with 77 Gy to the entire prostate in both arms. As dose constraints to organs at risk had priority over dose escalation and suboptimal planning could occur, we investigated how well the dose to the tumor was boosted. We developed an anatomy-based prediction model to identify plans with suboptimal tumor dose and performed replanning to validate our model.

## Materials and methods

We derived dose-volume parameters from planned dose distributions of 539 FLAME trial patients in four institutions and compared them between both arms. In the dose-escalated arm, we determined overlap volume histograms and derived features representing patient anatomy. We predicted tumor  $D_{98\%}$  with a linear regression on anatomic features and performed replanning on 21 plans.

## Results

In the dose-escalated arm, the median tumor  $D_{50\%}$  and  $D_{98\%}$  were 93.0 and 84.7 Gy, and 99% of the tumors had a dose escalation greater than 82.4 Gy (107% of 77 Gy). In both arms organs at risk constraints were met. Five out of 73 anatomic features were found to be predictive for tumor  $D_{98\%}$ . Median predicted tumor  $D_{98\%}$  was 4.4 Gy higher than planned  $D_{98\%}$ . Upon replanning, median tumor  $D_{98\%}$  increased by 3.0 Gy. A strong correlation between predicted increase in  $D_{98\%}$  and realized increase upon replanning was found ( $\rho = 0.86$ ).

## Conclusions

Focal dose escalation in prostate cancer was feasible with a dose escalation to 99% of the tumors. Replanning resulted in an increased tumor dose that correlated well with the prediction model. The model was able to identify tumors on which a higher boost dose could be planned. The model has potential as a quality assessment tool in focal dose escalated treatment plans.

# Introduction

Focal dose escalation to the tumor in prostate cancer radiotherapy has been hypothesized to improve patient outcome without increasing acute and late toxicities (Lips *et al.* 2011). In the multicenter randomized Focal Lesion Ablative Microboost in prostate cancer (FLAME) trial, patients in the dose-escalated arm received an escalated dose up to 95 Gy to the visible tumor. The aim of the trial was to increase the 5-year biochemical recurrence-free survival rate by 10%. To prevent increased toxicity compared to the standard arm, strict dose-volume constraints were imposed on the organs at risk (OARs). During treatment planning, these OAR constraints had priority over dose escalation. Indeed, no significant increase in toxicity was found up to two years after treatment (Monninkhof *et al.* 2018). Because of the OAR constraints being prioritized however, the planned dose escalation to the tumor was limited by the spatial separation between the tumor and the OARs. This raises the question as to how much dose escalation to the GTV was really achieved in the dose-escalated arm of the trial. For this reason, we investigated in the first part of this study to what extent a dose escalation to the visible tumor was realized via comparison of dose-volume parameters between both arms of the trial.

Besides patient anatomy, the degree of dose escalation to the tumor can also be affected by decisions made during optimization of the treatment plan. In clinical practice, it is difficult for a planner or radiation oncologist to assess if a treatment plan can be considered optimal. Over the past years Knowledge Based Planning (KBP) techniques have been introduced to enable automated plan quality assessment (QA) in radiotherapy (Wu *et al.* 2009, Moore *et al.* 2011, Appenzoller *et al.* 2012, Yuan *et al.* 2012, Good *et al.* 2013, Wang *et al.* 2013, Schreibmann *et al.* 2014, Nwankwo *et al.* 2015, Song *et al.* 2016, Shiraishi *et al.* 2016, Wall *et al.* 2018). These studies utilize a database of previously treated patients to guide treatment planning of a new patient, based on similarities of the patient's anatomy with existing ones in the database. For standard prostate treatment planning several KBP techniques were utilized to predict OAR dose from the patient anatomy (Yuan *et al.* 2012, Wang *et al.* 2017, Janssen *et al.* 2019). However, to date no KBP methods have been published to predict achievable focal dose escalation in prostate cancer. In the second part of the study we therefore developed an anatomy-based prediction model using our own database to predict the achievable dose in the tumor. We compared the predicted achievable tumor dose with the tumor dose realized in the clinical plans. We tested the validity of our model and the potential for a QA tool with a replanning of a subset of treatment plans based on our model's predictions.

In this work we present how much of the intended dose was actually planned, how much dose escalation could have been achieved, and how much of the predicted dose escalation could

be realized upon replanning. Focal dose escalation is a promising strategy in prostate cancer. By combining dosimetric evaluation with knowledge-based planning predictions, this study gives a comprehensive overview of the current feasibility and limitations of this dose escalation strategy for prostate cancer, as well as an indication of potential improvements that could be realized upon future clinical implementation.

# Materials and methods

## Patient characteristics

Data from 571 prostate cancer patients who participated in the randomized FLAME trial (clinicaltrials.gov NCT01168479) were used. All patients had biopsy-proven, clinically localized, intermediate or high-risk prostate cancer (Ash *et al.* 2000). Patients were excluded from the trial if they received previous pelvic radiation or underwent prostatectomy, if they had a World Health Organization (WHO) score > 2, an International Prostate Symptom Score (IPSS)  $\geq$  20, a transurethral resection of the prostate (TURP) less than three months prior to treatment, contraindications for MRI, or if they could not discontinue anti-coagulate usage which was required for implanting gold fiducial markers. We obtained approval from the institutional review board and written informed consent from all participating patients.

Patients were treated at four institutions: 320 patients at the University Medical Center Utrecht (UMCU), 93 at the University Hospitals in Leuven (UZL), 109 at the Netherlands Cancer Institute in Amsterdam (NKI), and 49 at the Radboud University Medical Center in Nijmegen (Radboudumc). At each institution patients were randomly and in a 1:1 ratio assigned to the standard and dose-escalated treatment arm. Treating physicians were not blinded for the randomization in order to evaluate and approve the treatment plans.

The primary endpoint of the trial was to achieve an increase in 5-year biochemical recurrence free survival rate of 10% among patients in the dose-escalated arm of the trial as compared to the standard arm patients. To identify biochemical recurrence, the Prostate Specific Antigen (PSA) level in the blood was measured twice per year, and biochemical recurrence was defined as a PSA rise of 2 n/mL above nadir PSA level, according to the Phoenix definition (Roach *et al.* 2006). Treatment-related acute and late toxicity, measures by the Common Toxicity Criteria for adverse events version 3.0 (CTCAE) (Trotti *et al.* 2003), as well as Quality of Life and disease-specific survival were secondary endpoints of the trial.

For this study we considered patients who were included in the per-protocol analyses of the trial (Monninkhof *et al.* 2018). Patients that did not receive the assigned treatment or decided to discontinue the treatment due to anxiety for increased toxicity in the dose-escalated arm were not included in the per-protocol analyses. From the patients eligible for the per-protocol analyses, we excluded three patients who were assigned to the standard treatment arm for which no bladder was delineated. In total 274 patients in the standard arm and 265 patients in the dose-escalated arm were available for analysis.

## Treatment planning and delivery

All patients received a planning CT scan and a pretreatment multiparametric (mp-) MRI exam, including a T2-weighted, diffusion weighted imaging and dynamic contrast-enhanced sequence. The prostate gland was delineated on the T2-weighted MRI by a radiation oncologist. The clinical target volume (CTV) consisted of the prostate gland and, depending on the risk of tumor involvement, the seminal vesicles (SV). For patients who were randomized to the dose-escalated treatment arm, any tumor tissue in the CTV that was visible on the mp-MRI was contoured and defined as gross tumor volume (GTV). After registration of MRI to CT, target volumes and organs at risk (OARs) were defined and delineated. The planning target volume (PTV) was defined as the CTV with a margin of 5 – 8 mm, according to institutional practice. Based on negligible dosimetric impact of PTV margins around intraprostatic GTVs, in this trial no margins were applied to the GTV (van Haaren *et al.* 2009, Lips *et al.* 2009).

The study protocol prescribed a radiation dose of 77 Gy to the PTV in 35 fractions, with an integrated boost up to 95 Gy to the identified tumors of patients in the dose-escalated arm. Depending on institutional practice, 55 to 77 Gy was prescribed to the SV whenever it was included in the CTV. Dose constraints to the OARs followed institutional practice and applied to both arms of the trial. In addition, dose constraints of 77 Gy to 1 cc of the rectum and 80 Gy to 1 cc of the bladder were included. One institution applied an endorectal balloon to further reduce dose to the rectal wall (Radboudumc).

Among the participating institutions different Treatment Planning Systems (TPSs) and delivery techniques were used. The UMCU used PLATO (Nucletron, Veenendaal, The Netherlands) and Monaco (Elekta, Stockholm, Sweden) to generate 7-beam intensity modulated radiation therapy treatment plans for 126 and 183 patients, respectively. The UZL generated 2-arc volumetric-modulated arc therapy treatment plans with Eclipse (Varian, Palo Alto, CA). The NKI and Radboudumc used Pinnacle TPS (Philips Radiation Oncology Systems, Fitchburg, WI) to generate 1- or 2- arc volumetric-modulated arc therapy plans.

## Dose evaluation

A prescription dose map was constructed using the CTV and GTV masks with corresponding prescription dose levels. All dose distributions were resampled to a 1 mm isotropic voxel grid. From the dose distributions we derived dose-volume parameters within the PTV, all GTVs in the prostate, the CTV minus GTV, the bladder and the rectum.

In both study arms we determined the near-maximum dose  $D_{2\%}$  and high-dose volume  $V_{107\%}$  in the CTV minus GTV. The  $V_{107\%}$  was calculated as the volume percentage with a dose escalation above 107% of the prescribed 77 Gy (82.4 Gy). We chose to evaluate the GTV coverage in terms of CTV prescription dose, since the trial prioritized organ at risk sparing over achieving GTV coverage and therefore GTV coverage was not explicitly required. Furthermore, we derived the  $V_{95\%}$  in the PTV and the near-maximum doses  $D_{1cc}$  and  $D_{2cc}$  in the bladder and the rectum. For plans in the dose-escalated arm, we evaluated to what extent we reached the prescribed dose escalation of 95 Gy. We determined the number of plans with a GTV  $D_{50\%}$  and  $D_{98\%}$  above 82.4 Gy. Statistically significant differences between both arms were examined with one-way ANOVA tests. Since we applied several tests, Bonferroni correction for multiple testing was used to correct the significance level.

At one institution (Radboudumc), an endorectal balloon was applied to reduce rectal wall dose and decrease inter- and intrafraction motion. For this relatively small patient group we merged rectal wall and balloon contours to represent the rectum, on which we report dose volume parameters to be in accordance with literature. We compared GTV and rectum dose-volume parameters between patients with and without endorectal balloon in situ to decide if both patient cohorts could be combined for development of a prediction model.

Another 25 patients from UMCU and NKI received adaptive treatment. For these patients a rigid registration of planning and adaptive CT scan was performed, and a weighted sum was applied to the co-registered planned dose distributions. The weights corresponded to the number of treatment fractions that each dose distribution was delivered. Three patients had a replanning CT. In addition to the rigid registration of CT scans, we extracted binary masks of prostate, bladder and rectum and pairwise deformably registered the masks between first and second planning session. The deformable registration involved an implementation of the b-spline deformation algorithm described by Rueckert *et al.* (1999). The normalized cross-correlation similarity measure was used for optimization, and registrations were visually assessed. Replanned dose distributions were mapped accordingly, resulting in locally deformed dose distributions for prostate, bladder and rectum. Planned dose distributions were weighted separately for prostate, bladder and rectum to allow for dose-volume parameter

derivation. In the second part of the study we only considered the initial treatment plans to develop a prediction model on.

## Prediction model

We developed a prediction model that calculated the highest achievable  $D_{98\%}$  in the GTV based on the anatomy of all patients. We chose to predict the near-minimum dose, as this was regarded to be most sensitive to trade-offs between OAR dose and tumor coverage.

We derived Overlap Volume Histograms (OVHs) of delineated PTV, GTV, bladder and rectum to encode the patient's anatomy (Wu *et al.* 2009, Yuan *et al.* 2012). We defined ten structure pairs (PTV→Bladder, PTV→Rectum, GTV→Bladder, GTV→Rectum, PTV→GTV, and vice versa) and derived the OVH of each structure pair on a 1 mm resolution. In case of multiple GTVs per patient, we performed our analysis per GTV to allow for a dose prediction per GTV. Per structure pair we combined the OVHs of all patients and performed Principal Component Analysis (PCA) to reduce dimensionality (Yuan *et al.* 2012). We determined the set of principal components (PCs) that described 90% of the variance in OVHs. We reconstructed the OVHs using the derived PCs and defined the obtained patient-specific coefficients as PC scores. In addition, we added radii  $r_{5\%}$ ,  $r_{50\%}$  and  $r_{95\%}$  corresponding with 5%, 50% and 95% fractional overlap between two structures. In contrast to the PC scores these radii were only dependent on the patient's individual OVHs.

The model we developed combined automatic feature selection with a modified linear regression algorithm to predict the  $D_{98\%}$  in the GTV. Given the complexity of a dose escalated treatment plan, it is difficult to manually assess if a treatment plan was made optimal in terms of highest GTV  $D_{98\%}$  for a given set of anatomical constraints. Due to the large size of the study, we expected the plans in our dataset to range between not optimal and close to optimal planned dose distributions. Therefore, we modified the regression algorithm such that for treatment plans with similar anatomy, a larger weighting was applied to tumors with higher planned  $D_{98\%}$ .

Depending on the anatomy, values for planned  $D_{98\%}$  are expected to lie in the range of 77 – 95 Gy. In some cases a GTV with a low  $D_{98\%}$  may be optimal given the anatomy. To account for a non-uniform distribution of  $D_{98\%}$  values over the dose range, we also applied a weighting of the planned  $D_{98\%}$  that compensated for the sparsity of data points at lower dose. Details on the model's training and validation scheme can be found in the supplementary material.

To verify if inclusion of patients with endorectal balloon in situ did not bias the performance of the model, we retrained the model after exclusion of patients with balloon and compared the pairwise difference between the predicted GTV  $D_{98\%}$  by the two models.

## Evaluation of the model

We determined the dose difference between predicted and planned  $D_{98\%}$  for all GTVs in the dose-escalated arm. We ranked the GTVs according to predicted dose difference in order to make a selection of treatment plans for replanning. We selected five treatment plans with the largest predicted dose difference and another 16 random plans: eight with a GTV with at least 10 Gy predicted dose difference and eight without. Among the largest predicted dose differences no bias towards any of the institutions was observed. Replanning was performed by planning specialists (DE, PR and RR) with ten, nine and three years of experience in treatment planning. The planning specialists were blinded for the predicted dose difference by the model, and instructed to plan the highest achievable dose to the GTVs while adhering to the existing target objectives and OAR constraints. New treatment plans were generated in the original treatment planning system, based on original delineations and according to the FLAME study treatment protocol. Because of decommissioning, seven treatment plans originally planned with the PLATO treatment planning system were replanned using Pinnacle. We compared our predicted tumor  $D_{98\%}$  with the  $D_{98\%}$  obtained upon replanning to evaluate our model. All analyses were performed in MATLAB (MathWorks, Natick, MA, USA).

# Results

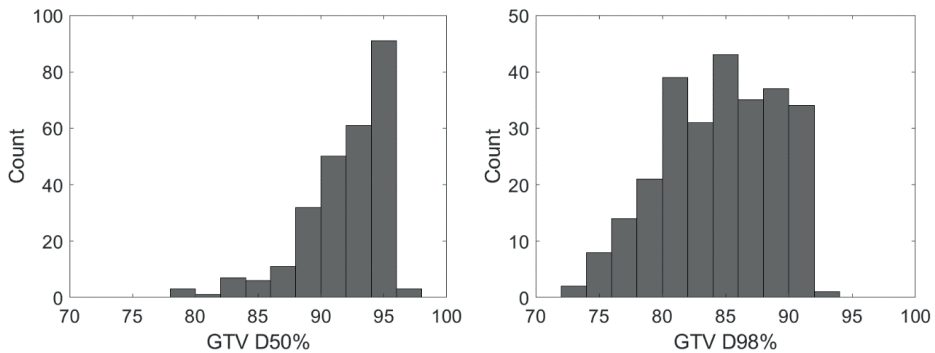
## Dose evaluation

The observed dose-volume parameters for the PTV, GTV, CTV minus GTV, bladder and rectum are described in Table 1. The median  $D_{50\%}$  to the GTV was 93.0 Gy, the median  $D_{98\%}$  was 84.7 Gy. The percentage of GTVs that received a  $D_{50\%}$  above 82.4 Gy (107% of 77 Gy) was 98.7%, and 70.4% received a  $D_{98\%}$  above that level. Histograms of the distribution of GTV  $D_{50\%}$  and  $D_{98\%}$  in the dose-escalated arm are shown in Figure 1.

**Table 1.** Comparison of dose-volume parameters in both arms of the FLAME trial. Median and IQR values are reported.

Structure	Dose-volume parameters	Standard arm (n =274)	Dose-escalated arm (n =265)	P-value*
PTV	V <sub>95%</sub> (%)	98.3 (95.5–98.8)	98.1 (95.3–98.7)	0.127
CTV – GTV	D <sub>2%</sub> (Gy)	79.3 (78.8–79.8)	91.2 (88.6–92.7)	< 0.001
	V <sub>107%</sub> (%)	0.7 (0.0–2.7)	25.9 (17.3–39.2)	< 0.001
GTV	D <sub>50%</sub> (Gy)		93.0 (90.3–94.5)	
	D <sub>98%</sub> (Gy)		84.7 (81.3–88.4)	
	V <sub>95%</sub> (%)		77.6 (50.6–92.0)	
Bladder	D <sub>1cc</sub> (Gy)	75.5 (74.4–76.7)	76.2 (75.0–77.6)	< 0.001
	D <sub>2cc</sub> (Gy)	74.6 (73.7–76.0)	75.2 (74.0–76.6)	0.009
Rectum	D <sub>1cc</sub> (Gy)	74.1 (73.5–74.8)	74.9 (73.7–75.9)	< 0.001
	D <sub>2cc</sub> (Gy)	73.3 (72.5–74.0)	73.5 (72.4–74.4)	0.037

\* Differences were tested with a one-way ANOVA test. A post hoc Bonferroni method was applied to correct the significance level for multiple testing.



**Figure 1.** Histograms of planned D<sub>50%</sub> (left) and D<sub>98%</sub> (right) of 265 patients in the dose-escalated arm.

The median V<sub>95%</sub> in the PTV was 98% in both study arms. The median near-maximum dose D<sub>2%</sub> and high-dose volume V<sub>107%</sub> in the CTV minus GTV were respectively 79.3 Gy and 0.7% in the standard arm, and 91.2 Gy and 25.9 % in the dose-escalated arm, and differed significantly between both arms. The difference is explained by dose gradients surrounding the GTVs in the dose-escalated arm. The median D<sub>1cc</sub> in the bladder and rectum were 75.5 and 74.1 Gy in the standard arm, and 76.2 and 74.9 Gy in the dose-escalated arm, respectively. The median

bladder and rectum  $D_{2cc}$  were 74.6 and 73.3 Gy in the standard arm and 75.2 and 73.5 Gy in the dose-escalated arm, respectively.

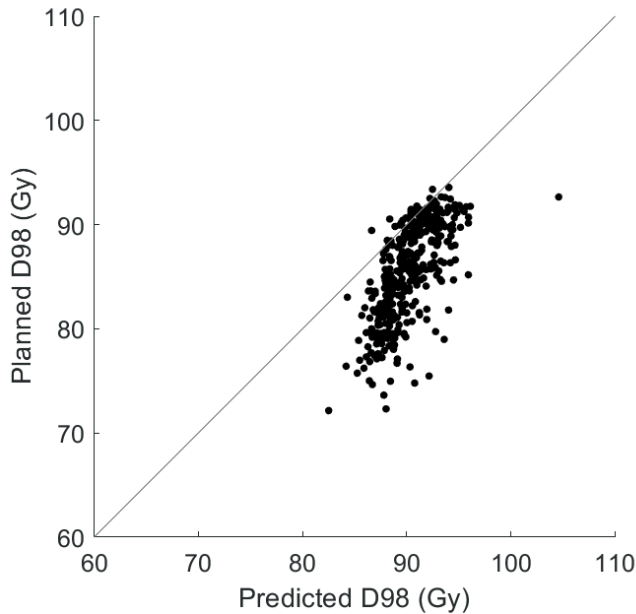
Between patients treated with and without an endorectal balloon in situ we observed minor differences in GTV  $D_{50\%}$  and  $D_{98\%}$  that were non-significant, and small differences in rectum  $D_{1cc}$  and  $D_{2cc}$  that were significant but not exceeding clinical dose constraints. The results of this comparison are presented in the supplementary material, Table A2.

## Prediction model

The model was trained on 382 GTVs. After PCA, four to five PCs were extracted from each OVH. The trained model consisted of five features, listed with corresponding coefficients in the supplementary material, Table A1. The predicted GTV  $D_{98\%}$  is plotted against the planned  $D_{98\%}$  in Figure 2. The influence of the larger contribution of plans with a higher planned GTV  $D_{98\%}$  is reflected by the small fraction of data points above the identity line. Planned GTV  $D_{98\%}$  values are observed up to 95 Gy, which reflects the aim of the trial. Predicted GTV  $D_{98\%}$  in some cases however exceeded the 95 Gy, suggesting that according to the model the anatomy of these patients would allow for further dose escalation. In one extreme case a GTV  $D_{98\%}$  of 104.6 Gy was predicted, which appeared to be a small tumor at a relative large distance from the rectum.

The median dose difference between predicted and planned  $D_{98\%}$  was 4.4 Gy, and dose differences ranged between -2.8 Gy and 16.7 Gy. In 135 of 265 patients who received a focal dose escalation, at least for one GTV an achievable increase of 5 Gy was predicted.

Between the prediction models trained with and without inclusion of patients treated with endorectal balloon in situ, we observed a median pairwise difference of 0.0 Gy (95% confidence interval -0.4 – 0.1 Gy), which justified the inclusion of patients with balloon in our presented prediction model. The difference between predicted and planned  $D_{98\%}$  of both models can be found in the supplementary material, as well as a scatter plot of the pairwise difference in predicted  $D_{98\%}$  between the two models (Figure A2).

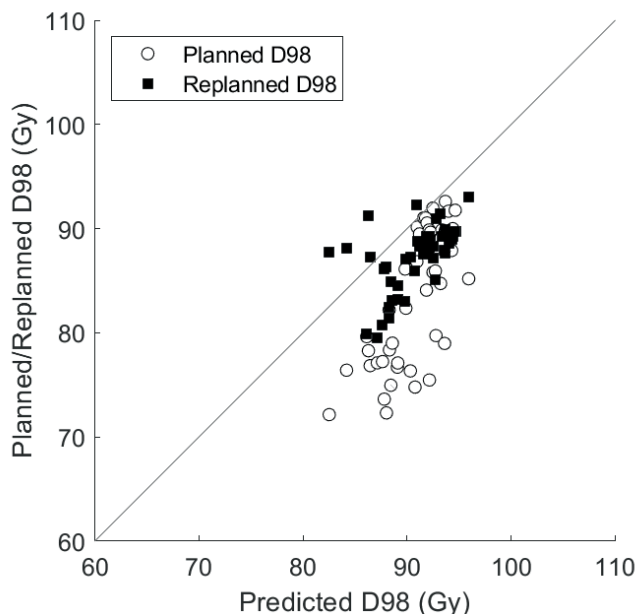


**Figure 2.** Scatterplot of predicted  $D_{98\%}$  versus planned  $D_{98\%}$  in the GTV after training of the modified linear regression model. The majority of data points can be observed below the identity line as a result of the asymmetric cost function.

## Evaluation of the model

The 21 treatment plans selected for replanning involved 43 GTVs. We plotted the predicted GTV  $D_{98\%}$  from our model versus the planned  $D_{98\%}$  and the replanned  $D_{98\%}$  in Figure 3. Before replanning we observed a median dose difference between predicted and planned GTV  $D_{98\%}$  of 7.5 Gy (0.5 – 16.7 Gy). After replanning the median dose difference between predicted and replanned GTV  $D_{98\%}$  was 3.8 Gy (-5.2 – 7.7 Gy). A strong correlation ( $\rho = 0.86$ ) was observed between the predicted increase in  $D_{98\%}$  and realized increase in  $D_{98\%}$  after replanning.

A median increase from planned to replanned GTV  $D_{98\%}$  of 3.0 Gy (-4.0 – 16.9 Gy) was found. For GTVs with a planned  $D_{98\%}$  below 80 Gy we observed a median increase of 10.4 Gy. GTVs with a planned  $D_{98\%}$  between 80 and 85 Gy had a median increase of 4.9 Gy, between 85 and 90 Gy a decrease of 0.7 Gy, and above 90 Gy a decrease of 1.7 Gy. Below 85.9 Gy, all GTV's



**Figure 3.** Scatterplot of the predicted  $D_{98\%}$  in the GTV versus the planned and replanned  $D_{98\%}$ . Upon replanning an increase in GTV  $D_{98\%}$  can be observed which correlates with the predicted increase in  $D_{98\%}$ .

$D_{98\%}$  increased upon replanning. In 16 out of 43 GTVs we observed a decreased  $D_{98\%}$  after replanning, with a median decrease of 1.4 Gy (range 0.2 – 4.1 Gy). These GTVs had a relative high median  $D_{98\%}$  of 90.1 Gy, which reduced the likelihood of improved tumor dose after replanning. For the five treatment plans that were selected based on largest predicted dose difference, the median difference between planned and replanned GTV  $D_{98\%}$  was 9.0 Gy. For the 16 randomly selected treatment plans this was 1.7 Gy.

We observed comparable median PTV  $V_{95\%}$  of 97.1% before and 97.9% after replanning. Median bladder  $D_{1cc}$  and  $D_{2cc}$  were respectively 77.0 and 75.8 Gy before, and 76.8 and 75.4 Gy after replanning, while median rectum  $D_{1cc}$  and  $D_{2cc}$  were respectively 74.3 and 73.3 Gy before, and 75.8 and 74.4 Gy after replanning. The small increase in dose to the rectum was expected to correlate with an increased dose to the GTV. Maximum bladder and rectum dose were still in accordance with clinical constraints.

# Discussion

We showed that integrated focal dose escalation in the prostate is feasible with a median dose above 107% of the standard dose of 77 Gy achieved in 99% of the patients. Observed dose-volume parameters show a median GTV  $D_{50\%}$  of 93.0 Gy, which was close to the intended 95 Gy, and a median  $D_{98\%}$  of 84.7 Gy.

We also developed a prediction model based on overlap volume histograms and planned  $D_{98\%}$  to identify GTVs for which a higher escalated dose was regarded feasible. After replanning of a subset of treatment plans, we observed a considerable increase in planned  $D_{98\%}$  which strongly correlated with the predicted increase by the model.

A recent trial on dose painting in prostate cancer (HEIGHT trial) with up to 89.3 Gy in 38 fractions found a GTV  $V_{95\%}$  (above 84.8 Gy) between 95.2% and 99.8% (Bossart *et al.* 2016). Because of the different fractionation scheme and level of dose escalation, a comparison with our results could not be made.

$D_{2\%}$  and  $V_{107\%}$  in the CTV minus GTV as well as  $D_{1cc}$  in the bladder and the rectum showed a significant increase of dose in the dose-escalated arm, while  $D_{2cc}$  in the bladder and the rectum did not. These findings can partially be explained by the study protocol that allowed for dose escalation in the healthy prostate, provided that dose-volume constraints to OARs were not violated.

In the dose-escalated arm there were 382 GTVs in 265 plans, which on average was 1.4 GTVs per plan. These findings are in agreement with Van Schie *et al.* (2018). A higher average of 2.0 GTVs per plan was observed in the replanning selection of 21 plans. The higher average in the replanning selection can partly be explained with statistics since a plan with multiple GTVs had an a priori higher chance of inclusion in the replanning selection. We also observed an overestimation of the achievable  $D_{98\%}$  as compared to the planned dose upon replanning. One explanation is the design of the trial, in which we aimed for an escalated dose up to 95 Gy. Our model however was not restricted by this dose constraint and, based on patient anatomy, could in principle predict a higher achievable escalated dose than 95 Gy. While both observations can partly be assigned to statistics and trial design, we do believe they are to some extent also explained by the limitation of our prediction model that did not consider the effect of multiple GTVs per prostate. The model determined the achievable GTV  $D_{98\%}$  for each GTV individually, which could lead to violation of OAR dose constraints in case of multiple GTVs within the prostate. During replanning, this likely has resulted in a reduced GTV  $D_{98\%}$ , since OAR dose constraints were prioritized.

We demonstrated that focal dose escalation was achieved in almost all patients in the dose-escalated arm of the trial. Although several trials have hypothesized clinical benefits of focal dose escalation, no KBP methods exist to predict the highest achievable integrated boost dose to the tumor. Here we demonstrated a novel methodology that, using anatomical features and based on a heterogeneous dataset, could predict the highest achievable dose in the GTV and allowed to identify GTVs for which the escalated dose could be improved. A limitation of existing KBP methods is that the predicted dose range reflects the range of clinical plans. In our model we introduced an upward bias to predict the highest achievable dose, by putting extra weight on the better optimized plans in the database. Since it was trained on data from multiple institutions, the model is robust to different treatment planning systems. We recognize that our model does not allow for a precise estimation of the achievable tumor dose. We do however believe that our model can assist as a QA tool to identify GTVs that could be planned with a higher escalated dose.

Focal dose escalation is a promising dose escalation strategy in prostate cancer. By combining dosimetric evaluation with knowledge-based planning predictions we were able to demonstrate the feasibility of focal dose escalation up to 95 Gy in the prostate, as well as presenting a methodology to potentially improve on focal dose escalation treatment plans in a clinical setting. Although developed for a novel dose escalation strategy in prostate cancer, we believe our methodology can be of general applicability to other treatment sites and radiation strategies as well.

## Conclusions

Focal dose escalation in prostate cancer was feasible in almost all GTVs, with an escalated dose well above the standard prescribed dose. We developed a prediction model to identify GTVs for which a higher escalated dose was considered achievable. Using this model to select plans for replanning, a considerable increase in  $D_{98\%}$  was found achievable, specifically for lower planned  $D_{98\%}$ . Our prediction model has potential as a QA tool and identify suboptimal GTV doses to be optimized via replanning.

# Supplementary Material

## Prediction model

### Materials and methods

To predict the highest achievable  $D_{98\%}$  in the GTV, we developed a model based on patient anatomy, expressed by principal component (PC) scores and fractional overlap scores from patient's OVHs. We applied a stepwise multiple regression model (SMRM) to find the optimal combination of anatomical features to predict the  $D_{98\%}$  (Yuan *et al.* 2012). We modified MATLAB's *stepwiselm* function available from the *Statistics and Machine Learning Toolbox*. The cost function was expanded with a penalty term such that for GTVs with comparable anatomy, the GTV with highest  $D_{98\%}$  would have a larger influence on the regression. A weighting of the planned  $D_{98\%}$  was added to compensate for sparsity of data at lower dose. The resulting asymmetric least squares optimization algorithm used the following cost function:

$$\text{Cost} = \sum_{i=1}^n (D_{\max} - D_{\text{plan},i})^2 \cdot (D_{\text{pred},i} - D_{\text{plan},i})^2 \cdot (\text{sgn}(D_{\text{pred},i} - D_{\text{plan},i}) - \alpha)^2. \quad (\text{A1})$$

The first term in this equation describes the weighting function, the second term corresponds to ordinary least squares optimization, and the third term penalizes predictions below the planned  $D_{98\%}$ .  $D_{\text{plan},i}$  and  $D_{\text{pred},i}$  are the planned and predicted  $D_{98\%}$  of data point  $i$ ,  $\text{sgn}$  is the sign function and  $\alpha$  the asymmetry parameter.  $D_{\max}$  was set to 102 Gy (107% of 95 Gy),  $\alpha$  was set to 0.9.

We had 265 treatment plans with 382 GTVs available and trained the prediction model on the entire dataset. During training, feature set selection was based on 10-fold cross validation, evaluating the Cost in the validation set. Since the SMRM performs feature set selection based on a preset p-value, we tested a range of p-values (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1). The range of p-values tested was determined during development of our model. For lower p-values the model would not include any feature, and for higher p-values calculation time increased drastically and resulted in decreased model's accuracy. For each p-value we identified the most frequently selected feature set and corresponding Cost. If feature sets were identical to the most frequent selected feature set, we selected the minimum of corresponding Costs. To be able to select the optimal p-value and optimal feature set, and increase robustness of the selection strategy, we repeated this procedure ten times. After ten repetitions we calculated the median Cost per p-value, and selected the p-value corresponding to the lowest median Cost. The final selected feature set was the feature set

that corresponded to the selected p-value and lowest Cost. Final weights were determined via fitting of the selected features to the complete data set using the asymmetric cost function.

To investigate the robustness of our training strategy, we also randomly designated 25% of the GTVs as test set, stratified by treatment plan. A schematic representation of our training strategy is depicted in Figure A1.

### Results

During OVH calculation, from most OVHs four PCs were extracted that explained 90% of the variance in the OVHs. For the structure pairs PTV→Bladder, PTV→Rectum, and Bladder→GTV these were five PCs. Together with the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> OVH curve fractional overlaps scores, this yielded 73 anatomical features that were used to train the prediction model. A p-value of 0.01 resulted in the lowest Cost of 0.627 during training of the model.

In the robustness test we found an average Cost of 0.609 in the training data, and an average Cost of 0.748 in the test data, demonstrating our model was robust when applied to different datasets.

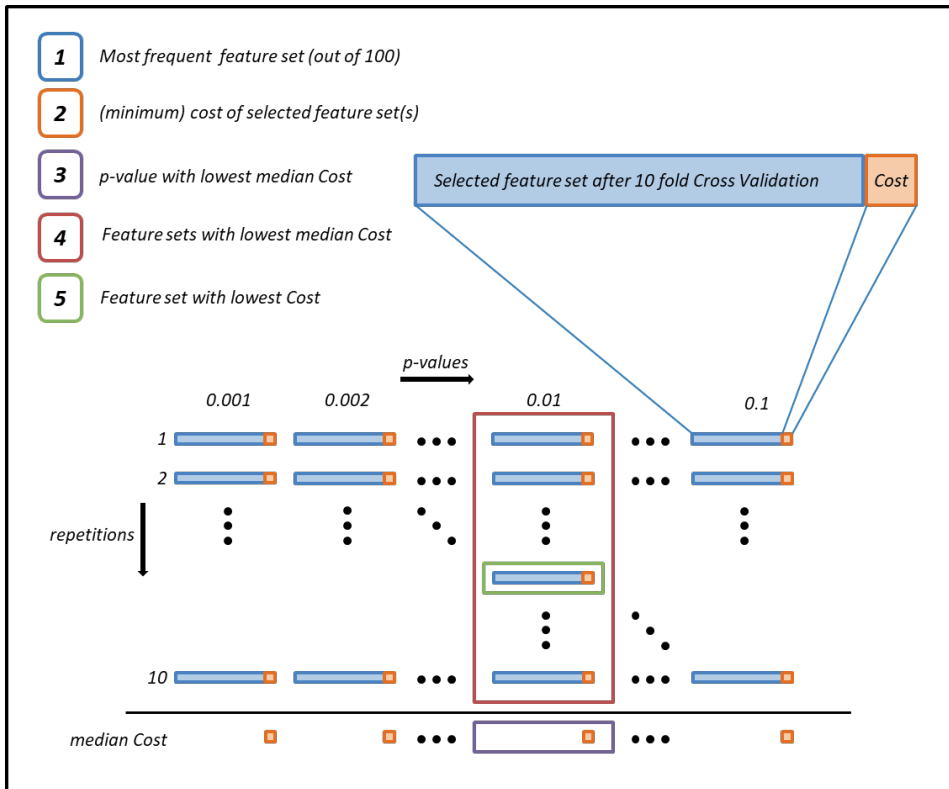
## Endorectal Balloon

### DVH parameters

In Table A2 dose-volume parameters in the GTV and rectum are compared between patients treated with and without endorectal balloon in situ. Significant dose differences are observed for the rectum, although  $D_{1cc}$  and  $D_{2cc}$  values were within accepted dose limits.

### Prediction model

In Table A3 we reported the selected features of the prediction model after exclusion of patients treated with an endorectal balloon in situ. The observed median difference between predicted and planned GTV  $D_{98\%}$  was 4.4 Gy (95% CI -0.2 – 10.7 Gy). For these patients, the median difference in the model trained on all patients was 4.3 Gy (95% CI of -0.3 – 10.4 Gy). The difference between predicted GTV  $D_{98\%}$  by both models was minimal: 0.0 Gy (95% CI -0.4 – 0.1 Gy). The difference between the scatter plot distributions of both models is shown in Figure A2. Considering the minimal differences in GTV  $D_{98\%}$  and scatter plot distributions, the model including all patients was favored for presentation in the article and selection of treatment plans for replanning.



**Figure A1.** Schematic representation of the training strategy of our prediction model. Steps in the training strategy are color coded in the top left corner.

**Table A1.** List of features in the trained prediction model, together with coefficients.  $D_{98\%}$  was determined using  $D_{98\%} = D_0 + \sum(\beta_i * x_i)$

Description of selected features $x_i$	Coefficients $\beta_i$
Offset	66.96 Gy
PC2 of OVH Bladder → GTV	0.048
PC4 of OVH Bladder → GTV	-0.092
$r_{5\%}$ of OVH GTV → Rectum	0.350
PC2 of OVH Rectum → GTV	0.076
PC3 of OVH Rectum → GTV	-0.405

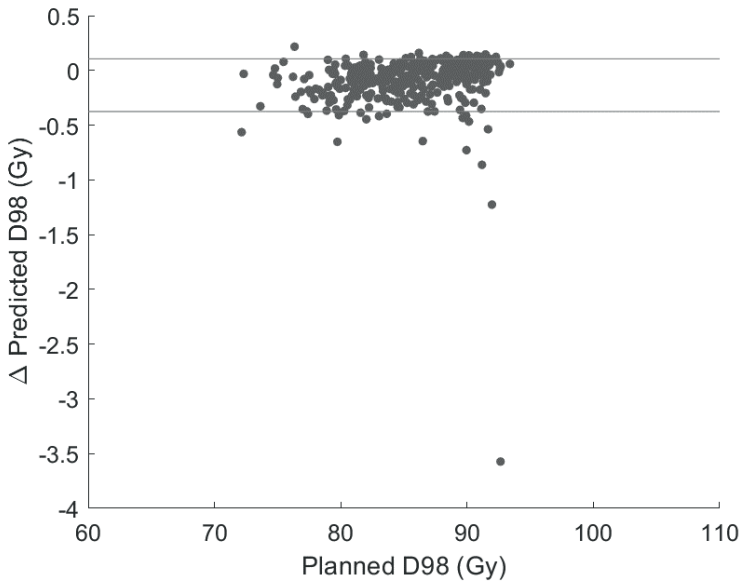
**Table A2.** Comparison of dose-volume parameters between patients treated with and without endorectal balloon in situ. Median and IQR values are reported.

Structure	Dose-volume parameter	Treatment arm	Without balloon (n=493)	With balloon (n=46)	P-value*
GTV	D <sub>50%</sub> (Gy)		84.8 (81.4 – 88.4)	82.5 (79.2 – 88.0)	0.05
	D <sub>98%</sub> (Gy)		92.8 (90.1 – 94.4)	93.9 (91.6 – 95.7)	0.29
Rectum	D <sub>1cc</sub> (Gy)	Standard	74.1 (73.5 – 74.7)	75.1 (74.1 – 76.3)	<b>&lt;0.001</b>
		Dose-escalated	74.8 (73.6 – 75.9)	75.3 (74.8 – 76.0)	0.14
	D <sub>2cc</sub> (Gy)	Standard	73.2 (72.5 – 73.8)	74.7 (73.7 – 76.0)	<b>&lt;0.001</b>
		Dose-escalated	73.3 (72.3 – 74.3)	74.6 (73.9 – 75.1)	<b>0.001</b>

\* Differences were tested with a one-way ANOVA test. A post hoc Bonferroni method was applied to correct the significance level for multiple testing.

**Table A3.** List of features and corresponding coefficients in the trained prediction model using patients treated without endorectal balloon in situ.

Description of selected features $x_i$	Coefficients $\beta_i$
Offset	66.57 Gy
PC2 of OVH Bladder → GTV	0.050
PC4 of OVH Bladder → GTV	-0.111
r <sub>5%</sub> of OVH GTV → Rectum	0.358
PC3 of OVH Rectum → GTV	-0.377



**Figure A2.** Scatter plot of the planned GTV  $D_{98\%}$  versus the difference in predicted GTV  $D_{98\%}$  between the trained prediction models using all patients and only patients treated without endorectal balloon in situ. A minimum difference of -3.6 Gy was observed, corresponding to the extreme case in Figure 2. Due to the relative small size and large distance from the rectum, this GTV was sensitive to the composition of the training dataset.